What species are the roots in these samples from the restored prairie treatment of a perennial bioenergy experiment, anyway?

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

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

The Prairie Chapter

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.

Introduction

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 outcomes has long been recognized (Weaver 1919, Bardgett et al. 2014). Knowing the physical placement of different species in a belowground community is thus fundamental to understanding ecosystem functions. Existing evidence suggests that grassland communities may be more diverse and more strongly structured belowground than aboveground (Kesanakurti et al. 2011, Hiiesalu et al. 2012, Wilson 2014) and that the physiological capacity to exploit soil resources is a strong predictor of competitive success (Tucker et al. 2011, Kroon et al. 2012, Hendriks et al. 2015), but it is not clear how universal these patterns are.

Progress in understanding these belowground dynamics is considerably 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 and Voigt 1950). Visual identification of roots based on anatomical or morphological characteristics is sometimes possible (Wardle and 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).

With the increasing availability of high-throughput DNA sequencing, molecular methods now provide a partial solution to this dilemma. DNA-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, sequencing potentially can 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 of mixed root species recovered from soil: 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 (~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 sample 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 restored prairie in central Illinois, and to use these identities to infer differences in species roles for water usage, carbon storage, and plant-microbe interactions by prairie plant communities. We used a DNA metabarcoding approach: we extracted fine root samples from soil cores bulked by depth and extracted whole-community DNA for ITS2 amplicon library construction (Fliudigm Access Array)

followed by sequencing (Illumina MiSeq 2x300), then identified the detected sequences against those from a checklist of species known present from aboveground surveys.

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 (Maize-soybean rotation, monocultures of *Miscanthus* $\x$ *giganteus* and *Panicum virgatum*, and the prairies decribed below) are grown side-by-side in a randomized complete-block design with five replicates: Four blocks of 0.7 Ha plots and a fifth block of 3.8 Ha plots that is instrumented for eddy-covariance measurements of carbon and water exchange balance. For this experiment we were interested in multispecies root communities and therefore sampled only from 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. In 2008, the prairie plots were seeded with a mix of 28 species native to Illinois (Pizzo and Associates, Leland IL), treated 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. 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).

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 120 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 3 cores within a 2 meter area. We pooled all three cores from each location by depth increment (0-10, 10-30, 30-50, 50-75, and 75-100 cm), collected a ~0.5-1 kg subsample of mixed roots and soil from each depth, and returned the remaining material to the holes. 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 individual genets, identified them to species by leaf and flower morphology, and used a trowel to extract known single-species 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, rhizosphere soil, and bulk 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. We considered any soil that remained attached to root hairs, was not chunky enough to break off with forcepts, and was not removed by a gentle shake, to be rhizosphere soil. The remaining material after all visible roots were removed was considered bulk soil. 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.

We then rinsed all roots in two changes of sterile water, recovered the rhizosphere soil by centrifugation, further cleaned the root surfaces by sonication in sterile water for 10 minutes and discarded the wash water, then lyophilized all three compoments (root, rhizosphere, and a 50-mL subsample of the bulk soil) and stored them at room temperature. Single-species root voucher samples were treated identically to the mixed root samples with the exception that all bulk soil had been removed at collection time, so no sieving or hand-picking steps were necessary.

DNA extraction and amplification

To maximize extraction 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

141 Bio; personal communication): We diluted the DNA to a volume of 100 µL with DNAse-
142 free water, added 50 µL of bead beating solution and inverted to mix, then added 25 µL
143 each of solutions C2 and C3, inverted to mix, and centrifuged at 10000 xg for 2 minutes.
144 We then collected the supernatant, added 2 volumes of solution C4, vortexed, and loaded
145 the sample onto a spin filter. Finally, we washed with 500 µL of solution C5 and eluted
146 with 50 µL of C6.
147 TK this paragraph needs many details TK: After extraction, we submitted whole
148 DNA to the W.M. Keck Center (Urbana, IL, USA) for amplification and sequencing. In
149 addition to DNA extracted from roots and rhizosphere soil, we included nominally-pure
150 extracts of root DNA from four species with high aboveground abundance (*Andropogon*
151 *gerardii*, *Sorghastrum nutans*, *Silphium perfoliatum*, and *Elymus nutans*; Feng 2014),
152 water extractions as a negative control on the DNA extraction + PCR + sequencing
153 process taken as a whole, and a mock community of DNA from 31 species combined in
154 equimolar quantity. Each sample was amplified by microfluidic PCR using the Fluidigm
155 Access Array chip to create amplicons from five primer sets targeting different regions of
156 the ribosomal RNA genes of diverse phylogenetic groups: Bacterial 16S V4 [cite TK],
157 fungal ITS [cite TK], SSU of Glomeromycota (arbuscular mycorrhizal fungi) (Van Geel
158 et al. 2014), eukaryotic 18S [cite TK], and plant ITS2 (Chen et al. 2010)). The resulting
159 amplicons were then barcoded [TK: Do we need to present full linker constructs?] and
160 sequenced by synthesis for 2x301 paired-end cycles (MiSeq, V3 chemistry, Illumina Inc,
161 San Diego CA). For the remainder of this paper, we discuss only the results obtained
162 from root samples using the plant ITS2 primers S2F (5'-

