# Taxonomic partitioning of root resources by depth in a prairie plant community, or: Watch those grasses, they may look like lightweights but they’ll go for your aquifer

#### …Sorry, just had to get that out of my system. The next draft will have a real title.

#### Christopher K. Black1,2, Scott A. Woolbright2,3, Taylor L. Pederson2, Christopher R. Sligar2, Evan H. DeLucia1,2

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

## 

## Running Head

Prairie taxonomic 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

## 

## Author contributions:

CKB, SAW, EHD designed the experiment; CKB, SAW, TLP, CRS collected data; SAW, TLP, CRS developed laboratory protocols; CKB, SAW developed data analysis methods; SAW, EHD provided funding; CKB, SAW, TLP, CRS, EHD wrote the paper.

## Methods

### Experimental site

Our experimental site is the University of Illinois Energy Farm (Urbana, Illinois, USA: 40.06N, 88.19W, elevation 220 m), a bioenergy research facility founded to compare the ecological and economic impacts of perennial grasses against those of conventional row crops when both are grown as feedstocks for fuel production. Four cropping systems are grown side-by-side; For this experiment we were interested in multispecies root communities and therefore sampled only from the five replicates (four 0.7 Ha plots one 3.8 Ha plot) of the prairie treatment. The site has a continental climate with a mean annual temperature of 11°C and approximately 1 m of precipitation annually. It is established on deep, highly fertile Mollisol soils (Argiudolls, mapped as Dana, Flanagan and Blackberry silt loam) and was used for agriculture for at least 100 years before establishment of the current experiment. For futher details on the establishment and management of the site, see previous work by (Zeri et al. 2011, Smith et al. 2013, Masters et al. 2016).

The prairie plots were established in 2008 by seeding with a mix of 28 species native to Illinois (**Table 1 TK: similar to Feng and Dietze 2013, but with abundance**). Species were chosen to balance four stated goals: Maximum biomass production, ecosystem resilience, trophic diversity, and locally adapted species. Seeds were obtained from Pizzo and Associates (Leland IL), treated at planting time with a mycorrhizal inoculum (“AM 120”; Reforestation Technologies International; Gilroy CA) that is primarily *Glomus intraradices* (Neal Anderson, RTI Inc.; personal communication), and overseeded with a spring oat (*Avena sativa*) cover crop. The plots were mowed after senescence each year and the above-ground biomass was baled and removed.25 of these 28 species remained detectable in 2011 (Feng and Dietze 2013) and in 2013 a total of 32 species were censused in the plots (**???**).

### Sample collection

To characterize the spatial distribution of species with depth, we collected mixed 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 bucket auger to collect roots and soil from five depth increments (0-10, 10-30, 30-50, 50-75, and 75-100 cm) of 3 cores within a 2 m radius. We pooled all three cores from each location, collected a ~0.5-1 kg subsample 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 1-gallon Ziplok 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 33 plant species present in aboveground surveys(X. Feng, unpublished data), we located 3-5 individuals, identified them to species by leaf and flower morphology, and used a trowel to extract roots still attached to their well-identified crown. We pooled all roots from each species, placed them in Ziplok 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 about 30-90 minutes 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 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 minutes of sonication to dislodge any residual soil from the 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 ~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. We then performed a post-extraction cleanup using materials from the same kit (E. Adams, Mo Bio; personal communication): 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 xg for 2 minutes. 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.

[TKTK This paragraph needs many details! Scott, can you help? Also maybe we don’t need to discuss the non-plant primers…] 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 roots and rhizosphere soil, we included nominally-pure extracts of root DNA from four species with high aboveground abundance (*Andropogon gerardii*, *Sorghastrum nutans*, *Silphium perfoliatum*, and *Elymus nutans*; Feng 2014), water extractions as a negative control on the DNA extraction + PCR + sequencing process taken as a whole, and a mock community of DNA from 29 species combined in equimolar quantity plus one species (*Heliopsis helianthoides*) at twice the concentration of the others. Each sample was amplified by microfluidic PCR using the Fluidigm Access Array chip to create amplicons from five primer sets targeting different regions of the ribosomal RNA genes of diverse phylogenetic groups: Bacterial 16S V4 [cite TK], fungal ITS [cite TK], SSU of Glomeromycota (arbuscular mycorrhizal fungi) (Van Geel et al. 2014), eukaryotic 18S [cite TK], and plant ITS2 (Chen et al. 2010)). The resulting amplicons were then barcoded with a unique 10-base oligonucleotide for each sample [TK: Do we need to present full linker constructs?] and sequenced by synthesis for 2x301 paired-end cycles (MiSeq, V3 chemistry, Illumina Inc, San Diego CA). For the remainder of this paper, we discuss only the results obtained from root samples using the plant ITS2 primers S2F (5’-ATGCGATACTTGGTGTGAAT) and S3R (5’-GACGCTTCTCCAGACTACAAT) (Chen et al. 2010).

