Supplementary Information (SI) for Environmental Science: Processes & Impacts.
This journal is © The Royal Society of Chemistry 2024

Supplementary Information

Fungal Diversity and Key Functional Gene Abundance in Iowa Bioretention Cells: Implications for Stormwater Remediation Potential

Erica A. Wiener,[1,2] *Jessica M. Ewald*, [1,2] and *Gregory H. LeFevre*[*,1,2]

[1]Department of Civil & Environmental Engineering, University of Iowa, 4105 Seamans Center, Iowa City, IA 52242, United States; [2]IIHR-Hydroscience &Engineering, 100 C. Maxwell Stanley Hydraulics Laboratory, Iowa City, IA 52242, United States

***Corresponding Author:**
GHL: gregory-lefevre@uiowa.edu; Phone: 319-335-5655; Fax: 319-335-5660; 4105 Seamans Center for Engineering, University of Iowa, Iowa City IA, United States

Includes: 26 total pages, 25 Figures, and 12 Tables.

Bioretention Cell Information

Table S.1: Site information for each bioretention cell sampled, including the number of replicate soil samples for the site, the runoff type the cell receives, planting categories, and dominant planting type. Plant categories were noted based on plants observed growing in the cell on the sampling date. Dominant planting type was determined via site notes and pictures.

Site ID	Number of Replicates Within Cell	Runoff Type	Planting Categories	Dominant Planting Type
cv_01	4	parking lot	Native grass, forbs, legumes	native grass
cv_02	4	parking lot	Native grass, forbs, legumes	native grass
cv_03	4	parking lot	Legumes, forbs, tree/shrub, native grass	native grass
cv_05	2	parking lot	Tree/shrub, native grass	Tree/shrub
cv_06	3	parking lot	Native grass, tree/shrub, forb	Tree/shrub
cv_07	3	minor road	Native grass, forb, sedge, decorative	mixed
cv_08	3	minor road	Native grass, forb, decorative	mixed
cv_09	3	roof	Forb, native grass	native grass
cv_10	3	roof	Forb, native grass, decorative	native grass
cv_11	3	roof	Annual/weed, forb, native grass	weed
cv_12	4	parking lot	Turf Grass	turf
cv_13	4	parking lot	Native grass	native grass
cv_14T	4	minor road	Forb, native grass, weeds	forb
cv_14B	4			forb
cv_15	4	parking lot	Annual/weed, legume	weed
cv_16	3	parking lot	Turf Grass	turf
cv_17	4	parking lot	Turf Grass	turf
cv_18	4	parking lot	Native grass, forb, weeds, decorative	native grass
cv_19	2	parking lot	Native grass, tree/shrub, decorative	mixed

cv_20	3	major road	Native grass, decorative	native grass
cv_21	4	major road	Tree/shrub, native grass	native grass
nl_1	4	major road	Legumes, forb	legume
nl_2	4	major road	Legumes, forb	legume
cr_1	4	major road	Forb, native grass, decorative	native grass
cr_2	4	major road	Forb, native grass	native grass
cr_3	4	parking lot	Forb, native grass, decorative	native grass
cr_4	4	parking lot	Native grass, forb, tree/shrub	mixed
ic_5	5	parking lot	Forb, native grass, decorative	native grass

Figure S.1: Picture of a bioretention cell. Bioretention cells are typically depressed areas to receive stormwater, filled with a mostly sandy media for quick infiltration, and often vegetated. This bioretention cell has an overflow drain structure (the circular grate).

Figure S.2: Picture of a bioretention cell (with co-author LeFevre) with different dominant plants outlined with white shapes as an example of a cell that would be classified as having a "mixed" dominant planting type.

Table S.2: Parameters requested for sequencing using MR DNA's services.

Parameter Name	Specification
ITS Amplicon Primers	Forward (fITS7[36], 5'-3'): GTGARTCATCGAATCTTG Reverse (ITS4[37], 5'-3'): TCCTCCGCTTATTGATATGC
Sequencing Technology	Illumina MiSeq (paired-end)
Number of Technical Replicates	2 PCR replicates per sample