163 ATGCGATACTTGGTGTGAAT) and S3R (5'-GACGCTTCTCCAGACTACAAT)
164 (Chen et al. 2010).
165 *Data processing*
166 The raw Illumina read files were separated into one file from each primer set and
167 PhiX reference reads were removed using CASAVA 1.8. We then used cutadapt 1.8.1
168 (Martin 2011) to trim primers, discard all reads that did not begin with the expected
169 primer, and trim 3' bases with a Phred quality score below 20. We then joined the
170 overlapping ends of each read using using the RDP maximum likelihood algorithm (Cole
171 et al. 2013) as implemented in Pandaseq 2.10 (Masella et al. 2012) using a minimum
172 alignment quality of 0.8, a minimum assembled length of 25 bases, and a minimum
173 overlap of at least 20 "bits saved" (corresponds to ~10 bases; see Cole et al. 2013).
174 We then used the `split_libraries.py` script in QIIME 1.9.1 (Caporaso et
175 al. 2010) to assign barcodes to sequence identities. We then dereplicated sequences and
176 removed singletons and suspected PCR chimeras using VSEARCH 2.0.4 (Rognes et al.
177 2016), extracted full-length ITS2 variable regions using ITSx 1.0.11 (Bengtsson Palme et
178 al. 2013), clustered OTUs using VSEARCH with a similarity threshold of 99%, and
179 assigned taxonomy using BLAST+ (Camacho et al. 2009) against the GenBank `nt`
180 database. After taxonomy assignment, we collapsed all OTUs assigned as the same
181 phylotype (species, genus, or family depending on the analysis of interest) into single
182 taxon groups, then removed taxa with a mean abundance of less than 1% of the reads per
183 sample.
184 Full analysis scripts and raw sequence data are available online at (Dryad URL
185 TK).

Statistical analysis: This section still TK.

Two possible approaches: rarefy samples to even depth and report fraction of reads observed in each sample (this throws away a lot of information and systematically understates variance), or alternately model raw counts with a negative binomial distribution to account for overdispersion (conceptually straightforward, computationally finicky).

Results and what Discussion there currently is

The sequencing run returned 1286163 plant ITS2 reads, of which 956812 were successfully end-paired, quality filtered, and assigned back from barcodes to samples. These clean reads contained 675898 unique sequences, of which 635969 were observed only once in the dataset (singletons). Of the 39929 sequences observed at least twice, 1537 were identified as probable PCR chimeras by `vsearch`. These chimeric sequences accounted for a total of 5643 reads, or 1.8% of the raw dataset. A further 1854 sequences were identified as incomplete ITS2 regions by `ITSx`, leaving a total of 36142 sequences to be clustered into OTUs.

We observed more reads from grasses at increasing depths, both relative to other functional types (Fig. 1) and in absolute abundance (Fig. 2). Forbs from the Asteraceae declined somewhat with depth, while reads from legumes were consistent across depths.

When grouped by genus, additional differences between forbs were visible (Fig. 3, Fig. 4): Within the Asteraceae, Ratidida (fibrous rooted) and Siplhium (tap-rooted) appear to trade off, with Ratidida less prevalent at middle depths than either shallow or deep layers while Silphium was rare at the surface but most common at middle depths. The fibrous-rooted genera Coreopsis, Solidago, Symphotrichum, and the

209 shallow-taprooted Taraxacum, all declined with depth. All genera of grasses were
210 generally higher in abundance at deeplayers than near the surface, but Panicum, Elymus,
211 and Sorghastrum had their greatest abundance at 75 cm while Andropogon and
212 Schizachrium both continued to increase all the way to 100 cm.

