Metagenomic sequencing of soil cores shows taxonomic partitioning of root placement by depth in a prairie plant community

Christopher K. Black[1,2], Scott A. Woolbright[2,3], Taylor L. Pederson[2], Christopher R. Sligar[2], Evan H. DeLucia[1,2,4]

1. Department of Plant Biology, University of Illinois at Urbana-Champaign
2. Carl. R. Woese Institute for Genomic Biology, University of Illinois
3. University of Arkansas at Little Rock
4. Institute for Sustainability, Energy, and Environment, University of Illinois

Running Head

Prairie root partitioning by depth

Corresponding Author

Evan H. DeLucia

Department of Plant Biology

University of Illinois at Urbana-Champaign

265 Morrill Hall

505 South Goodwin Ave.

Urbana, IL 61801

Phone: (217) 333-6177

Fax: (217) 244-7246

Email: delucia@illinois.edu

Section information

Text Section	Words	Non-text Section	Count	In color?
Summary	196	Figures	5	Figs. 2, 4, 5
Introduction	695	Tables	1	No
Materials and Methods	1519	Supporting Figures	5	Figs. S2, S4, S5
Results	741			
Discussion	866			
Acknowledgements	147			
Total in main text	3968			

Summary

- The ways diversity affects ecosystem resilience are central to climate change responses, yet poorly known for belowground contexts. In particular, it is unclear to what extent different plants change their rooting behavior in response to other taxa and whether these changes improve assemblage partitioning of functional and spatial niches.
- In mixed root samples from varying depths of a tallgrass prairie restoration managed for biomass production, we identified taxa by high-throughput microfluidic PCR and sequencing of ITS2 barcodes, and asked whether roots from different functional groups had recognizable spatial roles.
- Grass prevalence increased with depth while forbs from the Asteraceae decreased, making grasses and asters show negative co-occurrence (spatial segregation). Most other taxa were found equally across depths and showed random or positive co-occurrence, suggesting this site has little overall spatial partitioning. Aboveground cover and belowground abundance were positively correlated, but with less root per shoot for grasses than for dicots.
- In this managed, high-evenness system on uniform, fertile soils, the relatively small fraction of biomass found as grass roots in deep soil may be disproportionately important for functions such as water uptake, N retention, and C storage that affect the resilience of the whole ecosystem.

Keywords

biodiversity, DNA metabarcoding, grassland, niche partitioning, root community, tallgrass prairie

Introduction

In a rapidly changing world, one of the most pressing questions in ecology is which ecosystem properties promote stability and which lead to rapid state changes. It is now well recognized that diversity generally promotes resilience in ecosystem productivity (Tilman *et al.*, 2001), partly through facilitation but primarily through functional complementarity that allows more effective partitioning and therefore more complete exploitation of niche space, producing a more stable overall allocation of resources across the community (Fornara & Tilman, 2008, 2009; De Deyn *et al.*, 2008). This implies that the resilience benefits of niche partitioning will only be realized if different species in the community possess functions that are actually complementary to each other. For example, N and water availability are often strongly anticorrelated in grassland soils, leading *Festuca ovina* and *Achillea millefolium* to extract N from different depths when competing in mixed-species stands than when they are grown alone, while *Phleum pratense* and *Trifolium pratense* showed no such vertical partitioning (Jumpponen *et al.*, 2002).

In addition to functional complementarity, the resilience of grassland systems may rely on diversity in *spatial* partitioning as well. Prairie plant communities are noted for maintaining a large proportion of their biomass below ground (Jackson *et al.*, 1996), and the importance of belowground interactions between species with differing root traits for determining community productivity has long been recognized (Weaver, 1919; Bardgett *et al.*, 2014). Grassland plant communities may be more strongly structured belowground than aboveground (Kesanakurti *et al.*, 2011; Hiiesalu *et al.*, 2012; Wilson, 2014) and the physiological capacity to exploit soil resources may be a strong predictor of competitive success (Tucker *et al.*, 2011; de Kroon *et al.*, 2012; Hendriks *et al.*, 2015). Therefore, to understand the potential responses of multispecies communities to destabilized conditions such as a changing climate, it is necessary to understand the physical placement of the different species in the community as well as the between-species differences in resource allocation and niche partitioning strategies.

Progress in understanding these belowground dynamics is slowed by the challenges of studying belowground environments, especially for mixed-species samples. Because harvesting root systems requires digging and root systems show extreme spatial

variability that increases the sample size needed to characterize a system, all methods are laborious and most require destructive harvests, often in the form of massive excavations to identify roots by physically tracing them to the plant crown (Weaver & Voigt, 1950). Visual identification of roots based on anatomical or morphological characteristics is sometimes possible (Wardle & Peltzer, 2003) but is an inherently low-throughput method that requires a trained specialist and becomes much more challenging for very fine roots or highly diverse communities. Minirhizotron methods can provide some information on root placement, but have limited success at distinguishing between species (Rewald *et al.*, 2012).

The ever decreasing costs of massively parallel DNA sequencing technologies provide a partial solution to this dilemma. Sequencing-based methods can use smaller samples and therefore less destructive sampling techniques, allow distinctions between visually identical roots, and permit quantitative comparisons. Additionally, because DNA persists to some extent in dead tissue and freely in the environment (i.e. environmental or eDNA), sequencing can potentially integrate longer timescales by detecting species that are present in the community but were dormant or senesced at the time of sampling.

The internal transcribed spacer between the 5.8S and 26S regions of the nuclear ribosomal DNA (ITS2) is a popular choice for barcoding mixed root species recovered from soil because it is easily extracted from all plant groups, widely used in root sequencing (Jackson *et al.*, 1999) and therefore well represented in databases, and short enough (*c.* 200 bases) to sequence on an Illumina platform. Although higher taxonomic resolution can be obtained from multiple barcodes (e.g. *rcbL* and *matK*; Staats *et al.*, 2016), assigning identities to multiple barcodes in mixed samples is challenging (Rewald *et al.*, 2012) and ITS2 alone provides comparable resolution to other single barcodes (Chen *et al.*, 2010; Fahner *et al.*, 2016).

The objective of this study was to identify the root species present at varying soil depths (0-100 cm) in a constructed prairie in central Illinois, and to use these identities to infer differences in species roles for water usage, N uptake, and carbon storage. We used a DNA metabarcoding approach in which we collected fine root samples by coring and extracted whole-community DNA for ITS2 amplicon library construction (Fluidigm Access Array) followed by sequencing (Illumina MiSeq 2x300), then identified the

111 detected sequences to genus level by database search and compared the resulting
112 abundance estimates against those from aboveground surveys.
113
114 **Materials and Methods**
115 *Experimental site*
116 Our experimental site is the University of Illinois Energy Farm (Urbana, Illinois,
117 USA: 40.06N, 88.19W, elevation 220 m), where five plots of tallgrass prairie were
118 planted in 2008 (four 0.7 ha plots plus one 3.8 ha plot). The site has a continental climate
119 with a mean annual temperature of 11°C and *c.* 1 m of precipitation annually. It is
120 established on deep, highly fertile Mollisol soils (Argiudolls, mapped as Dana, Flanagan
121 and Blackberry silt loam) with organic C contents more than 1% throughout the top 50
122 cm (Fig. 1). The site was used for agriculture for at least 100 years before establishment
123 of the current experiment. For further details on the establishment and management of the
124 site, see previous work by Zeri et al. (2011), Smith et al. (2013), and Masters et al.
125 (2016).
126 The prairie plots were established by seeding with a mix of 28 species native to
127 Illinois (Table 1). Species were chosen to balance four stated goals: Maximum biomass
128 production, ecosystem resilience, trophic diversity, and locally adapted species. Seeds
129 were obtained from Pizzo and Associates (Leland, IL), treated at planting time with a
130 mycorrhizal inoculum ('AM 120'; Reforestation Technologies International; Gilroy CA)
131 that is primarily *Glomus intraradices* (Neal Anderson, RTI Inc.; personal
132 communication), and overseeded with a spring oat (*Avena sativa*) cover crop. The plots
133 were mowed after senescence each year and the aboveground biomass was baled and
134 removed. Of these 28 species, 25 were detectable in 2011 (Feng & Dietze, 2013) and in
135 2012 a total of 32 species, five of them spontaneously established, were censused in the
136 plots. The five most common species in 2012 were *Andropogon gerardii*, *Sorghastrum*
137 *nutans*, *Ratibida pinnata*, *Helianthus grosseserratus*, and *Coreopsis tripteris*, collectively
138 providing *c.* 75% of total cover (Table 1).
139 *Sample collection*
140 To characterize the spatial distribution of species with depth, we collected mixed
141 root samples on July 15-18 of 2013, after most late-season grasses were well emerged but

before most early-season plants began to senesce. At 24 locations (one in each quadrant of each 0.7 ha plot, two in each quadrant of the 3.8 ha plot), we used an 8 cm diameter bucket auger to collect roots and soil from five depth increments (0-10, 10-30, 30-50, 50-75, and 75-100 cm) of three cores within a 2 m radius. We pooled all three cores from each location, collected a subsample (*c.* 0.5-1 kg) of mixed roots and soil from each depth, and returned the remaining material to the holes. The resulting 120 samples were stored on ice in one-gallon polyethylene bags for transport to the laboratory, then frozen at -80 °C the same day and stored until further analysis.

To characterize the genetic diversity of our target species and generate a mock community for use as a sequencing control, we collected voucher specimens on August 31 and September 1, 2013. For each of the 32 plant species present in aboveground surveys, we located 3-5 individuals, identified them to species, and used a trowel to extract roots still attached to these well-identified stems. We pooled all roots from each species, placed them in polyethylene bags, placed them on ice for transport to the laboratory, and froze them at -80 °C the same day for storage until further analysis.

Root recovery

To separate roots from soil, we thawed mixed samples overnight at 4 °C, then screened them through a 2 mm sieve followed by manually picking all visible roots using forceps. The picking process took *c.* 30-90 min per sample and all sieves, forceps, and gloves were wiped with ethanol immediately before use to minimize contamination by non-sample DNA. After root picking, a subsample of the root-free bulk soil was collected and lyophilized, then ground and combusted to determine total carbon and nitrogen content (4010 CHNSO Elemental Analyzer; Costech, Valencia CA).