### Data processing

The raw Illumina read files were separated into one file from each primer set and PhiX reference reads were removed using CASAVA 1.8. 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 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 ~10 bases; see Cole et al. 2013). We then used the split\_libraries.py script in QIIME 1.9.1 (Caporaso et al. 2010) to assign barcodes to sequence identities.

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 and 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 mutivariate analyses.

To visualize the effects of depth and C/N content on root community composition, we performed nonmetric multidimensional scaling followed by permutational MANOVA (adonis) as implemented by the vegan 2.4-1 package (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; ‘plot:location’ as a blocking effect.

Full analysis scripts and raw sequence data are available online at (Dryad URL TK).

## References

Bardgett, R. D., L. Mommer, and F. T. De Vries. 2014. Going underground: root traits as drivers of ecosystem processes. TRENDS in Ecology and Evolution 29:692–699.

Bengtsson Palme, J., M. Ryberg, M. Hartmann, S. Branco, Z. Wang, A. Godhe, P. Wit, M. Sánchez García, I. Ebersberger, F. Sousa, A. Amend, A. Jumpponen, M. Unterseher, E. Kristiansson, K. Abarenkov, Y. J. K. Bertrand, K. Sanli, K. M. Eriksson, U. Vik, V. Veldre, and R. H. Nilsson. 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.

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

Caporaso, J. G., J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Peña, J. K. Goodrich, J. I. Gordon, G. A. Huttley, S. T. Kelley, D. Knights, J. E. Koenig, R. E. Ley, C. A. Lozupone, D. McDonald, B. D. Muegge, M. Pirrung, J. Reeder, J. R. Sevinsky, P. J. Turnbaugh, W. A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, and R. Knight. 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7:335–336.

Chen, S., H. Yao, J. Han, C. Liu, J. Song, L. Shi, Y. Zhu, X. Ma, T. Gao, X. Pang, K. Luo, Y. Li, X. Li, X. Jia, Y. Lin, and C. Leon. 2010. Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species. Plos One 5:e8613.

Cole, J. R., Q. Wang, J. A. Fish, B. Chai, D. M. McGarrell, Y. Sun, C. T. Brown, A. Porras-Alfaro, C. R. Kuske, and J. M. Tiedje. 2013. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Research 42:D633–D642.

Fahner, N. A., S. Shokralla, D. J. Baird, and M. Hajibabaei. 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; University of Illinois At Urbana-Champaign.

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

Hendriks, M., J. M. Ravenek, A. E. Smit Tiekstra, J. W. Paauw, H. Caluwe, W. H. Putten, H. Kroon, and L. Mommer. 2015. Spatial heterogeneity of plant–soil feedback affects root interactions and interspecific competition. New Phytologist 207:830–840.

Hiiesalu, I., M. Öpik, M. Metsis, L. LILJE, J. Davison, M. Vasar, M. Moora, M. Zobel, S. D. Wilson, and M. Pärtel. 2012. Plant species richness belowground: higher richness and new patterns revealed by next‐generation sequencing. Molecular Ecology 21:2004–2016.

Jackson, R. B., J. G. Canadell, J. R. Ehleringer, H. A. Mooney, O. E. Sala, and E. D. Schulze. 1996. A global analysis of root distributions for terrestrial biomes. Oecologia 108:389–411.

Jackson, R. B., L. A. Moore, W. A. Hoffmann, W. T. Pockman, and C. R. Linder. 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.

Kesanakurti, P. R., A. J. Fazekas, K. S. Burgess, D. M. Percy, S. G. Newmaster, S. W. Graham, S. C. H. Barrett, M. Hajibabaei, and B. C. Husband. 2011. Spatial patterns of plant diversity below-ground as revealed by DNA barcoding. Molecular Ecology 20:1289–1302.

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

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

Masella, A. P., A. K. Bartram, J. M. Truszkowski, D. G. Brown, and J. D. Neufeld. 2012. PANDAseq: paired-end assembler for illumina sequences. BMC Bioinformatics 13:31.

Masters, M. D., C. K. Black, I. B. Kantola, K. P. Woli, T. Voigt, M. B. David, and E. H. DeLucia. 2016. Soil nutrient removal by four potential bioenergy crops: Zea mays, Panicum virgatum, Miscanthus× giganteus, and prairie. Agriculture, Ecosystems & Environment 216:51–60.