213 However, some caution is necessary when interpreting these identities. When
214 multiple species from the same genus were detected, they tended to co-occur in the same
215 samples (Fig. 5; note the vertical stripes within each grouped genus). Since most species
216 within each of the genera known from the site are very similar to each other in functional
217 type and growth habit, it is possible that this indicates environmental filtering, where
218 related species tend to grow in the same locations. Or, more parsimoniously, these
219 "multiple species" may be bioinformatic artifacts, either from sequencer error or from
220 genuine biological variation, and all or most of the reads in these groups actually derive
221 from same taxa. Replotting the same heatmap after removing the rarest taxa (sequences
222 assigned to species with mean abundance less than 1% per sample) reduces this effect
223 (Fig. 6) but does not remove it entirely. Furthermore, note that the lines for species from
224 genera represented by multiple species in the aboveground checklist (Rudbeckia,
225 Silphium, Solidago) are less similar to each other than those from genera that have only
226 one representative in the checklist. Some refinement of the barcoding pipeline will
227 probably be needed before our taxonomic assignments are trustworthy below the genus
228 level.

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

252 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species. Plos
253 One 5:e8613.
254 Cole, J. R., Q. Wang, J. A. Fish, B. Chai, D. M. McGarrell, Y. Sun, C. T. Brown, A.
255 Porras-Alfaro, C. R. Kuske, and J. M. Tiedje. 2013. Ribosomal Database Project:
256 data and tools for high throughput rRNA analysis. Nucleic Acids Research
257 42:D633–D642.
258 Fahner, N. A., S. Shokralla, D. J. Baird, and M. Hajibabaei. 2016. Large-Scale
259 Monitoring of Plants through Environmental DNA Metabarcoding of Soil:
260 Recovery, Resolution, and Annotation of Four DNA Markers. Plos One
261 11:e0157505.
262 Feng, X. 2014. Productivity, physiology, community dynamics, and ecological impacts of
263 a grassland agro-ecosystem: integrating field studies and ecosystem modeling.
264 PhD thesis, University of Illinois At Urbana-Champaign; University of Illinois At
265 Urbana-Champaign.
266 Hendriks, M., J. M. Ravenek, A. E. Smit Tiekstra, J. W. Paauw, H. Caluwe, W. H.
267 Putten, H. Kroon, and L. Mommer. 2015. Spatial heterogeneity of plant–soil
268 feedback affects root interactions and interspecific competition. New Phytologist
269 207:830–840.
270 Hiiesalu, I., M. Öpik, M. Metsis, L. LILJE, J. Davison, M. Vasar, M. Moora, M. Zobel,
271 S. D. Wilson, and M. Pärtel. 2012. Plant species richness belowground: higher
272 richness and new patterns revealed by next□generation sequencing. Molecular
273 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

296 mays, Panicum virgatum, Miscanthus× giganteus, and prairie. Agriculture,
297 Ecosystems & Environment 216:51–60.
298 Rewald, B., C. Meinen, M. Trockenbrodt, J. E. Ephrath, and S. Rachmilevitch. 2012.
299 Root taxa identification in plant mixtures – current techniques and future
300 challenges. Plant and Soil 359:165–182.
301 Rognes, T., T. Flouri, B. Nichols, C. Quince, and F. Mahé. 2016. VSEARCH: a versatile
302 open source tool for metagenomics. Peerj.
303 Smith, C. M., M. B. David, C. A. Mitchell, M. D. Masters, K. J. Anderson-Teixeira, C. J.
304 Bernacchi, and E. H. DeLucia. 2013. Reduced Nitrogen Losses after Conversion
305 of Row Crop Agriculture to Perennial Biofuel Crops. Journal of Environment
306 Quality 42:219.
307 Staats, M., A. J. Arulandhu, B. Gravendeel, A. Holst-Jensen, I. Scholtens, T. Peelen, T.
308 W. Prins, and E. Kok. 2016. Advances in DNA metabarcoding for food and
309 wildlife forensic species identification. Analytical and Bioanalytical Chemistry
310 408:4615–4630.
311 Tucker, S. S., J. M. Craine, and J. B. Nippert. 2011. Physiological drought tolerance and
312 the structuring of tallgrass prairie assemblages. Ecosphere 2.
313 Van Geel, M., P. Busschaert, O. Honnay, and B. Lievens. 2014. Evaluation of six primer
314 pairs targeting the nuclear rRNA operon for characterization of arbuscular
315 mycorrhizal fungal (AMF) communities using 454 pyrosequencing. Journal Of
316 Microbiological Methods 106:93–100.
317 Wardle, D. A., and D. A. Peltzer. 2003. Interspecific interactions and biomass allocation
318 among grassland plant species. Oikos 100:497–506.