We then rinsed all roots in three changes of sterile water, with the final rinse including 10 min of sonication to dislodge any residual soil from the root surface, then lyophilized all samples and stored them at room temperature. Single-species root voucher samples were treated identically to the mixed root samples, with the exception that bulk soil had been removed at collection time and therefore no sieving or hand-picking steps were necessary.

DNA extraction and amplification

To maximize extraction of DNA from tough root tissue, we ground all samples once in a dry mortar and pestle at room temperature, then again in liquid nitrogen to a very fine powder. We then weighed *c.* 100 mg of tissue from each sample and extracted whole DNA using a Powersoil-htp isolation kit (Mo Bio Laboratories, Carlsbad CA) according to the manufacturer's directions, including an optional initial bead-beating step (GenoGrinder; 4 x 30 s at 1750 RPM, 2 min between rounds) followed by a 60 min incubation at 65 °C with 1% proteinase K. We then performed a post-extraction cleanup using materials from the same kit (E. Adams, Mo Bio; personal communication) as follows. We diluted the DNA to a volume of 100 µl with DNAse-free water, added 50 µl of bead beating solution and inverted to mix, then added 25 µl each of solutions C2 and C3, inverted to mix, and centrifuged at 10000 ***g*** for 2 min. We then collected the supernatant, added 2 volumes of solution C4, vortexed, and loaded the sample onto a spin filter. Finally, we washed with 500 µl of solution C5 and eluted with 50 µl of C6.

After extraction, we submitted whole DNA to the W.M. Keck Center (Urbana, IL, USA) for amplification and sequencing. In addition to DNA extracted from mixed samples, we included water extractions as a negative control on the DNA extraction + PCR + sequencing process taken as a whole (Fig. S1), a mock community of DNA from 29 species combined in equimolar quantity plus one species (*Heliopsis helianthoides*) at twice the concentration of the others (Fig. S2), and nominally pure extracts of root DNA (Fig. S3) from four species with high aboveground abundance (*Andropogon gerardii*, *Sorghastrum nutans*, *Silphium perfoliatum*, and *Elymus nutans*; Feng, 2014). The second internal transcribed spacer of the nuclear ribosomal gene (ITS2) was amplified from each sample by microfluidic PCR using the Fluidigm Access Array chip and sequenced by synthesis for 2x301 paired-end cycles (MiSeq, V3 chemistry, Illumina Inc, San Diego CA). The 5' linker construct for each sequence consisted of Illumina flowcell-binding primer i5, Fluidigm linker pad CS1, and plant-specific ITS2 primer S2F (Chen *et al.*, 2010) for a final forward linker sequence of `5'-AATGATACGGCGACCACCGAGATCT-ACACTGACGACATGGTTCTACA-ATGCGATACTTGGTGTGAAT`. The 3' linker construct for each sequence consisted of the Illumina flowcell-binding primer i7, a 10-base oligonucleotide barcode that was unique for each sample, Fluidigm linker pad CS2, and

plant-specific ITS2 primer S3R (Chen *et al.*, 2010) for a final reverse linker sequence of 5'-CAAGCAGAAGACGGCATACGAGAT-XXXXXXXXXX-TACGGTAGCAGAGACTTGGTCT-GACGCTTCTCCAGACTACAAT, where XXXXXXXXXX indicates the variable barcode sequence.

Data processing

The raw Illumina read files were preprocessed by sequencing center staff before delivery by using CASAVA 1.8 to remove sequencing primers, PhiX reference reads, and reads from unrelated samples sequenced in the same flowcell lane. We then used cutadapt 1.8.1 (Martin, 2011) to trim primers, discard all reads that did not begin with the expected primer, and trim 3' bases with a Phred quality score below 20. We then joined the overlapping ends of each read using the RDP maximum likelihood algorithm (Cole *et al.*, 2013) as implemented in Pandaseq 2.10 (Masella *et al.*, 2012) using a minimum alignment quality of 0.8, a minimum assembled length of 25 bases, and a minimum overlap of at least 20 'bits saved' (corresponds to *c.* 10 bases; see Cole *et al.*, 2013). We then used the `split_libraries_fastq.py` script in QIIME 1.9.1 (Caporaso *et al.*, 2010) to assign sequences to samples by barcode matching.

To assign sequences to taxonomic units, we used a de novo clustering approach. We dereplicated sequences and removed singletons and suspected PCR chimeras using VSEARCH 2.0.4 (Rognes *et al.*, 2016), extracted full-length ITS2 variable regions using ITSx 1.0.11 (Bengtsson-Palme *et al.*, 2013), clustered the results using VSEARCH with a similarity threshold of 99%, and assigned taxonomy to each cluster using BLAST+ (Camacho *et al.*, 2009) against the GenBank `nt` database. We then mapped taxonomy for the full dataset by using VSEARCH to search the full sequence file (including singletons and chimeras) against these cluster centroids, also with a similarity threshold of 99%. After taxonomy assignment, we collapsed all clusters assigned as the same phylotype (species, genus, or family depending on the analysis of interest) into single taxon groups using the `phyloseq` 1.16.2 package (McMurdie & Holmes, 2013) in R 3.3.1 (R Core Team, 2016), then corrected for between-sample differences in sequencing depth by transforming raw read counts for each taxon group into sample proportions. Taxa with a mean abundance less than 1% per sample were removed from plots, but included in multivariate analyses.

To visualize the effects of depth and C/N content on root community composition, we performed nonmetric multidimensional scaling followed by permutational MANOVA using the `adonis` function of vegan 2.4-1 (Oksanen *et al.*, 2016) in R 3.31 (R Core Team, 2016), using Jaccard distance as the response variable; depth, C, and N as environmental variables; and 'plot:location' as a blocking effect. To assess species co-occurrence patterns, we used the package `cooccur` 1.3 (Griffith *et al.*, 2016) in R 3.31.

Full analysis scripts and raw sequence data are available online (https://github.com/infotroph/Prairie_seq) and will be archived in Dryad (http://datadryad.org).

Results

The MiSeq run and sequence cleanup produced ample, high-quality sequences for analysis: The raw file contained 1286163 plant ITS2 reads, of which 730235 were successfully end-paired, quality filtered, and assigned back from barcodes to samples. These clean reads contained 494505 unique sequences, of which 459986 were observed only once in the dataset (singletons). Of the 34519 sequences observed at least twice, 2219 were identified as probable or borderline PCR chimeras by VSEARCH. These chimeric sequences accounted for a total of 8526 reads, or *c.* 1.2% of the raw dataset. A further 75 sequences were identified as incomplete or undetectable ITS2 regions by ITSx, leaving a total of 32225 ITS2 sequences to be clustered at 99% similarity into 1347 unique operational taxonomic units (OTUs), which we then used as reference sequences at 99% similarity to map the 730235 reads (singletons included) of the cleaned sequence file, for a final sequence-by-sample table of 576650 reads. Sample coverage was excellent, with 110 of the 120 mixed root samples having more than 1000 reads (1355-8998), sufficient for confident analysis. There was no apparent change in total read count between samples from differing depths.

After obtaining best-match barcode identities for each of the 1347 OTUs clustered at 99%, we collapsed OTUs to phylotypes by combining groups that all yielded the same taxon as their top BLAST hit. When we collapsed at the species level, we obtained 158 phylotypes that were identified as originating from 70 genera across 16 families, which is

notably higher than the 32 species, 22 genera, and 6 families known from site vegetation surveys (Table 1). Many samples contained high read counts from several different 'species' of genera that have only one known species at the site (Fig. S2), possibly indicating that ITS2 barcode identities may be more reliable at the genus rather than the species level.

Of the genera detected both belowground and aboveground, per-plot mean belowground read proportion was positively correlated with percent aboveground cover, but the slope of the relationship was much lower for grasses than for any dicot group (Fig. 2), possibly indicating a lower sensitivity to detect grasses by this ITS method. Similarly, the number of reads obtained from grasses in our mock community was consistently lower than calculated from input DNA concentrations, and neither of two species from the Lamiaceae was ever detected either in the mock community or the mixed samples (Fig. S2).

To test for spatial partitioning between species, we examined how often pairs of species were found together in a given sample. Most species co-occurred with each other at rates broadly compatible with random occurrence. Of the pairs that appeared nonrandom, more were positively correlated (overdispersion) than were negatively correlated (spatial partitioning), and there were no evident differences in co-occurrence rate between samples from different depths (Fig. S4). However, these effects differed between functional types: Species from the Asteraceae showed neutral to positive co-occurrence with species from all families, while Poaceae species had positive co-occurrence with other Poaceae but negative co-occurrence with most species from other families (Fig. 3).

The overall taxonomic makeup of roots shifted with depth: The proportion of reads identified as grasses increased from *c.* 10% of near-surface reads to *c.* one-third of reads in the 75-100 cm layer. Forbs from the Asteraceae declined somewhat with depth, while legumes were relatively consistent across depths (Fig. 4). These partitionings were consistent across groups within each functional type: When grouped by genus, all grasses were more abundant in deep layers than near the surface (Fig. S5). Within the Asteraceae, only *Silphium* was more abundant in middle layers rather than shallow, possibly

attributable to being the only genus at the site with a mix of tap-rooted (*S. lacinatum*, *S. terebinthinaceum*) and fibrous-rooted (*S. integrifolium*, *S. perfoliatum*) species.

Our observation of increasing grass dominance in deeper layers was supported by multivariate analyses: Non-metric multidimensional scaling produced a first axis highly correlated with depth (Fig. 5), with grasses associated with deeper layers, Asteraceae with shallow layers, and legumes orthogonal to the depth axis. The largest single explanatory variable was sampling location, which we treated as a conditioning variable and which explained *c.* 40% of total inertia; after accounting for location, depth added a small (*c.* 8%) but significant (PERMANOVA pseudo-F = 8.7; P = 0.001) further increase in inertia explained. Organic C, N, and C:N were all strongly anticorrelated with depth and neither they nor their interactions with depth added any further explanatory power (PERMANOVA; all $R^2 < 0.01$; all $P > 0.4$).