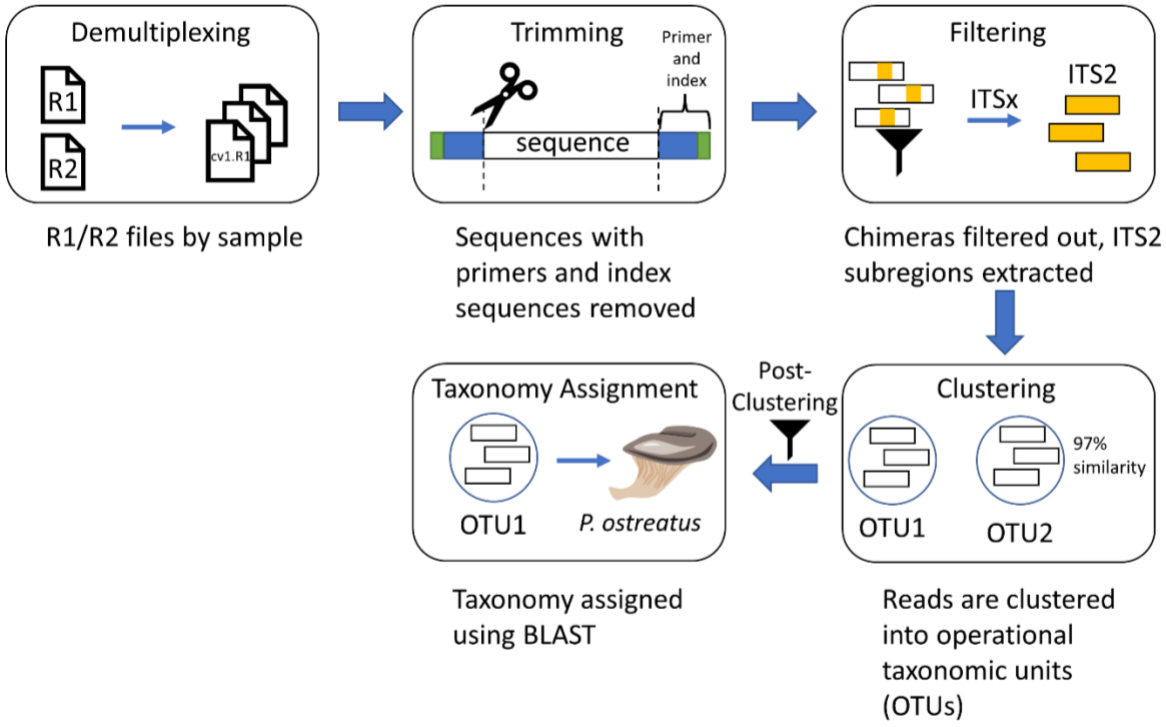

Figure S.3: Flowchart of bioinformatics pipeline used within PipeCraft2 (excepting the Demultiplexing step). The steps with PipeCraft2 names, in order, are: Cut Primers, Merge Reads, Quality Filtering, Chimera Filtering, ITSX, Clustering, Post-Clustering, Filter Tag Jumps, and Assign Taxonomy.

Table S.3: Primers used for qPCR.

Gene Target	Function	Sequences (5'→3')	Reference
nirKf	fungal nitrite reductase	Forward: TACGGGCTCATGTAYGTNSARCC Reverse: AGGAATCCCACASCNCCYTTNTC	[48]
Cu1	basidiomycete laccase	Forward: CAYTGGCAYGGNTTYTTYCA Reverse: GRCTGTGGTACCAGAANGTNCC	[49]
Cu1A	ascomycete laccase	Forward: ACMWCBGTYCAYTGGCAYGG Reverse: GRCTGTGGTACCAGAANGTNCC	[50]

Table S.4: qPCR amplification conditions (i.e., component compositions) for all three target genes.

Component	Target Volume and Final Concentration (as applicable)
Power SYBR Green PCR Master Mix (Applied Biosystems)	10 µL
Forward Primer	0.18 µL (900 nM)
Reverse Primer	0.18 µL (900 nM)
Bovine Serum Albumin (New England BioLabs, 20 mg/mL stock concentration)	0.4 µL (0.4 mg/mL)
Ultrapure water	7.24 µL
Template	2 µL
Total Volume	20 µL

(Note that the *Power* SYBR Green PCR Master Mix contains ROX reference dye and AmpliTaq Gold DNA Polymerase, UP.)

Soil Physicochemical Data

Table S.5: Data received from MVTL for soil physicochemical characteristics. Buffer pH is an indicator of "stored acidity" and is often measured for agricultural purposes to indicate if liming is necessary. Phosphorus was measured using two different methods. Bray Phosphorus is often used for soils with pH less than 7.4, and Olsen for soils with pH greater than 7.4.

Sample ID	pH	Buffer pH	Nitrate (ppm)	Organic Matter (%)	Phosphorus (Bray I Method, ppm)	Phosphorus (Olsen Method, ppm)	Potassium (ppm)	Zinc (ppm)	Copper (ppm)	Manganese (ppm)	Ammonium (ppm)
cv_cell01	7.5	7.6	8.5	5.5	197	44	64	10.2	1.4	1.3	1.8
cv_cell02	7.1	7.5	19.8	11	152	46	68	12	3.6	1.4	1.7
cv_cell03	7.4	7.5	19.8	7.1	194	47	81	10.7	1.6	1.8	3.1
cv_cell05	7.7	7.5	6	10	70	51	140	6.3	0.5	2.1	3.7
cv_cell06	7.3	7.6	2.7	2.2	80	29	69	2.4	0.2	1.1	2
cv_cell07	7.6	7.6	0.3	5.8	78	42	162	3.1	0.6	2.3	2.8
cv_cell08	7.4	7.4	16.1	2.1	94	47	298	3.6	0.7	1.4	2.5
cv_cell09	6.6	7.5	10.1	1.8	106	24	23	3.6	0.8	0.6	2.3
cv_cell10	7.1	7.5	7.3	2.2	99	31	26	4.8	1	0.8	3.2
cv_cell11	7	7.5	5.4	1.5	30	10	45	3.1	0.2	1.1	2.2
cv_cell12	6.8	7.4	54	3.2	38	22	205	2.2	0.4	1.4	2.3
cv_cell13	7.6	7.6	26.6	3.3	19	11	113	2.5	0.6	1.5	1.7
cv_cell14	7.6	7.6	17.6	3.6	61	28	89	3.2	0.4	1.2	2.4
cv_cell15	7.3	7.5	54.1	4.3	23	10	61	11.6	0.6	1.9	1.9
cv_cell16	7.4	7.4	44.1	3.2	58	41	136	4	1.4	1.8	2.4
cv_cell17	7.5	7.5	45	2.3	44	16	64	4.8	0.7	1	1.9
cv_cell18	7.6	7.6	28.9	3.1	44	17	55	7.7	0.7	1.2	2.1
cv_cell19	8	7.7	31.8	2.7	52	20	81	4.3	0.6	1.1	1.6
cv_cell20	7.8	7.6	24	4.2	26	12	49	10.1	1.1	1.4	1.8
cv_cell21	7.9	7.6	13.8	2.5	14	7	48	5.8	0.8	1	1.3
nl_cell_1	7.7	7.5	18.7	5.1	96	55	144	6.2	1	0.9	2.3
nl_cell_2	7.9	7.5	24.8	3.4	107	42	144	5.9	0.9	0.7	2.1
cr_cell_1	7.1	7.3	2.7	5.6	55	21	109	5	1.2	2.3	3.1
cr_cell_2	7.3	7.4	16.5	5.9	51	24	95	7.8	1.8	2.5	2.9
cr_cell_3	7.6	7.6	29.7	4.4	30	12	48	10.9	0.7	2.1	2.7
cr_cell_4	7.7	7.6	0.3	9.6	37	20	70	7.4	0.6	1.6	2.9
ic_cell_5	7.4	7.6	20.9	2.2	84	21	41	3.7	0.5	0.8	1.7