319 Weaver, J. E. 1919. The Ecological Relations of Roots. Carnegie Institution of
320 Washington, Washington.
321 Weaver, J. E., and J. W. Voigt. 1950. Monolith method of root-sampling in studies on
322 succession and degeneration. Botanical Gazette 111:286–299.
323 Wilson, S. D. 2014. Below□ground opportunities in vegetation science. JOURNAL OF
324 VEGETATION SCIENCE 25:1117–1125.
325 Zeri, M., K. J. Anderson-Teixeira, G. C. Hickman, M. D. Masters, E. H. DeLucia, and C.
326 J. Bernacchi. 2011. Carbon exchange by establishing biofuel crops in Central
327 Illinois. Agriculture, Ecosystems & Environment 144:319–329.

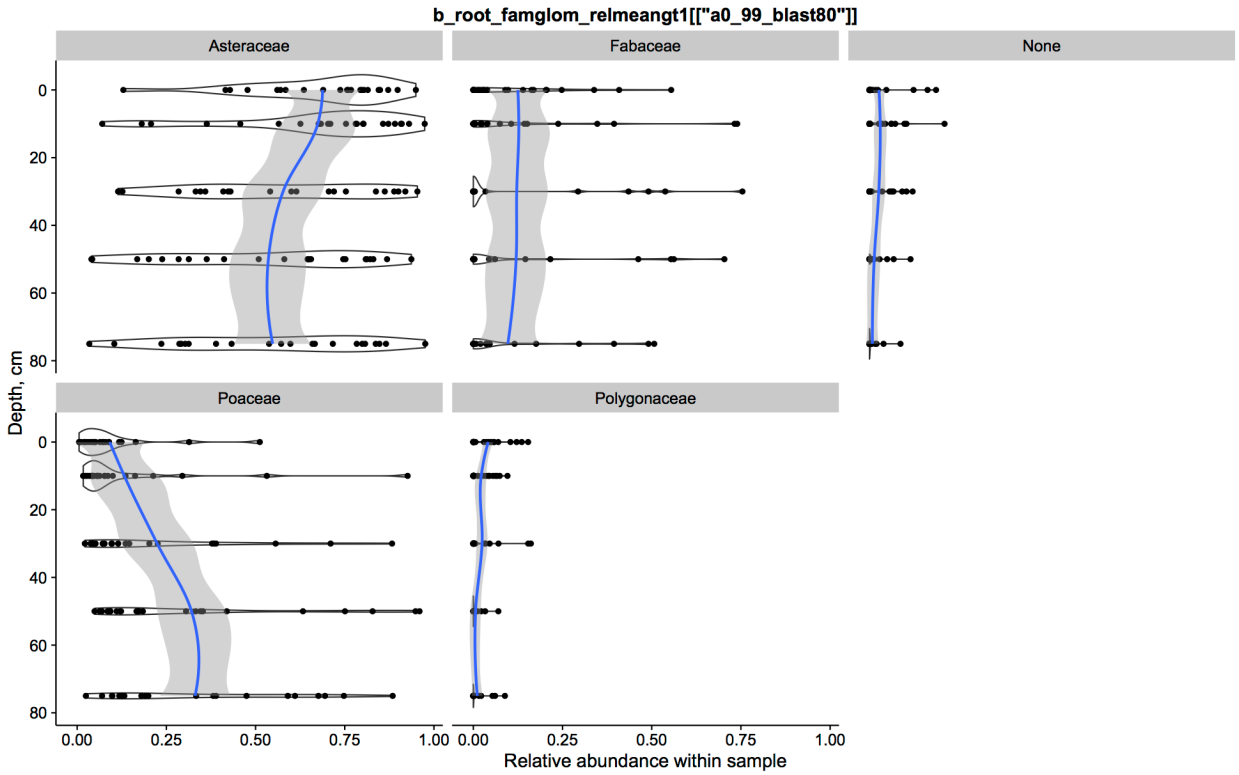

Figure 1: Relative abundance (fraction of reads from each sample) as a function of sample depth for each observed plant family. Sequences were clustered at 99% similarity and identified to species according to the closest BLAST match against the Genbank `nt` database. Taxa were then collapsed by family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