McMurdie, P. J., and S. Holmes. 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. Plos One 8:e61217.

Oksanen, J., F. G. Blanchet, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O’Hara, G. L. Simpson, P. Solymos, M. H. H. Stevens, E. Szoecs, and H. Wagner. 2016. vegan: Community ecology package.

R Core Team. 2016. R: A language and environment for statistical computing. Version 3.3.1. R Foundation for Statistical Computing, Vienna, Austria.

Rewald, B., C. Meinen, M. Trockenbrodt, J. E. Ephrath, and S. Rachmilevitch. 2012. Root taxa identification in plant mixtures – current techniques and future challenges. Plant and Soil 359:165–182.

Rognes, T., T. Flouri, B. Nichols, C. Quince, and F. Mahé. 2016. VSEARCH: a versatile open source tool for metagenomics. Peerj.

Smith, C. M., M. B. David, C. A. Mitchell, M. D. Masters, K. J. Anderson-Teixeira, C. J. Bernacchi, and E. H. DeLucia. 2013. Reduced Nitrogen Losses after Conversion of Row Crop Agriculture to Perennial Biofuel Crops. Journal of Environment Quality 42:219.

Staats, M., A. J. Arulandhu, B. Gravendeel, A. Holst-Jensen, I. Scholtens, T. Peelen, T. W. Prins, and E. Kok. 2016. Advances in DNA metabarcoding for food and wildlife forensic species identification. Analytical and Bioanalytical Chemistry 408:4615–4630.

Tucker, S. S., J. M. Craine, and J. B. Nippert. 2011. Physiological drought tolerance and the structuring of tallgrass prairie assemblages. Ecosphere 2.

Van Geel, M., P. Busschaert, O. Honnay, and B. Lievens. 2014. Evaluation of six primer pairs targeting the nuclear rRNA operon for characterization of arbuscular mycorrhizal fungal (AMF) communities using 454 pyrosequencing. Journal Of Microbiological Methods 106:93–100.

Wardle, D. A., and D. A. Peltzer. 2003. Interspecific interactions and biomass allocation among grassland plant species. Oikos 100:497–506.

Weaver, J. E. 1919. The Ecological Relations of Roots. Carnegie Institution of Washington, Washington.

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

Wilson, S. D. 2014. Below‐ground opportunities in vegetation science. JOURNAL OF VEGETATION SCIENCE 25:1117–1125.

Zeri, M., K. J. Anderson-Teixeira, G. C. Hickman, M. D. Masters, E. H. DeLucia, and C. J. Bernacchi. 2011. Carbon exchange by establishing biofuel crops in Central Illinois. Agriculture, Ecosystems & Environment 144:319–329.

Figure 1: Correlation between percent aboveground cover (horizonal axis) and read proportion per root sample (vertical axis). Points are genus means for one experimental plot (N=5 plots); bars are 1 standard errors in each direction. Root proportions are averaged across all depths.

Figure 1: Correlation between percent aboveground cover (horizonal axis) and read proportion per root sample (vertical axis). Points are genus means for one experimental plot (N=5 plots); bars are 1 standard errors in each direction. Root proportions are averaged across all depths.

Figure 2: 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 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.

Figure 2: 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 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.

Figure 3: Nonmetric multidimensional scaling plot showing centroids for all detected species. Red: Poaceae. Green: Asteraceae. Blue: Fabaceae. Black: Other families. Grey crosses: Low-abundance species, unlabeled for figure clarity. Pink ellipses show centroid SD for each depth and pink polygons show enclosing hulls for each depth. TK: Show depth, C, N as arrows instead.

Figure 3: Nonmetric multidimensional scaling plot showing centroids for all detected species. Red: Poaceae. Green: Asteraceae. Blue: Fabaceae. Black: Other families. Grey crosses: Low-abundance species, unlabeled for figure clarity. Pink ellipses show centroid SD for each depth and pink polygons show enclosing hulls for each depth. TK: Show depth, C, N as arrows instead.

Figure 4: Figure still TK: Total root mass, soil texture and C/N content by depth. Root mass is taken from Black et al (submitted); soil texture is from (Smith et al. 2013); C and N were collected for this paper.

## Appendix: Supplemental figures

Figure 5: Read counts by assigned genus for water controls. Each panel shows reads from a separate 2015 µL 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) and that the scale differs between panels.

Figure S1: Read counts by assigned genus for water controls. Each panel shows reads from a separate 2015 µL 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) and that the scale differs between panels.

Figure 6: Genus identities of 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.