PCR Standard Information

Table S.6: Sequences for gBlock standards for qPCR. Primer binding sites are bolded.

Gene Target	Sequence	GC Content (%)	Sequence Length (bp)	GenBank accession number
nirKf	GCATG**TACGGCCTCATGTACGT GCAGCC**GGAAGGCAATGACCT ACCACCCGTGGACAAGGAGTAT TATGTTATGCAGAGCGAGTTCT ACCATGAACCACCAGAGGTTGA TGACGACGGCCGACGATCCGA AATTGTCGAGTTCTCATATCCCA ATGGTCTTCGCGAAGAACCTCA AGTCGTTGCCTTCAACGGTAGT GAGTCTGCCTTGACAAGAGATC ATCCTCTCAAGGCTCATGTTGG TGACGATGTGAGAATCTTCTTC GGAAATGCTGGACCAAACCTGA CTAGCTCATTCCATATCATCGG CACGCACTTCAAGAATGTGTAC CGAGATGGTGGCGTGACCAGC AACCCATCAAAGGGTATTCAGA CCGTTTCTGTTCCTTGCGGTGG GTCAACGATCGTCGACTTGAAG ATGGCTGTCCCAGGAACGTATA CACTGGTTGATCATTCCATTTTC CGCTTG**GACAAGGGAGCAGTG GGTTTTCT**CAATG	50.3	495	EF600898.1
Cu1	CGCAG**CATTGGCACGGCTTCTT CCA**GAAGGGTACTAACTGGGCT GATGGAGCTGCCTTCGTCAACC AGTGCCCTATCGCGACGGGGA ACTCTTTCCTTTACGACTTCACC GCGACGGACCAAGCAGGTCAG TGCCTGTGGCGCTTATGTTTTC CCGTAATCAGCAGCTAACACTC CGCACCCACA**GGCACCTTCTG GTACCACAGTC**ACTTG	56.6	212	X84683.1
Cu1A	**ACAACTGTTCACTGG**CACGGC GTGTGCCAGAAAGGGACGAAT GCTATGGATGGGCCTATTGTTG CTACACAATGCCCTGTGGCCCC GGATCATAGTATTGTGTATGATT TTGTTGTCAGTCCTCTAGATCT GTCTAGAAACTAATAATGCATCA AAATCATTACTATTTGGGAATCT CAAGATATTGATTTGGTGAAATA TGTTGATCAACCA**GGGTCCTTC TGGTACCACAGCC**	43.4	235	AJ715437.1

Quality Control Using Conventional PCR

To confirm appropriate use of each gBlock as a standard, we ran conventional PCR prior to qPCR. The reaction and thermocycling conditions are detailed in Table B-4 and Table B-5, respectively. All PCR reactions were prepared in a UV-irradiated, bleach-wiped laminar flow biosafety cabinet and using DNase/RNase-free materials. Agarose gels (1%) were prepared with TAE buffer, SYBR Safe DNA Gel Stain, and Blue Juice loading dye (Invitrogen). Gel electrophoresis was run for 40 minutes at 100V for each agarose gel. Gels were viewed under a UV transilluminator.

Table S.7: Conventional PCR reaction conditions for 25 µL reactions. These reaction conditions were used for all functional gene targets.