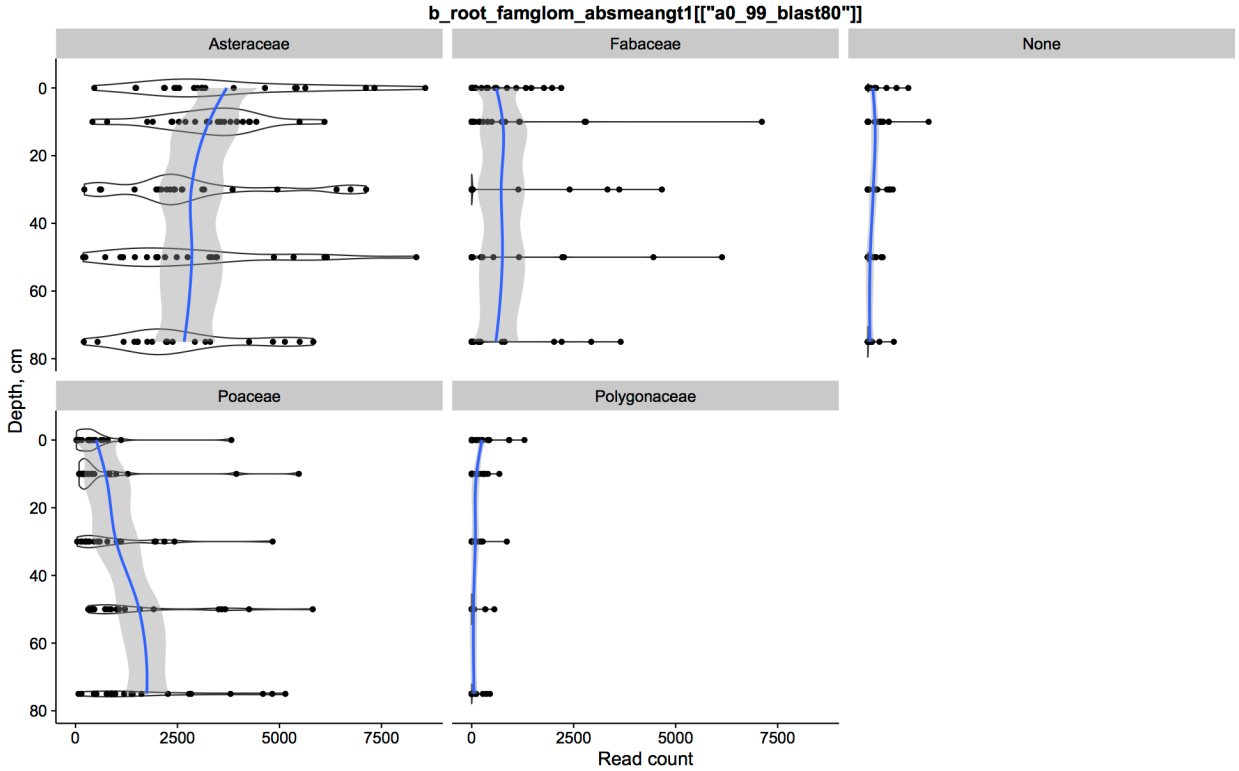

Figure 2: Read count from each sample as a function of sample depth for each observed plant family. Sequences were clustered at 99% similarity and identified to species according to the closest BLAST match against the Genbank `nt` database. Taxa were then collapsed by family and groups with a mean abundance less than 1% per sample were removed for plotting.