Discussion

Root placement in our prairie plots appears to be partitioned by taxonomy, with grasses increasing in abundance relative to forbs at greater depths (Fig. 4) but little evidence of spatial patterning among other functional groups (Fig. 3). The aboveground and belowground abundances of most species appeared to scale together (Fig. 2), but with very different slopes for grasses than for dicots. Given the high overall proportion of reads identified as Asteraceae and especially given their dominance in the shallow soil layers that contain over 80% of the total root mass in this system (Fig. 1), it is likely that much of the root mass in these plots originates from forbs rather than grasses. However, even if grass root mass in the deep soil layers is small relative to the surface and relative to other taxa, it may still be ecologically important for access to water, nutrient economy, and carbon storage.

One goal of this study was to infer the species composition of root mixtures, to allow richer insights from other studies where bulk biomass samples are collected without identifying their component mixtures. We found that most genera were at least occasionally present at all depths, that the majority of the sequences obtained were from forbs in the Asteraceae, that these were especially dominant in the shallow soil layers, and that despite grasses dominating aboveground measurements of cover and biomass we

detected few reads from grass roots near the surface and only some in the deeper layers. Considering biomass and taxon identity together, it is thus likely that relatively shallow forb roots account for a substantial fraction of the root C input to this prairie. However, this does not necessarily mean that grasses are a minor component of total root mass: To attempt a quantitative conversion from reads to tissue mass requires knowing the potentially species-specific relationships between root biomass and DNA yield (Rewald *et al.*, 2012; Zeng *et al.*, 2015), which we did not attempt to characterize here. The lower slope of grasses in the aboveground-belowground relationship (Fig. 2) and the apparent lower detection of grasses in our mock community (Fig. S2) suggest a possible extraction or sequencing bias against grasses, but even if present we expect this to be constant across samples and thus should not affect comparisons of relative functional group dominance between depths.

An additional caution for our method is that the precision and accuracy of our barcode identifications are limited by the taxonomic resolution of the ITS2 region, the potential for intraspecific variation (Álvarez & Wendel, 2003), and the completeness of available sequence databases. Following previous workers (e.g. Fahner *et al.*, 2016), we regard all species identities from this dataset to be tentative, and in a few cases even genus assignment was uncertain. This was particularly evident in the C_4 grasses, for many of which we obtained BLAST ties with multiple best results widely spread across the PACMAD clade (sensu Sánchez-Ken & Clark, 2010). However, because our site has only a few congeneric species (Table 1) and our hypotheses were focused at the level of phylogenetically distinct functional groups, further refinements of sequence identity are unlikely to change the conclusions. Using multiple primer sets may help to increase taxonomic resolution and would be straightforward to add to future work using the Fluidigm PCR platform (Brown *et al.*, 2016), but in our mixed-sample setting the bioinformatic challenge of resolving multiple markers would likely be prohibitive (Fahner *et al.*, 2016).

Previous work on the spatial structure of grassland root communities has found mixed results. An old-field site in Ontario (Kesanakurti *et al.*, 2011) was similar to ours in finding positive correlations between aboveground and belowground abundance and in finding fewer roots from grasses than expected from aboveground data, but differed in

finding very strong spatial partitioning. This difference may be attributable to differences in spatial scale (they extracted individual roots from 5x5x5-cm soil blocks, we pooled roots from multiple cores in each *c.* 2 m sampling area) or to differing environmental controls in a heterogeneous old-field environment than in our uniformly planted and actively managed site. By contrast, two grasslands with differing soil water regimes in Wyoming showed co-occurrence patterns much like those we observed, with most species present in all samples and only weak changes in detection with depth (Frank *et al.*, 2015). In addition, mapping of individual roots at the millimeter scale in these same Wyoming grasslands was consistent with our finding of more positive than negative species co-occurrences (Frank *et al.*, 2015). Taken together, these findings support the conclusion of Price and co-workers (2012) that belowground community structure is largely determined by *abiotic* filtering; perhaps we should not be surprised to find little spatial structure in our highly fertile, previously tilled, evenly planted, annually mowed site.

Although root mass in the deep soil was low relative to the surface layers, these roots are likely to be ecologically important. By allocating proportionally more roots to the deep soil than other groups, the grasses in this prairie may gain access to more reliable water supplies and reduced competition for nutrients that help maintain their aboveground dominance. Even small increases in the amount of deep root can contribute to drought survival of grasses (Nippert *et al.*, 2012), reduce N leaching losses (Smith *et al.*, 2013), and increase soil C storage (De Deyn *et al.*, 2008; Anderson-Teixeira *et al.*, 2013), implying that the taxonomic partitioning we observed is likely to increase the management value of prairie restorations as well as their ecological resilience.

Acknowledgements

We thank Xiaohui Feng for plant species identifications, Xiaohui Feng and Michael Dietze for providing the aboveground survey data, Mark Band and Alvaro Hernandez (W. M. Keck Center) for technical advice, library preparation, and sequencing, James Doroghazi, Shawn P. Brown, Elizabeth Bach, and Aleel Grennan for valuable discussions, Michael D. Masters for soil CN analysis and lab logistics, and

386 Robert A. Cachro, Jacob N. Rosenthal, Taylor A. Wright, and Derick A. Carnazzola for
387 assistance with sample processing.
388 Author contributions: CKB, SAW, EHD designed the experiment; CKB, SAW,
389 TLP, CRS collected data; SAW, TLP, CRS developed laboratory protocols; CKB, SAW
390 developed data analysis methods; SAW, EHD provided funding; CKB, SAW, TLP, CRS,
391 EHD wrote the paper.
392 This project was partially funded by the Energy Biosciences Institute and by an
393 award to SAW through the Institute for Genomic Biology Fellows program, Genomic
394 Ecology of Global Change theme.

References

Anderson-Teixeira KJ, Masters MD, Black CK, Zeri M, Hussain MZ, Bernacchi CJ, DeLucia EH. **2013**. Altered belowground carbon cycling following land-use change to perennial bioenergy crops. *Ecosystems* **16**: 508–520.

Álvarez I, Wendel JF. **2003**. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* **29**: 417–434.

Bardgett RD, Mommer L, de Vries FT. **2014**. Going underground: root traits as drivers of ecosystem processes. *TRENDS in Ecology and Evolution* **29**: 692–699.

Bengtsson-Palme J, Ryberg M, Hartmann M, Branco S, Wang Z, Godhe A, De Wit P, Sánchez-García M, Ebersberger I, de Sousa F *et al.* 2013. Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. *Methods in Ecology and Evolution* **4**: 914–919.

Brown SP, Ferrer A, Dalling JW, Heath KD. **2016**. Don't put all your eggs in one basket: a cost-effective and powerful method to optimize primer choice for rRNA environmental community analyses using the Fluidigm Access Array. *Molecular Ecology Resources* **16**: 946–956.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. **2009**. BLAST+: architecture and applications. *BMC Bioinformatics* **10**: 421.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, González Peña A, Goodrich JK, Gordon JI *et al.* 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**: 335–336.

Chen S, Yao H, Han J, Liu C, Song J, Shi L, Zhu Y, Ma X, Gao T, Pang X *et al.* 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE* **5**: e8613.

Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, Brown CT, Porras-Alfaro A, Kuske CR, Tiedje JM. **2013**. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Research* **42**: D633–D642.

De Deyn GB, Cornelissen JHC, Bardgett RD. **2008**. Plant functional traits and soil carbon sequestration in contrasting biomes. *Ecology Letters* **11**: 516–531.

de Kroon H, Hendriks M, van Ruijven J, Ravenek J, Padilla FM, Jongejans E, Visser EJW, Mommer L. **2012**. Root responses to nutrients and soil biota: drivers of species coexistence and ecosystem productivity. *Journal of Ecology* **100**: 6–15.

Fahner NA, Shokralla S, Baird DJ, Hajibabaei M. **2016**. Large-scale monitoring of plants through environmental DNA metabarcoding of soil: Recovery, resolution, and annotation of four DNA markers. *PLoS ONE* **11**: e0157505.

Feng X. **2014**. *Productivity, physiology, community dynamics, and ecological impacts of a grassland agro-ecosystem: integrating field studies and ecosystem modeling*. PhD thesis, University of Illinois at Urbana-Champaign, Urbana, IL, USA.

Feng X, Dietze M. **2013**. Scale dependence in the effects of leaf ecophysiological traits on photosynthesis: Bayesian parameterization of photosynthesis models. *New Phytologist* **200**: 1132–1144.

Fornara DA, Tilman D. **2008**. Plant functional composition influences rates of soil carbon and nitrogen accumulation. *Journal of Ecology* **96**: 314–322.

Fornara DA, Tilman D. **2009**. Ecological mechanisms associated with the positive diversity-productivity relationship in an N-limited grassland. *Ecology* **90**: 408–418.

Frank DA, Pontes AW, Maine EM, Fridley JD. **2015**. Fine-scale belowground species associations in temperate grassland. *Molecular Ecology* **24**: 3206–3216.

Griffith DM, Veech JA, Marsh CJ. **2016**. cooccur: Probabilistic species co-occurrence analysis in R. *Journal of Statistical Software, Code Snippets* **69**: 1–17.

Hendriks M, Ravenek JM, Smit-Tiekstra AE, van der Paauw JW, de Caluwe H, van der Putten WH, de Kroon H, Mommer L. **2015**. Spatial heterogeneity of plant–soil feedback affects root interactions and interspecific competition. *New Phytologist* **207**: 830–840.

Hiiesalu I, Öpik M, Metsis M, Lilje L, Davison J, Vasar M, Moora M, Zobel M, Wilson SD, Pärtel M. **2012**. Plant species richness belowground: higher richness

and new patterns revealed by next-generation sequencing. *Molecular Ecology* **21**: 2004–2016.

Jackson RB, Canadell JG, Ehleringer JR, Mooney HA, Sala OE, Schulze ED. **1996**. A global analysis of root distributions for terrestrial biomes. *Oecologia* **108**: 389–411.