Component	Volume; Target Concentration (if applicable)
Taq Master Mix (QIAGEN)	12.5 µL
Nuclease-free water	5 µL
Forward Primer	3 µL; 1.2 µM
Reverse Primer	3 µL; 1.2 µM
BSA	0.5 µL; 0.4 µg/µL
Template	1 µL

Table S.8: Thermocycling parameters for fungal functional genes

Gene Target	Initial Denaturing	Denaturing	Annealing	Extension	Final Extension	Number of Cycles	Expected amplicon size (bp)
nirKf	95℃, 10min	95℃, 30s	54℃, 30s	72℃, 30s	72℃, 10min	30	~480bp
Cu1AF/Cu2R	94℃, 3min	94℃, 30s	48℃, 30s	72℃, 120s	72℃, 10 min	35	140-300
Cu1F/Cu2R	94℃, 3min	94℃, 30s	50℃, 30s	72℃, 2min	72℃, 10min	35	140-300

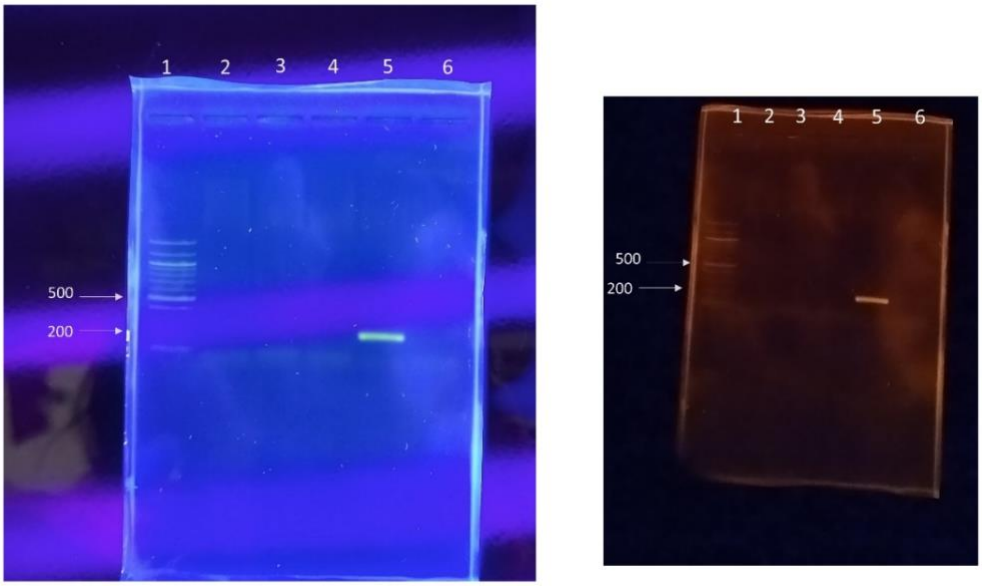

Figure S.4: Gel images for the basidiomycete-targeting laccase primers (Left, taken without filter and Right, taken with a UV imaging box with filter). Both images are of the same gel. In order, Lane 1: 100bp ladder, Lane 2: sample, Lane 3: CV sample, Lane4: sample, Lane5: positive control (laccase from T versicolor), Lane 6: negative control. Only Lane 5, the positive control, exhibited a distinct bright band. This band is at the expected size for the gBlock.

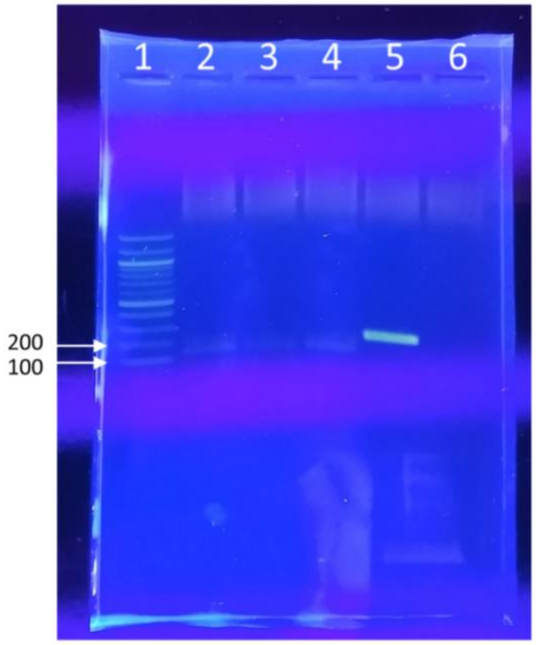

Figure S.5: Gel image for the ascomycete-targeting laccase primers. In order, Lane 1: 100bp ladder, Lanes 2-4: sample, Lane 5: positive control, Lane 6: negative control. Lane 5, the positive control, exhibited a distinct bright band at around 200bp. This band is at the expected size for the gBlock. There were faint bands between 100 and 200bp for the samples, indicating possible laccase sequences in the samples. Smears below the 100bp line were due to marks on the transilluminator cover and not from the gel.