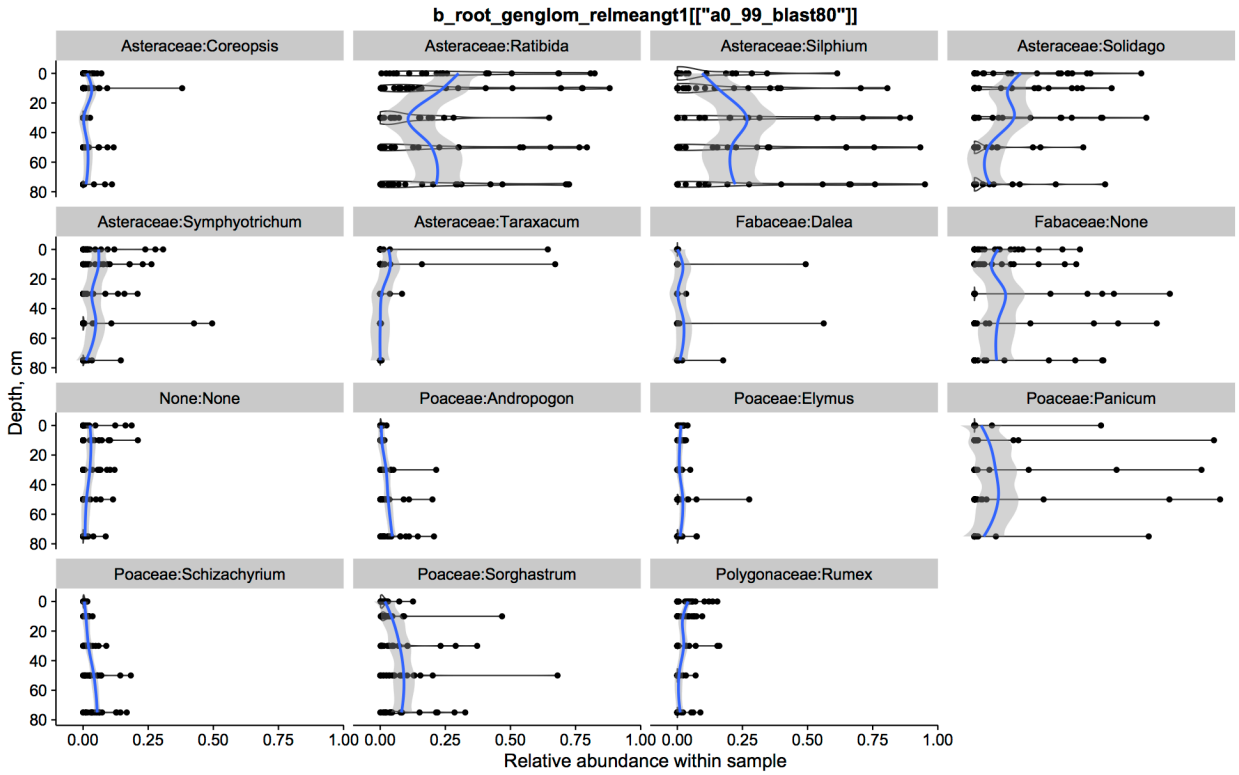

Figure 3: Relative abundance (fraction of reads from each sample) as a function of sample depth for each observed plant genus. Sequences were clustered at 99% similarity and identified to species according to the closest BLAST match against the Genbank `nt` database. Taxa were then collapsed by genus and groups with a mean abundance less than 1% per sample were removed for plotting.