Jackson RB, Moore LA, Hoffmann WA, Pockman WT, Linder CR. **1999**. Ecosystem rooting depth determined with caves and DNA. *Proceedings of the National Academy of Sciences of the United States of America* **96**: 11387–11392.

Jumpponen A, Högberg P, Huss-Danell K, Mulder CPH. **2002**. Interspecific and spatial differences in nitrogen uptake in monocultures and two-species mixtures in north European grasslands. *Functional Ecology* **16**: 454–461.

Kesanakurti PR, Fazekas AJ, Burgess KS, Percy DM, Newmaster SG, Graham SW, Barrett SCH, Hajibabaei M, Husband BC. **2011**. Spatial patterns of plant diversity below-ground as revealed by DNA barcoding. *Molecular Ecology* **20**: 1289–1302.

Martin M. **2011**. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**: 10–12.

Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. **2012**. PANDAseq: PAired-eND Assembler for Illumina sequences. *BMC Bioinformatics* **13**: 31.

Masters MD, Black CK, Kantola IB, Woli KP, Voigt T, David MB, DeLucia EH. **2016**. Soil nutrient removal by four potential bioenergy crops: *Zea mays*, *Panicum virgatum*, *Miscanthus* × *giganteus*, and prairie. *Agriculture, Ecosystems & Environment* **216**: 51–60.

McMurdie PJ, Holmes S. **2013**. phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**: e61217.

Nippert J, Wieme R, Ocheltree T, Craine JM. **2012**. Root characteristics of C_4 grasses limit reliance on deep soil water in tallgrass prairie. *Plant and Soil* **355**: 385–394.

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O'Hara RB, Simpson GL, Solymos P *et al.* 2016. *vegan: Community*

485 *ecology package*. R package version 2.4-1. [WWW document] URL
486 https://CRAN.R-project.org/package=vegan [accessed 6 October 2016].
487 **Price JN, Hiiesalu I, Gerhold P, Pärtel M**. **2012**. Small-scale grassland assembly
488 patterns differ above and below the soil surface. *Ecology* **93**: 1290–1296.
489 **R Core Team**. **2016**. *R: A language and environment for statistical computing*. Version
490 3.3.1. Vienna, Austria: R Foundation for Statistical Computing. [WWW
491 document] URL https://www.R-project.org/ [accessed 6 October 2016].
492 **Rewald B, Meinen C, Trockenbrodt M, Ephrath JE, Rachmilevitch S**. **2012**. Root
493 taxa identification in plant mixtures – current techniques and future challenges.
494 *Plant and Soil* **359**: 165–182.
495 **Rognes T, Flouri T, Nichols B, Quince C, Mahé F**. **2016**. VSEARCH: a versatile open
496 source tool for metagenomics. *PeerJ* **4**: e2584.
497 **Sánchez-Ken JG, Clark LG**. **2010**. Phylogeny and a new tribal classification of the
498 Panicoideae s.l. (Poaceae) based on plastid and nuclear sequence data and
499 structural data. *American Journal of Botany* **97**: 1732–1748.
500 **Smith CM, David MB, Mitchell CA, Masters MD, Anderson-Teixeira KJ, Bernacchi**
501 **CJ, DeLucia EH**. **2013**. Reduced nitrogen losses after conversion of row crop
502 agriculture to perennial biofuel crops. *Journal of Environmental Quality* **42**: 219–
503 228.
504 **Staats M, Arulandhu AJ, Gravendeel B, Holst-Jensen A, Scholtens I, Peelen T,**
505 **Prins TW, Kok E**. **2016**. Advances in DNA metabarcoding for food and wildlife
506 forensic species identification. *Analytical and Bioanalytical Chemistry* **408**:
507 4615–4630.
508 **Tilman D, Reich PB, Knops J, Wedin D, Mielke T, Lehman C**. **2001**. Diversity and
509 productivity in a long-term grassland experiment. *Science* **294**: 843–845.
510 **Tucker SS, Craine JM, Nippert JB**. **2011**. Physiological drought tolerance and the
511 structuring of tallgrass prairie assemblages. *Ecosphere* **2**: 48.
512 **Wardle DA, Peltzer DA**. **2003**. Interspecific interactions and biomass allocation among
513 grassland plant species. *Oikos* **100**: 497–506.
514 **Weaver JE**. **1919**. *The Ecological Relations of Roots*. Washington, D.C.: Carnegie
515 Institution of Washington.

Weaver JE, Voigt JW. **1950**. Monolith method of root-sampling in studies on succession and degeneration. *Botanical Gazette* **111**: 286–299.

Wilson SD. **2014**. Below-ground opportunities in vegetation science. *Journal of Vegetation Science* **25**: 1117–1125.

Zeng W, Zhou B, Lei P, Zeng Y, Liu Y, Liu C, Xiang W. **2015**. A molecular method to identify species of fine roots and to predict the proportion of a species in mixed samples in subtropical forests. *Frontiers in Plant Science* **6**: 313.

Zeri M, Anderson-Teixeira KJ, Hickman GC, Masters MD, DeLucia EH, Bernacchi CJ. **2011**. Carbon exchange by establishing biofuel crops in Central Illinois. *Agriculture, Ecosystems & Environment* **144**: 319–329.

Table 1: Plant species planted or found present (*) during aboveground vegetation surveys of permanent quadrats in prairie restoration plots at the University of Illinois Energy Farm. Mean and standard deviation of stem abundance and percent cover across growing season 2012 were recalculated from X. Feng (unpublished data).

Functional type	Family	Species	% cover	sd
C_3 grass	Cyperaceae	*Carex bicknellii*	1.4	2.3
C_3 grass	Poaceae	*Elymus canadensis*	2.2	3.4
C_4 grass	Poaceae	*Andropogon gerardii*	34.3	13.7
C_4 grass	Poaceae	*Schizachyrium scoparium*	3.0	3.4
C_4 grass	Poaceae	*Sorghastrum nutans*	21.2	10.5
Forb	Asteraceae	*Coreopsis palmata*	0.1	0.5
Forb	Asteraceae	*Coreopsis tripteris*	4.9	6.1
Forb	Asteraceae	*Echinacea pallida*	0.1	0.4
Forb	Asteraceae	*Felicia hirta*[*]	0.3	0.8
Forb	Asteraceae	*Helianthus grosseserratus*	7.2	8.2
Forb	Asteraceae	*Heliopsis helianthoides*	0.1	0.3
Forb	Asteraceae	*Parthenium integrifolium*	0.4	1.2
Forb	Asteraceae	*Ratibida pinnata*	8.3	5.4
Forb	Asteraceae	*Rudbeckia hirta*[*]	0.0	0.2
Forb	Asteraceae	*Rudbeckia subtomentosa*	2.9	4.2
Forb	Asteraceae	*Silphium integrifolium*	1.6	2.8
Forb	Asteraceae	*Silphium laciniatum*	0.2	0.8
Forb	Asteraceae	*Silphium perfoliatum*	0.5	2.0
Forb	Asteraceae	*Silphium terebinthinaceum*	0.1	0.7
Forb	Asteraceae	*Solidago canadensis*[*]	3.2	6.4
Forb	Asteraceae	*Solidago rigida*	1.3	2.7
Forb	Asteraceae	*Symphyotrichum novae-angliae*	4.0	3.8
Forb	Asteraceae	*Taraxacum officinale*[*]	0.1	0.5
Forb	Lamiaceae	*Monarda fistulosa*	1.0	2.0
Forb	Lamiaceae	*Pycnanthemum virginianum*	0.6	0.8
Forb	Plantaginaceae	*Penstemon digitalis*	0.5	2.2
Forb	Plantaginaceae	*Veronicastrum virginicum*	0.2	0.8
N fixer	Fabaceae	*Astragalus canadensis*	0.0	0.2
N fixer	Fabaceae	unidentified *Baptisia* sp.[*]	0.1	0.4
N fixer	Fabaceae	*Baptisia alba*	0.7	1.5
N fixer	Fabaceae	*Dalea purpurea*	0.2	0.6
N fixer	Fabaceae	*Desmodium canadense*	4.8	5.7
N fixer	Fabaceae	*Lespedeza capitata*	0.2	0.5

[*]Indicates a species not present in the seed mix planted during plot establishment in 2008.

Figure legends

Fig. 1: (a) total root mass of prairie plots measured in mid-August of 2011 (replotted from Anderson-Teixeira *et al.*, 2013) and 2014 (replotted from Black *et al*., submitted). Error bars show mean ± 1 standard deviation of 24 cores. Remaining panels show means of soil properties measured when the plots were established in 2008 (replotted from Smith *et al.*, 2013): (b) soil texture; (c) soil organic C and N content; (d) soil bulk density.

Fig. 2: Correlation between percent aboveground cover (horizontal axis) and percent of reads per root sample (vertical axis). Each point shows mean ± 1 standard error for one genus in one experimental plot (N = 5 plots). Belowground read percentages are averaged across all depths. Separate regression fits are shown for dicot and monocot genera; shaded areas show 95% confidence bands for the line of fit.

Fig. 3: Standardized effect sizes for observed co-occurrence rates. Each point is one pair of species; violin shapes show density distribution for each group. Vertical axes show family identity of the first species in the pair, panel labels show the family identity of the second species. Co-occurrence was defined as both species providing more than 1% of internal transcribed spacer (ITS2) reads from the same mixed root sample. Larger positive effect sizes indicate species pairs that are found together in more samples than expected for independently occurring species; negative values indicate pairs that are found together less than expected.

Fig. 4: Relative abundance (fraction of reads from each sample) as a function of sample depth for each observed plant family. Sequences were clustered at 99% similarity and identified to species according to the closest BLAST match against internal transcribed spacer (ITS2) sequences in the Genbank `nt` database. Taxa were then collapsed by family and groups with a mean abundance less than 1% per sample were removed for plotting.

Fig. 5: Nonmetric multidimensional scaling plot showing centroids for all detected species. Each text label indicates the centroid for one species, identified by short species code and colored by family. Each cross indicates the centroid for a low-abundance species, colored by family but left unlabeled for figure clarity. Black arrows show best fits for environmental vectors of depth in soil, organic C and N content, and C:N ratio.