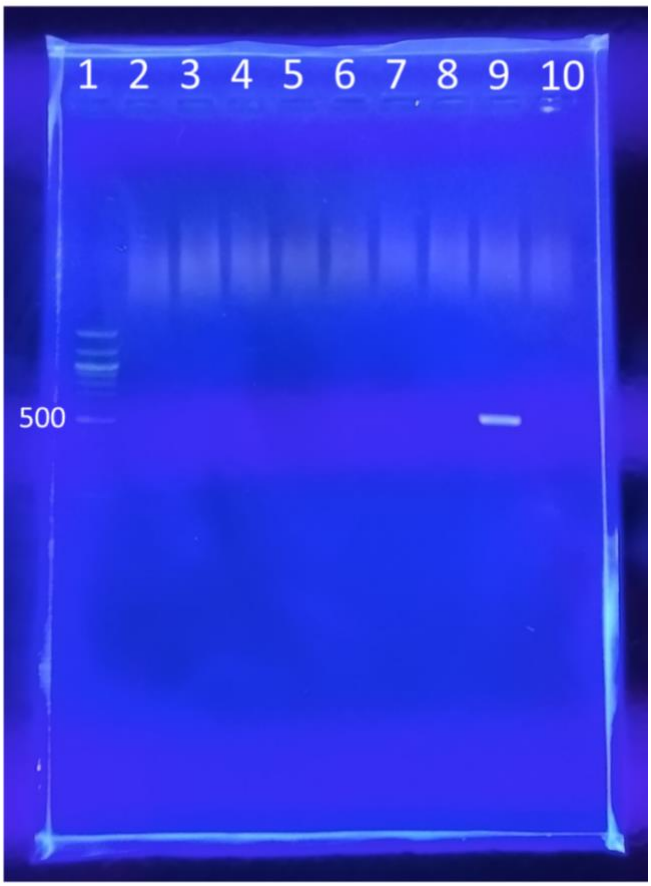

Figure S.6: Gel image for Class II peroxidase and fungal nitrite reductase primers. In order, Lane 1: 100bp ladder, Lanes 2-4: sample (peroxidase target), Lane 5: positive control (peroxidase), Lane 6: negative control (peroxidase), Lanes 7-8: sample (*nirKf* target), Lane 9: positive control (*nirKf*), Lane 10: negative control (*nirKf*). Lane 9, the positive control for *nirKf*, exhibited a distinct bright band at around 500bp. This band is at the expected size for the gBlock. There were no other distinct bands in any other lanes. Given failure of amplification of Class II peroxidase standards, that functional gene target was removed from further analysis.

qPCR Thermocycling Parameters

Table S.9: Thermocycling parameters for all functional gene targets.

Gene Target	Initial Denaturing	Denaturing	Annealing	Extension	Number of Cycles
Cu1 and Cu1A	95 ℃, 3 min	95 ℃, 30 sec	50 ℃, 30 sec	72 ℃, 2 min	40
nirKf	95 ℃, 10 min	95 ℃, 30 sec	54 ℃, 30 sec	72 ℃, 30 sec	40

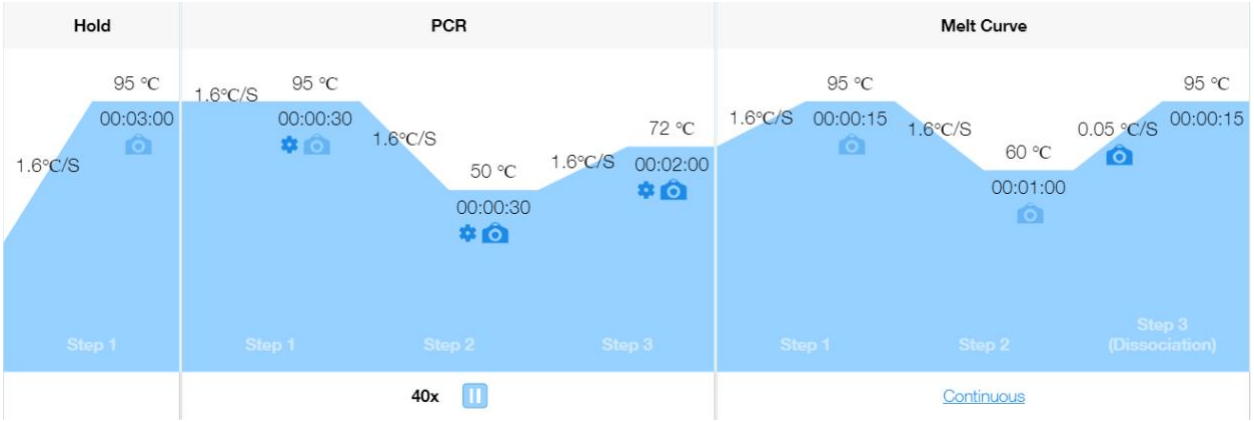

Figure S.7: Visual representation of thermocycling parameters for laccase genes

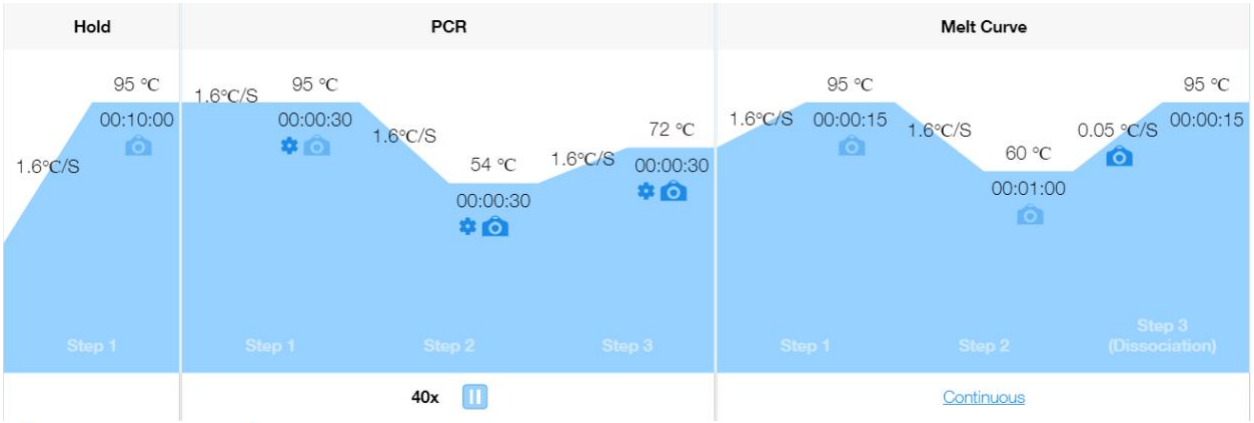

Figure S.8: Visual representation of thermocycling parameters for *nirKf* gene

qPCR Quality Control Results

Standard curves and melt curves for each target functional gene are provided in this section. No amplification was observed in no-template controls (C_q data provided in Table S.10).