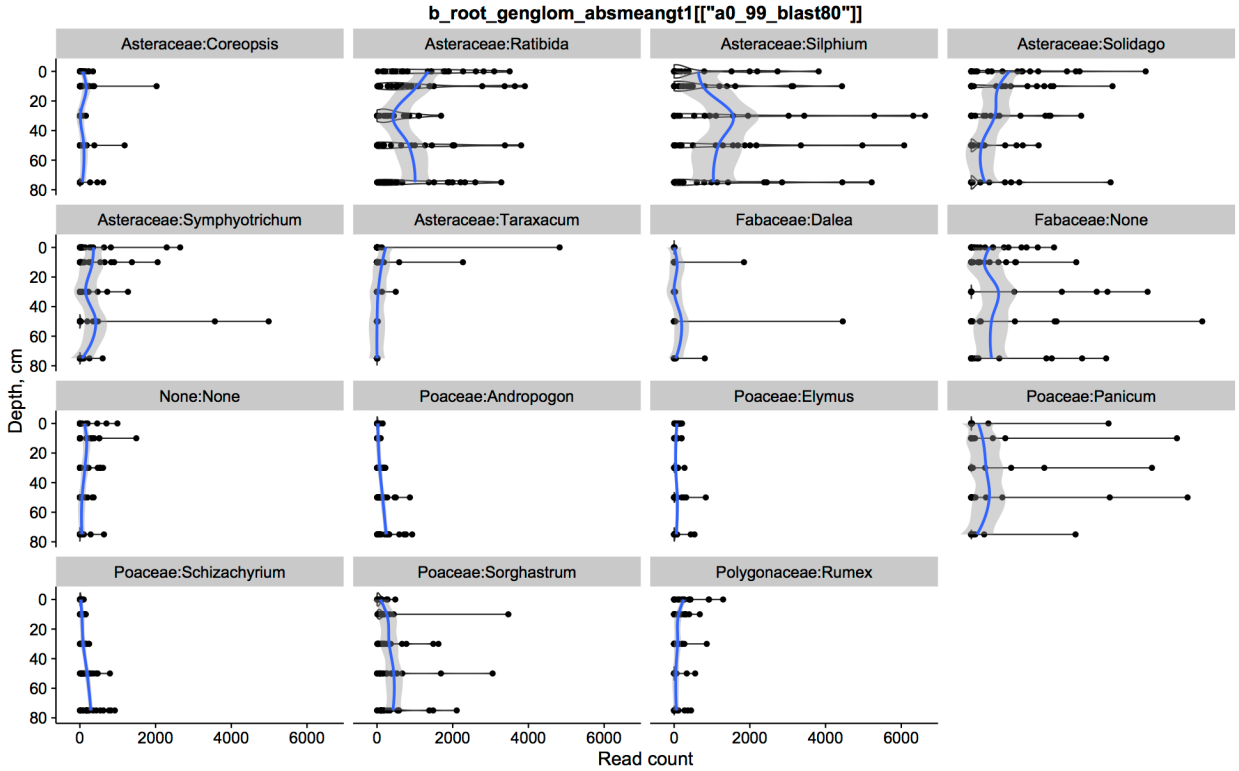

Figure 4: Read count from each sample as a function of sample depth for each observed plant genus. Sequences were clustered at 99% similarity and identified to species according to the closest BLAST match against the Genbank `nt` database. Taxa were then collapsed by genus and groups with a mean abundance less than 1% per sample were removed for plotting.