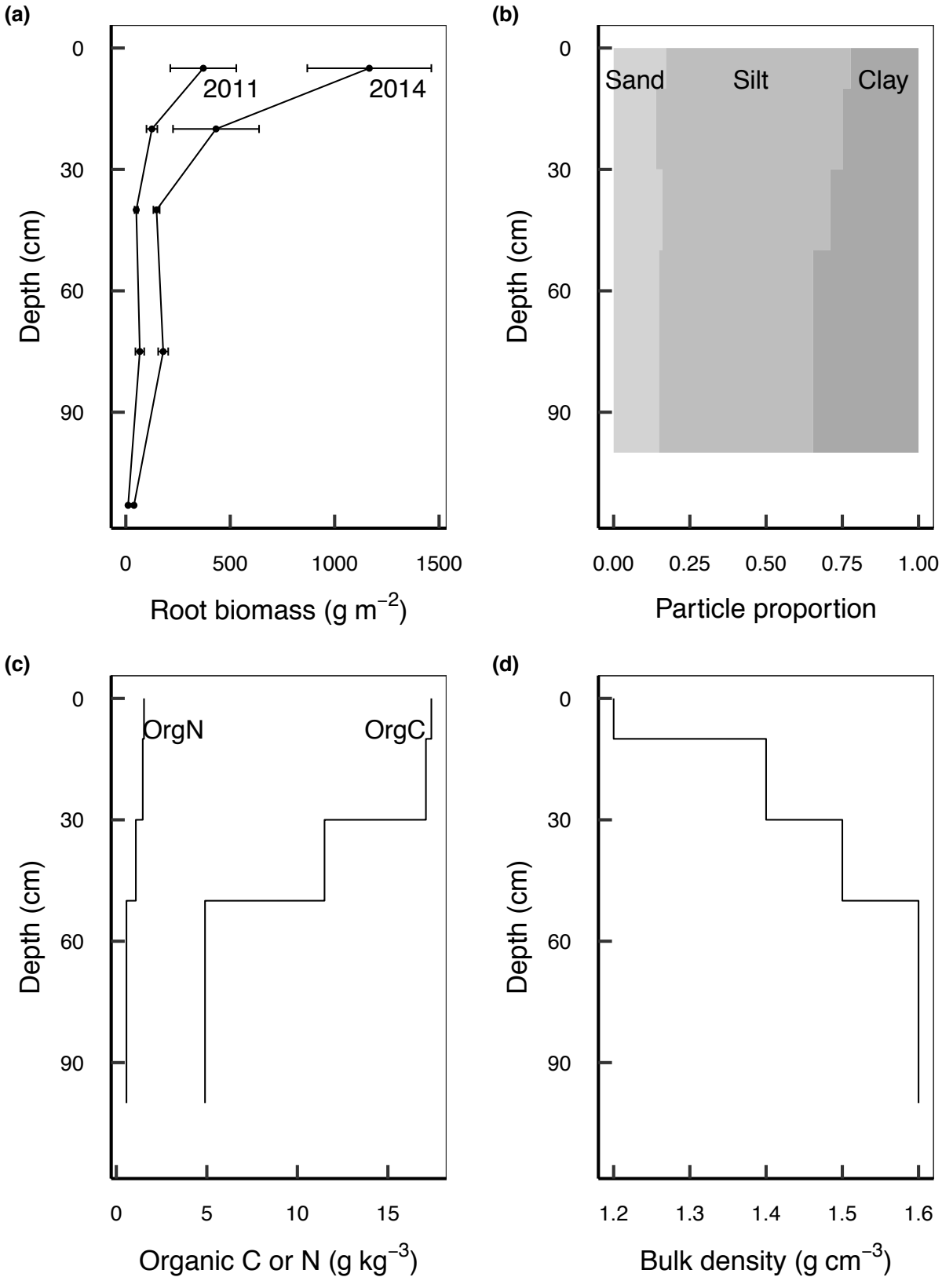
(a)
0
30
60
90
Depth (cm)
2011
2014
0
500
1000
1500
Root biomass (g m^{-2})
(b)
Sand
Silt
Clay
0.00
0.25
0.50
0.75
1.00
Particle proportion
(c)
OrgN
OrgC
0
5
10
15
Organic C or N (g kg^{-3})
(d)
1.2
1.3
1.4
1.5
1.6
Bulk density (g cm^{-3})

Fig. 1.

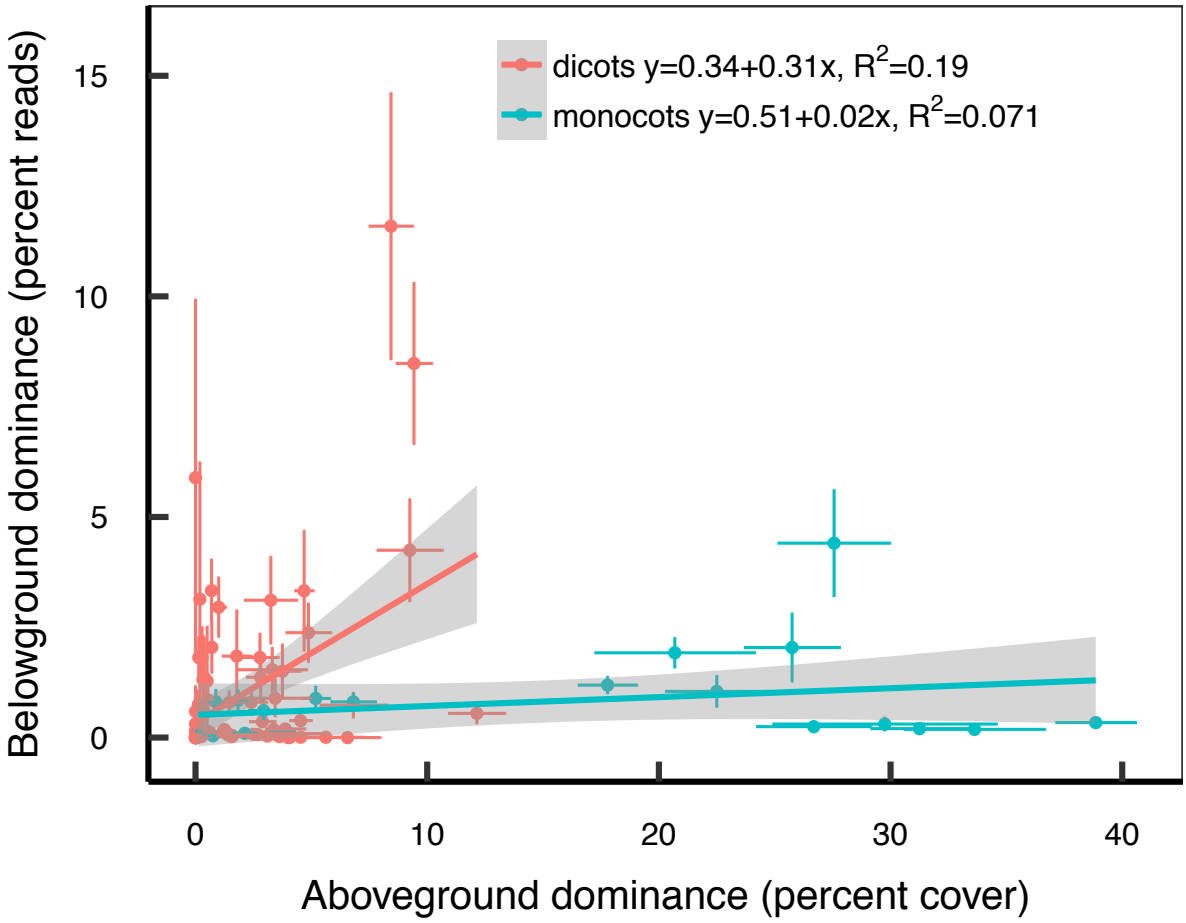
dicots y=0.34+0.31x, R2=0.19
monocots y=0.51+0.02x, R2=0.071
Belowground dominance (percent reads)
Aboveground dominance (percent cover)
0
5
10
15
0
10
20
30
40

Fig. 2.

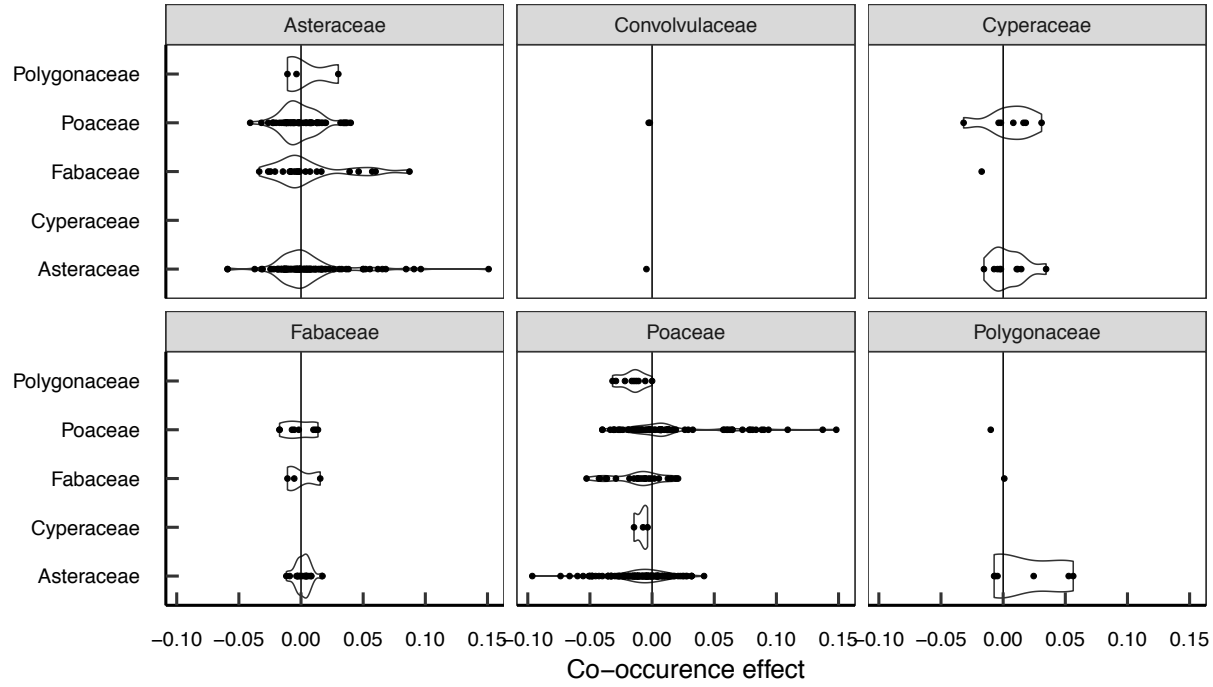
Asteraceae
Convolvulaceae
Cyperaceae
Fabaceae
Poaceae
Polygonaceae
Polygonaceae
Poaceae
Fabaceae
Cyperaceae
Asteraceae
−0.10 −0.05 0.00 0.05 0.10 0.15
Co−occurence effect

Fig. 3.