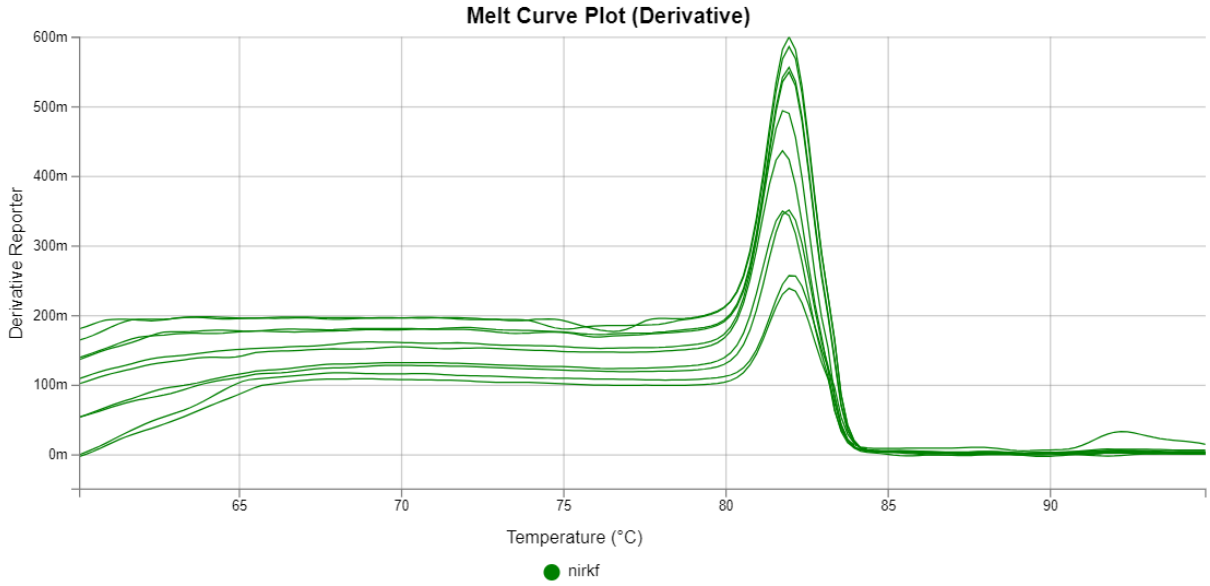

Figure S.9: Melt curve plot for gBlock standards with the *nirKf* primer pair.

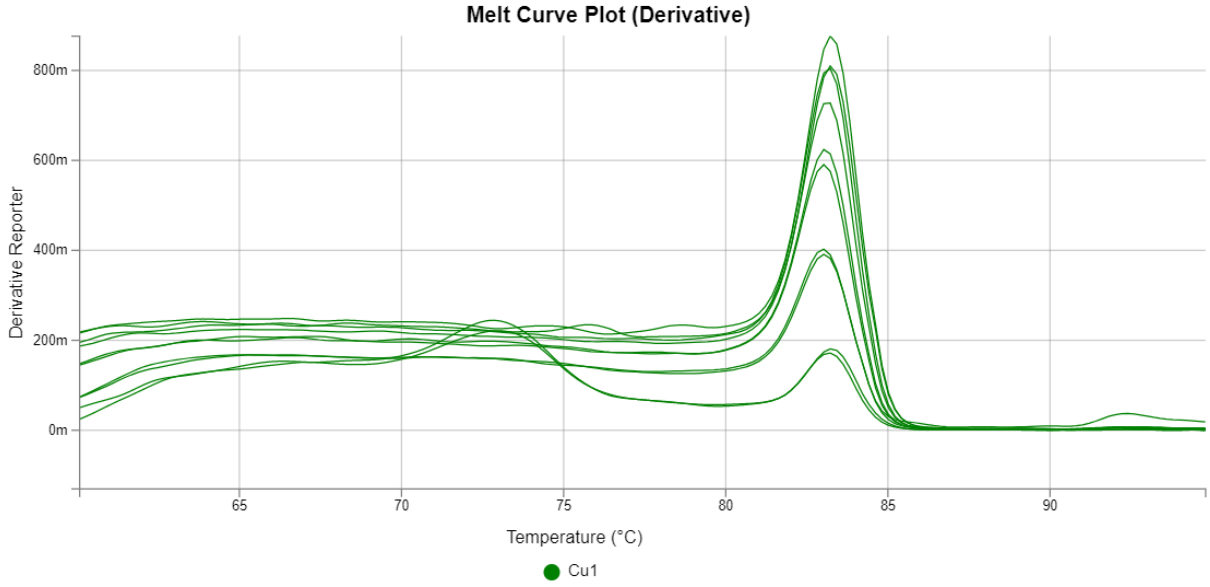

Figure S.10: Melt curve plot for gBlock standards with the *Cu1* laccase primer pair.