Figure S2: Genus identities of 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.

Figure 7: Species identities of 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. Notice that the barcoding assignment identifies the dominant sequence from the Andropogon gerardii sample as Solidago, not a grass!

Figure S3: Species identities of 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. Notice that the barcoding assignment identifies the dominant sequence from the “*Andropogon gerardii*” sample as *Solidago*, not a grass!

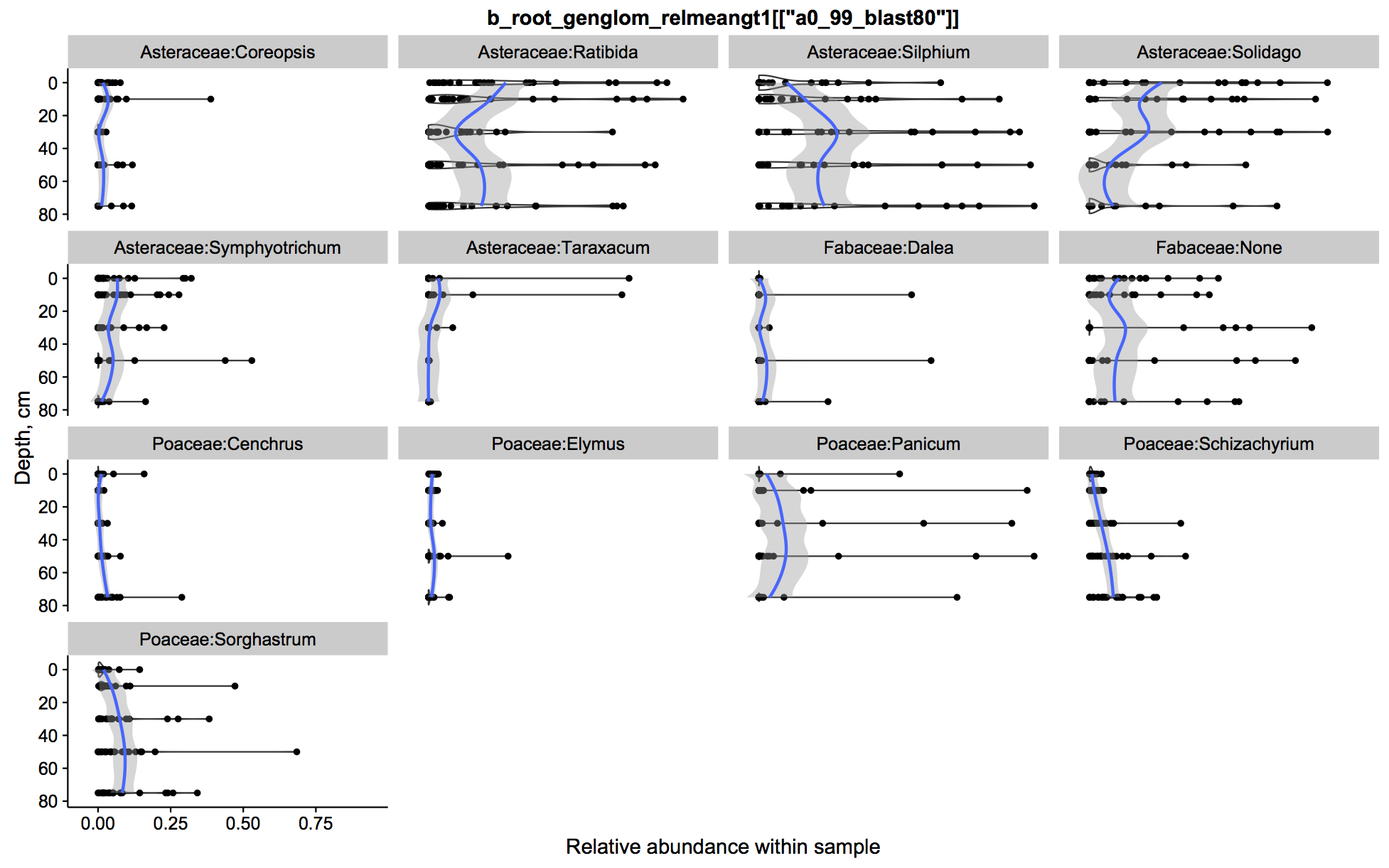


Figure S4: Relative abundance (fraction of reads from each sample) as a function of sample depth for each observed plant genus. Sequences were clustered at 99% similarity and identified to species according to the closest BLAST match against the Genbank nt database. Taxa were then collapsed by genus and groups with a mean abundance less than 1% per sample were removed for plotting.