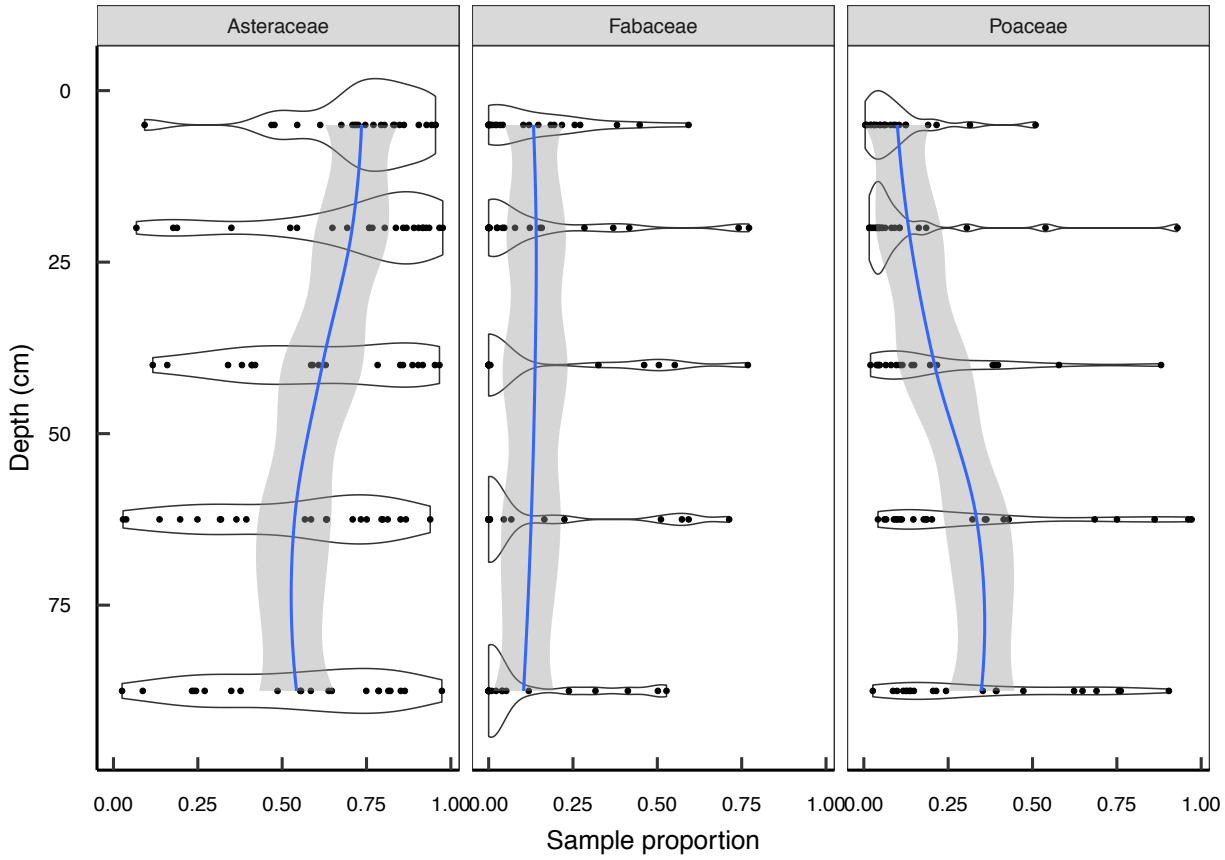
Asteraceae
Fabaceae
Poaceae
Depth (cm)
0
25
50
75
0.00
0.25
0.50
0.75
1.00
Sample proportion

Fig. 4.

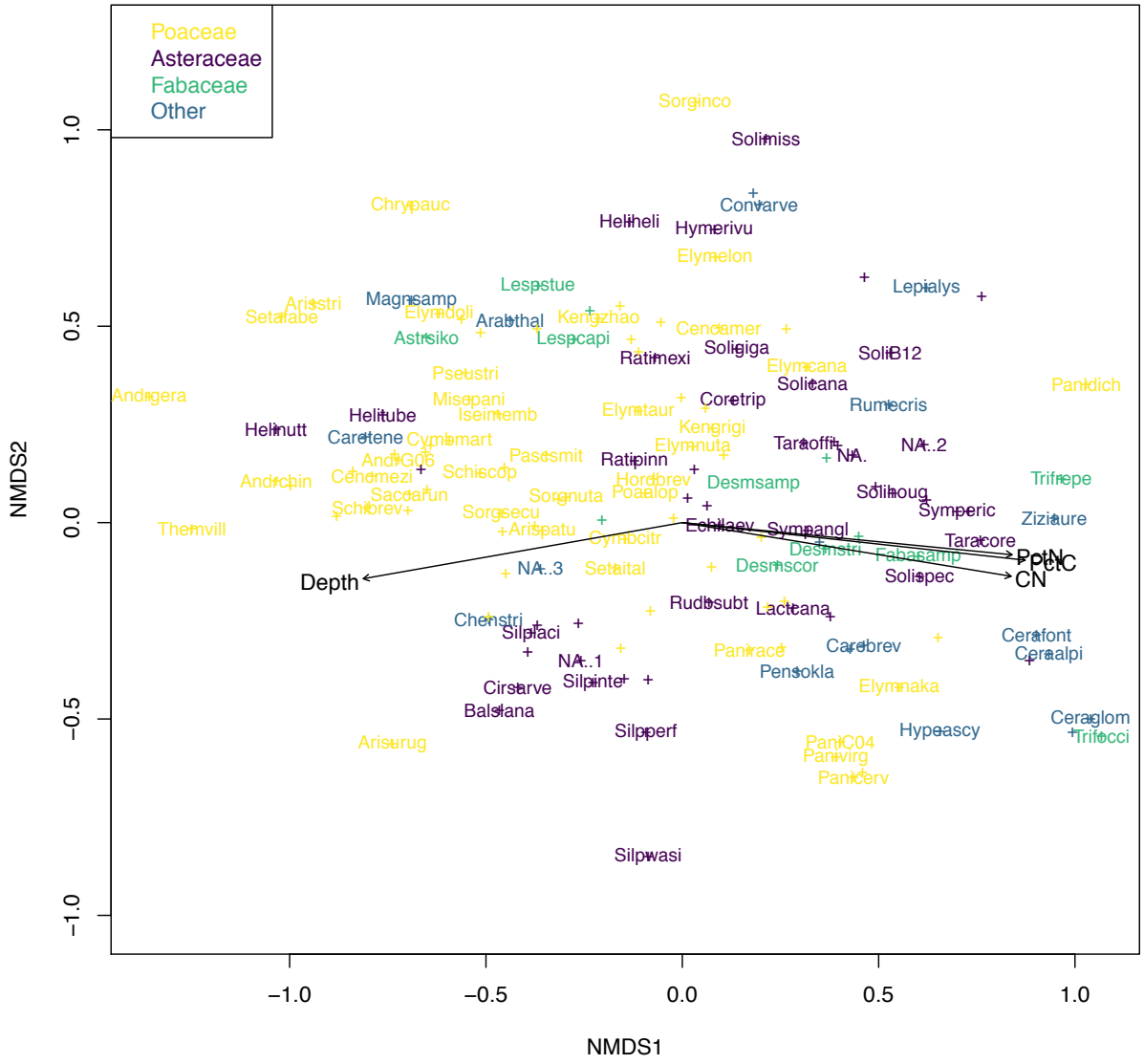
Poaceae
Asteraceae
Fabaceae
Other
NMDS1
NMDS2
−1.0
−0.5
0.0
0.5
1.0
Depth
PctN
PctC
CN
Sorginco
Solimiss
Chrypauc
Convarve
Heliheli
Hymerivu
Elymelon
Lepialys
Lespstue
Magnsamp
Arisstri
Setafabe
Elymcali
Arabthal
Kengzhao
Cencamer
Astrsiko
Lespcapi
Soligiga
SoliB12
Ratimexi
Elymcana
Pseustri
Solicana
Andrgera
Miscpani
Coretrip
Rumecris
Panidich
Helinutt
Helitube
Iseimemb
Elymtaur
Kengrigi
Carrtene
Cymbmart
Elymnuta
Taraoffi
NA..2
AndrG06
Pascsmit
Ratipinn
NA.
Andrchin
Cenomezi
Schiscop
Hordbrev
Desmsamp
Trifrepe
Saccarun
Sorgnuta
Poaalop
Solihoug
Schibrev
Sorgsecu
Symperic
Zizaaure
Themvill
Arispatu
Echilaev
Sympangl
Cymbcitr
Taracore
Desmstri
Fabasamp
NA..3
Setaital
Desmscor
Solispec
Rudbsubt
Lactcana
Chenstri
Silpraci
Ceratont
Panivirg
Carebrev
Ceraalpi
NA..1
Pensokla
Cirsarve
Silpinte
Elymnaka
Balsfana
Ceraglom
Silpperf
Hypoascy
Arisurug
PaniC04
Trifocci
Panivirg
Panicerv
Silpwasi

Fig. 5.