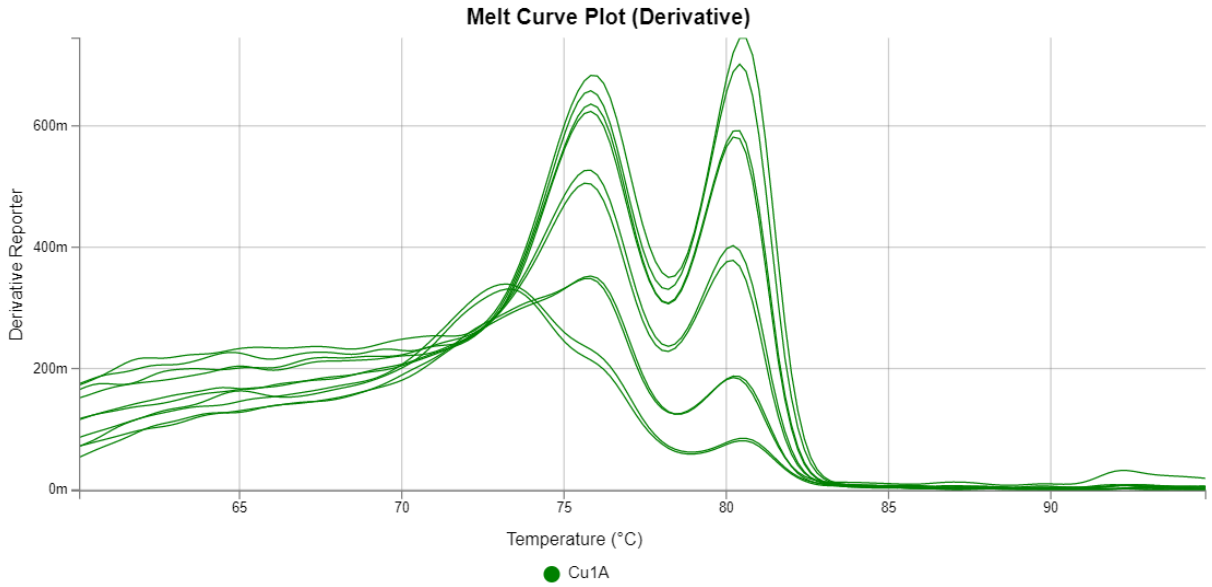

Figure S.11: Melt curve plot for gBlock standards with the *Cu1A* laccase primer pair. Though the melt curve implicates non-specific amplification with dual peaks, the initial PCR and gel electrophoresis under the same reaction and thermocycling conditions yielded a single band in the gBlock standards. Given the totality of information gathered, we deemed *Cu1A* qPCR results to still be of acceptable quality.

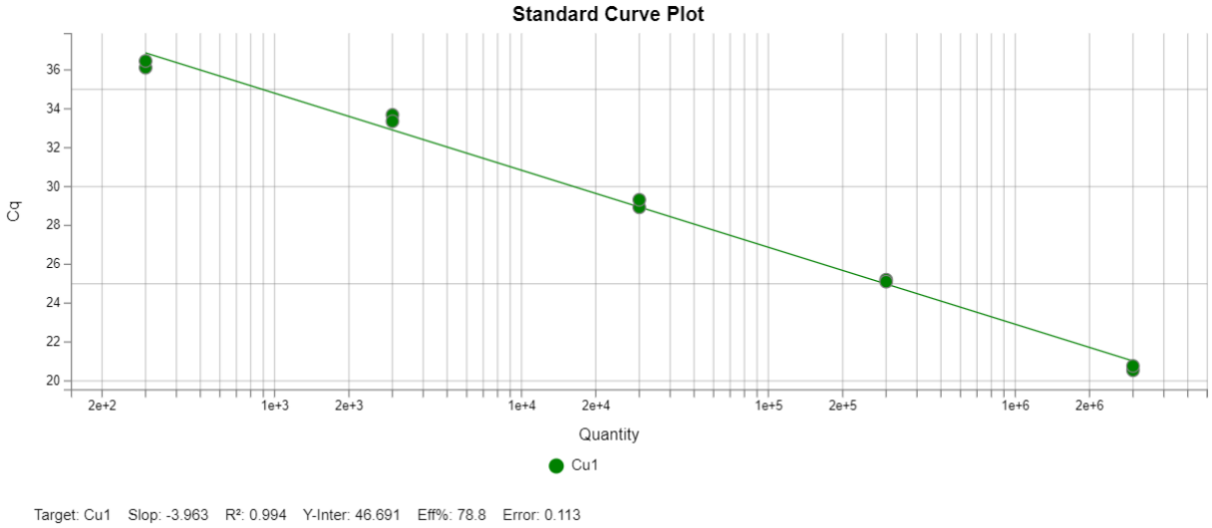

Figure S.12: Standard curve plot and accompanying slope, R^2, Y-intercept, efficiency (%), and error for the *Cu1* laccase primer pair.