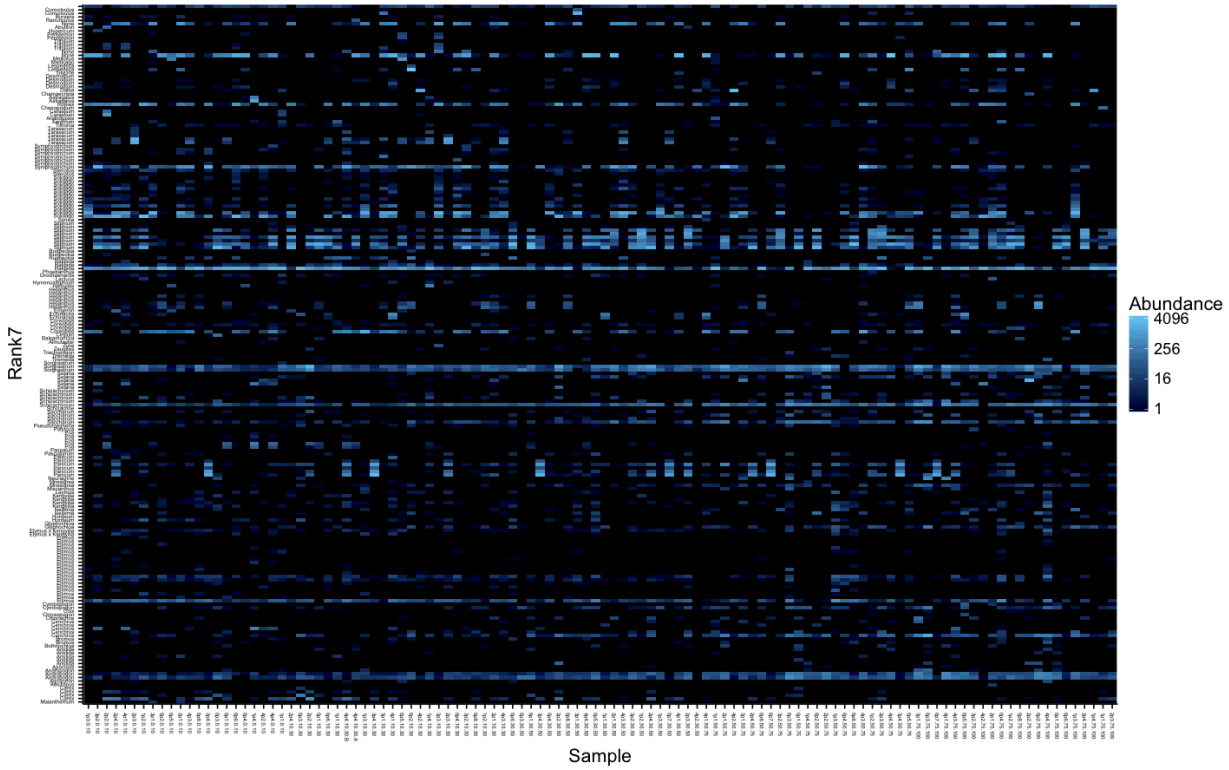

Figure 5: Heat map showing observed read counts per sample (x axis, ordered by depth) for sequences classified to 200 species (y axis, ordered & labeled by genus).

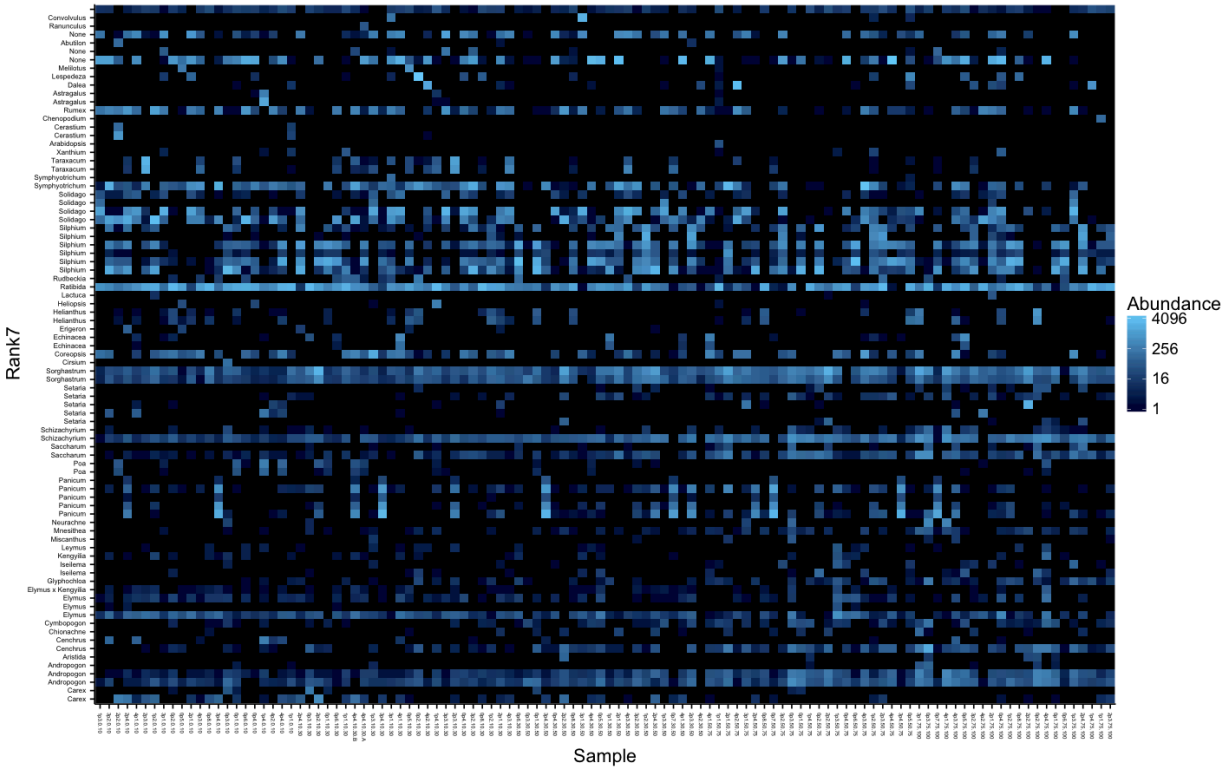

Figure 6: Heat map showing observed read counts per sample (x axis, ordered by depth) for sequences classified to 83 species (y axis, ordered & labeled by genus). Sequences were clustered at 99% similarity and identified to species according to the closest BLAST match against the Genbank `nt` database. Taxa were then collapsed by assigned species and groups with a mean adbundance less than 1% per sample were removed for plotting.