Supporting Information

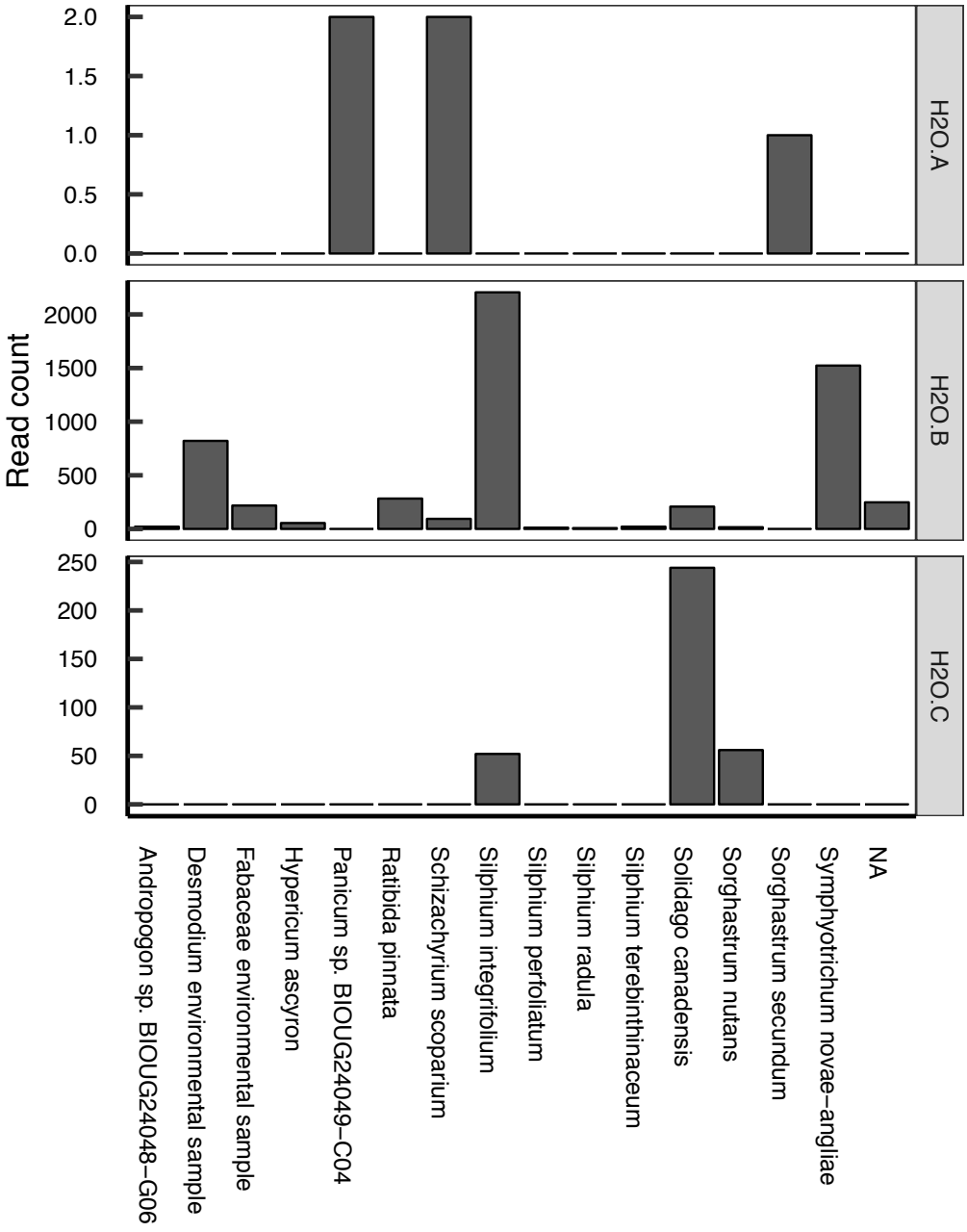

Fig. S1: Read counts of internal transcribed spacer (ITS2) sequences obtained from water controls, binned by assigned genus. Each panel shows reads from a separate aliquot of DNA-free water processed simultaneously through the same DNA extraction, PCR, and sequencing pipeline as the root samples. Notice that the y-axis shows raw read counts (not sample proportions as in other figures) and that the scale differs between panels.

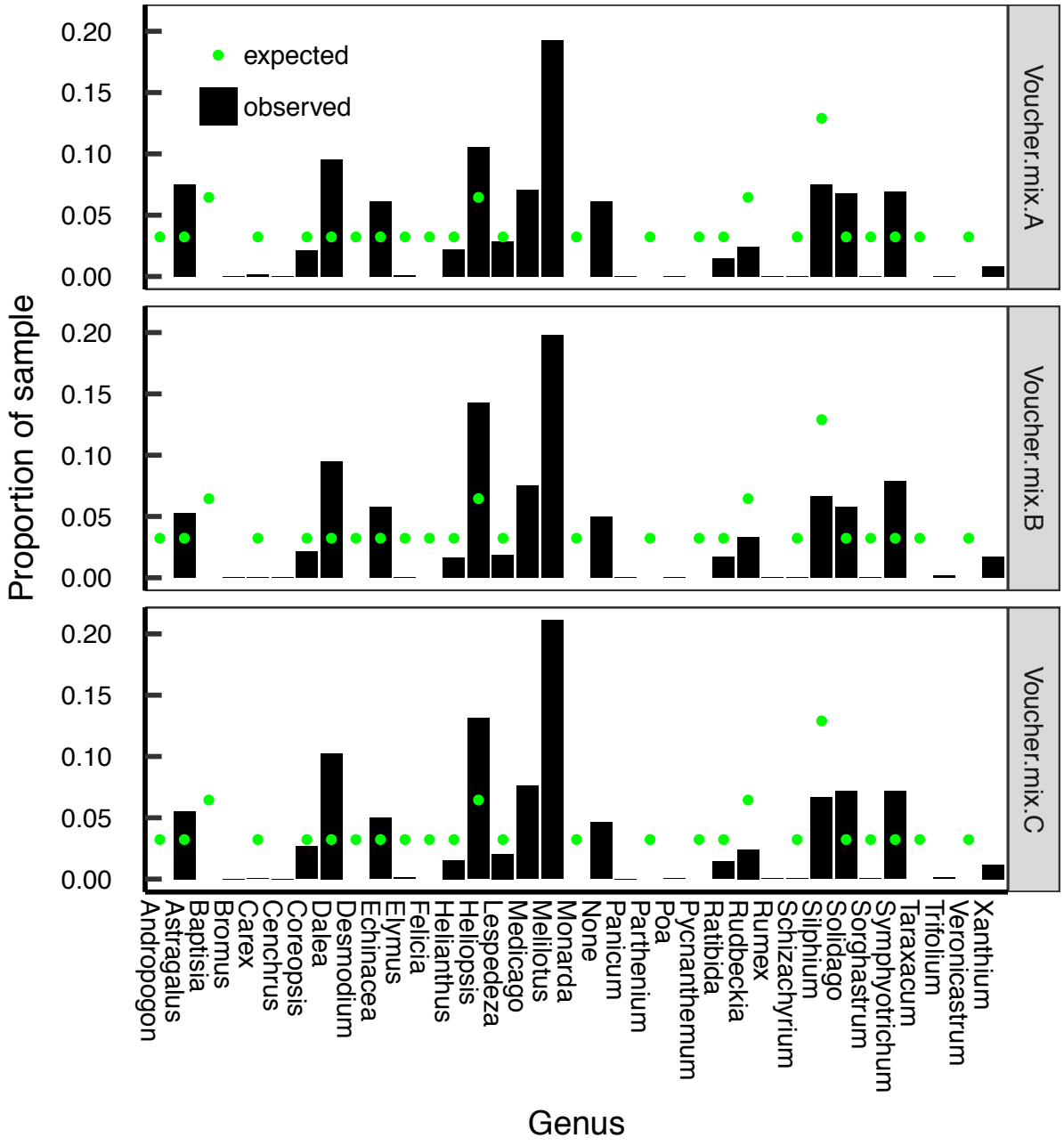

Fig. S2: Genus identities of internal transcribed spacer (ITS2) reads obtained from a mock community of DNA obtained from known-species root samples. Green dots show proportion of template DNA added to the mixture. Black bars show the proportion of reads obtained. Panels are technical replicates, each separately amplified and sequenced from the same aliquot of mixed DNA.

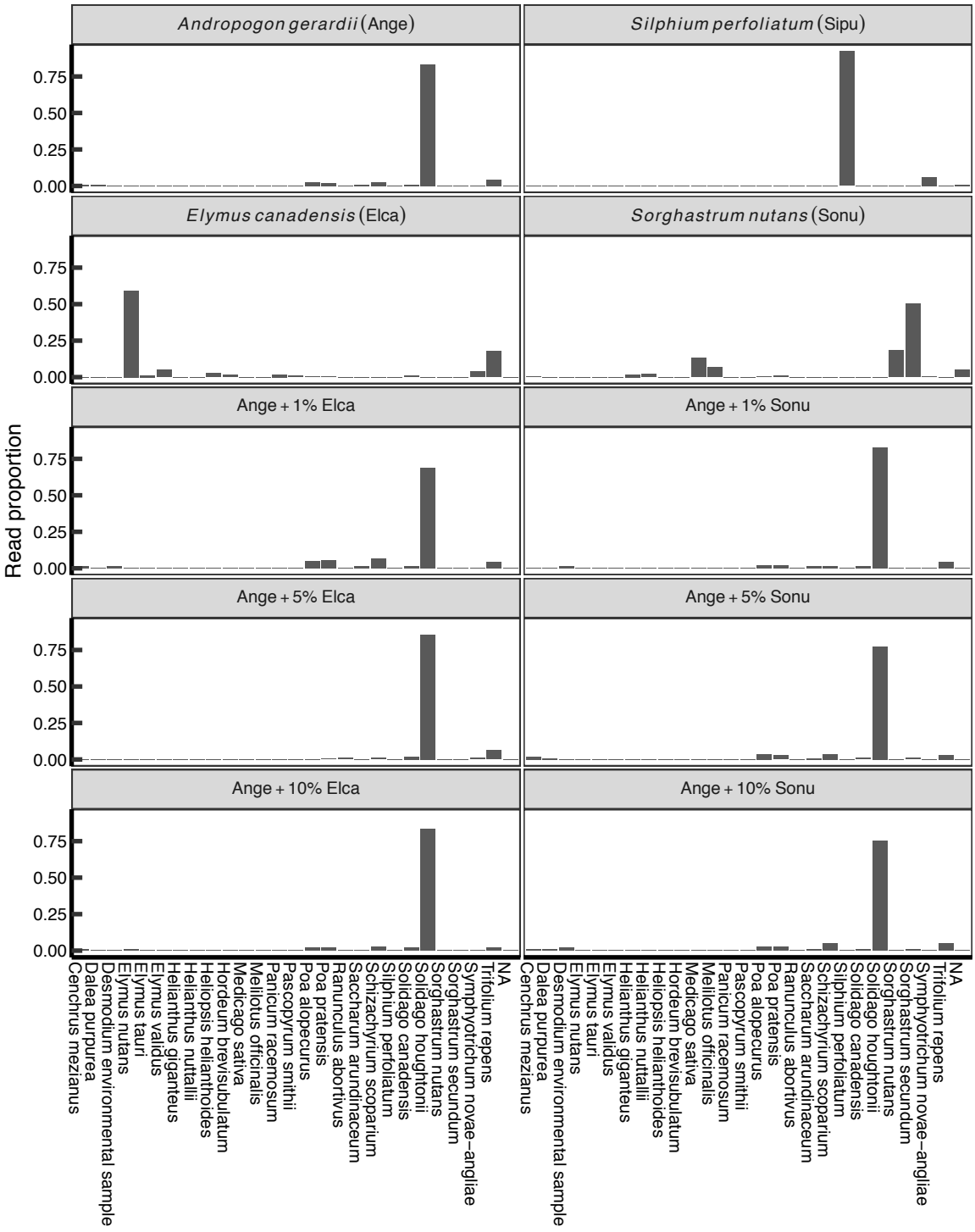

Fig. S3: Species identities of internal transcribed spacer (ITS2) reads obtained from (top row) DNA extracted from single-species root samples, and (second two rows) DNA from the '*Andropogon gerardii*' sample spiked with 1%, 5%, or 10% *Elymus canadensis* or *Sorghastrum nutans* DNA.

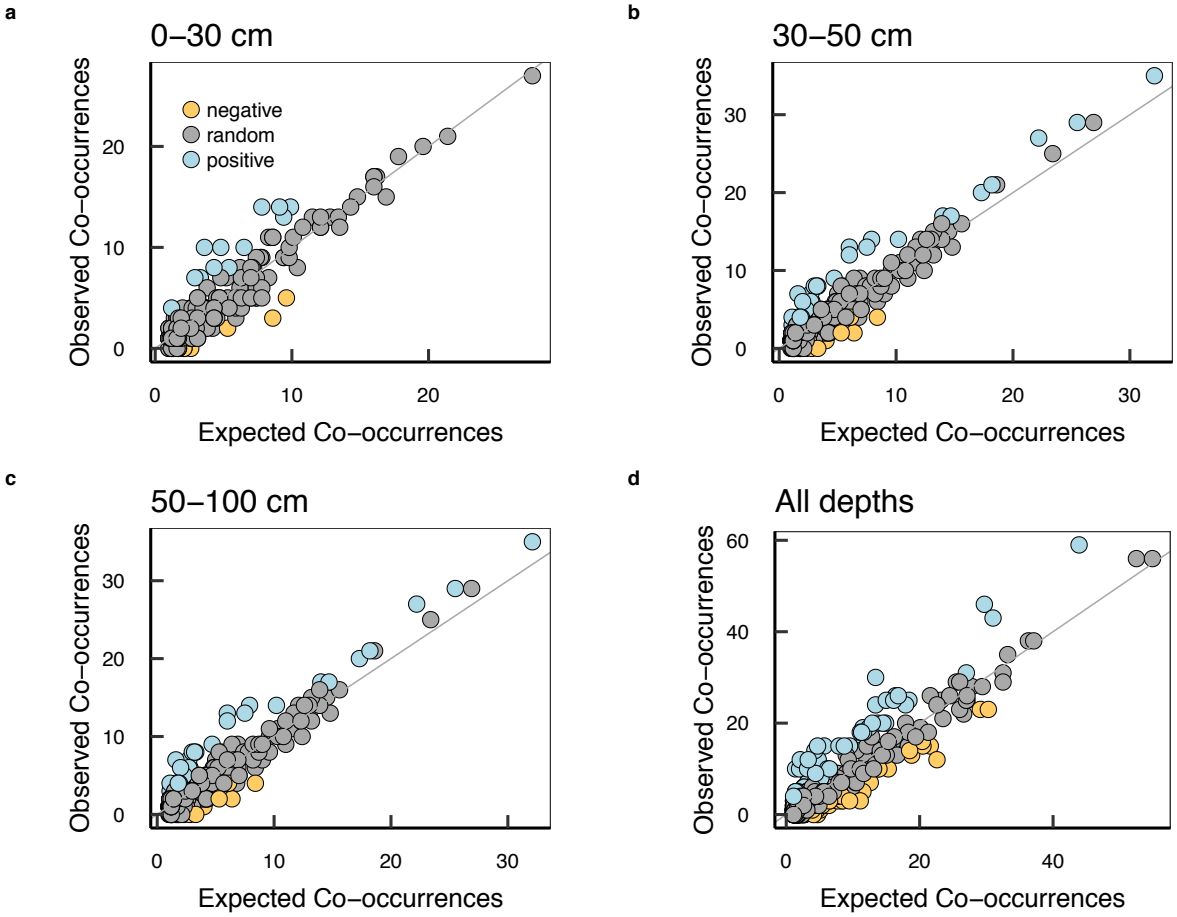

Fig. S4: Observed species co-occurrence rates versus the rates expected for independent, randomly distributed species of the same abundance, calculated separately for (a-c) shallow, medium, and deep soil layers, or (d) all samples from all depths. Each point represents one pair of species and the observed co-occurrence rate is the number of root samples with at least 1% of reads attributed to each species in the pair. Colors indicate species pairs that co-occur significantly more (blue) or less (yellow) often than expected under the null model ($P < 0.05$).

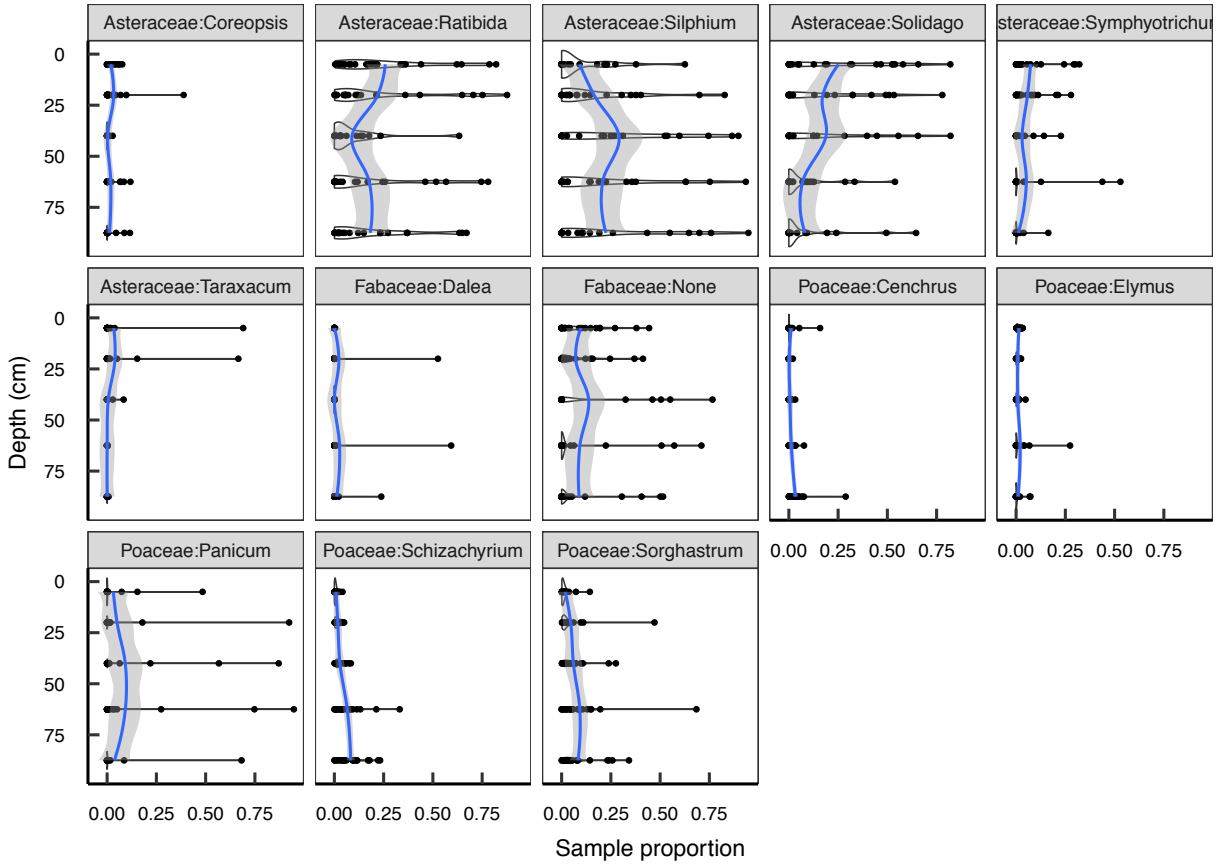

Fig. S5: Relative abundance (fraction of reads from each sample) as a function of sample depth for each observed plant genus. Sequences were clustered at 99% similarity and identified to species according to the closest BLAST match against internal transcribed spacer (ITS2) sequences in the Genbank `nt` database. Taxa were then collapsed by genus and groups with a mean abundance less than 1% per sample were removed for plotting.