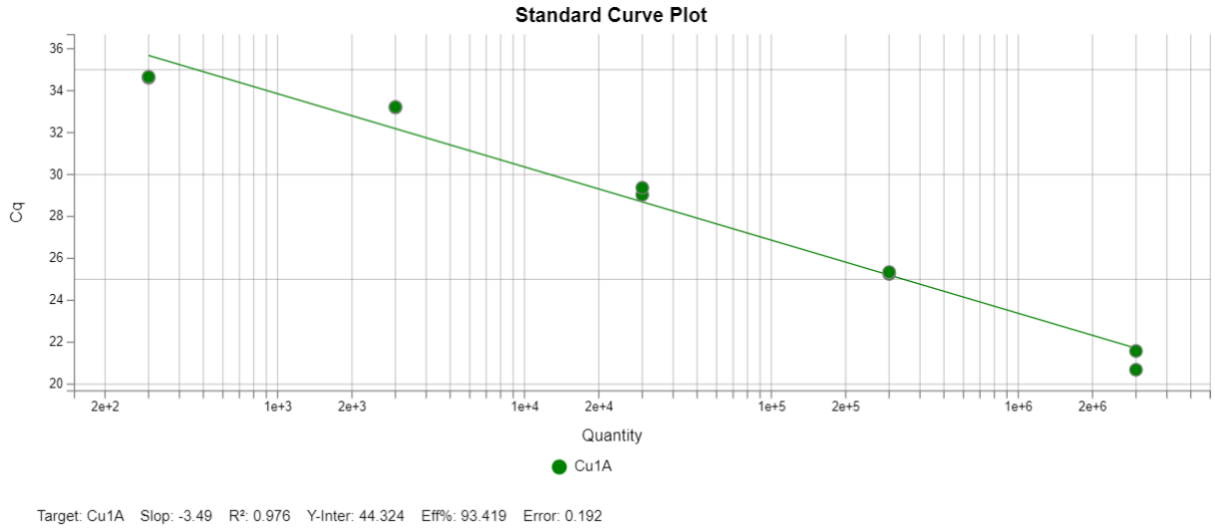

Figure S.13: Standard curve plot and accompanying slope, R^2, Y-intercept, efficiency (%), and error for the *Cu1A* laccase primer pair.

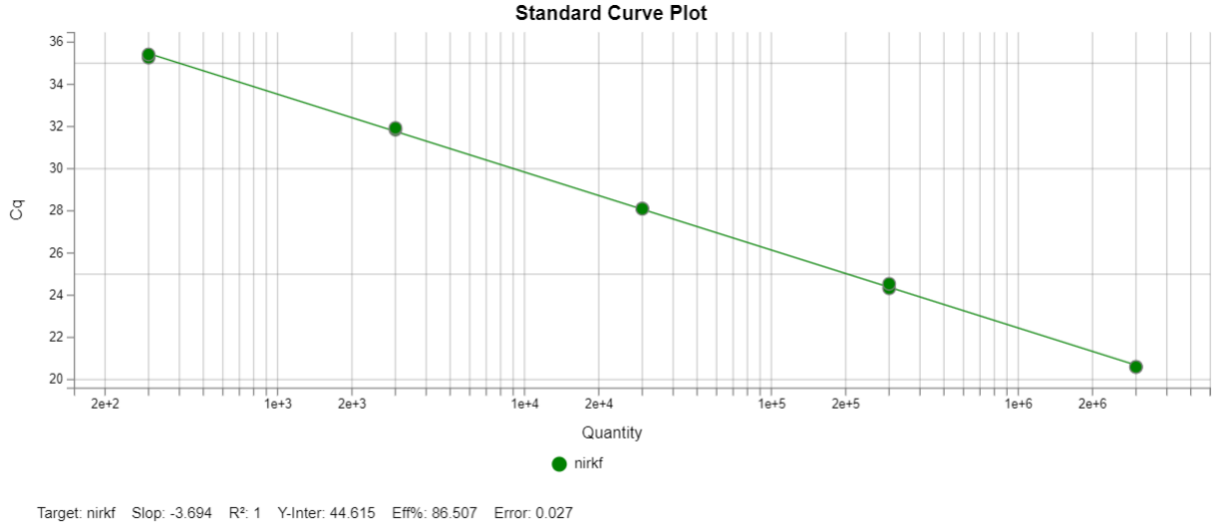

Figure S.14: Standard curve plot and accompanying slope, R^2, Y-intercept, efficiency (%), and error for the *nirKf* primer pair.

Table S.10: Average C_q values for each functional gene target.

Functional Gene Target	Average C_q for No-Template Control (n=2)
nirKf	Undetermined
Cu1	Undetermined
Cu1A	34.27

PERMDISP Analysis Results

As noted in the main text, a PERMDISP analysis is warranted when using a PERMANOVA test in situations where the sampling design is unbalanced (i.e., uneven sampling within the groups). Thus, PERMDISP was run using the betadisper function in the R 'vegan' package for each statistically significant variable from the PERMANOVA. Results are summarized below, and PCoA plots with 90% confidence intervals are plotted for the analysis (intersections of the ellipses are quick visual ways to observe if dispersions might be significantly different between groups).

Table S.11: Results from PERMDISP analyses of each explanatory variable that was statistically significant from the PERMANOVA tests.

Explanatory Variable from PERMANOVA	*p* Value from PERMDISP test
Presence of Legumes	0.009
Dominant Planting	0.21
Presence of Native Grasses	0.02
Presence of Forbs	0.12

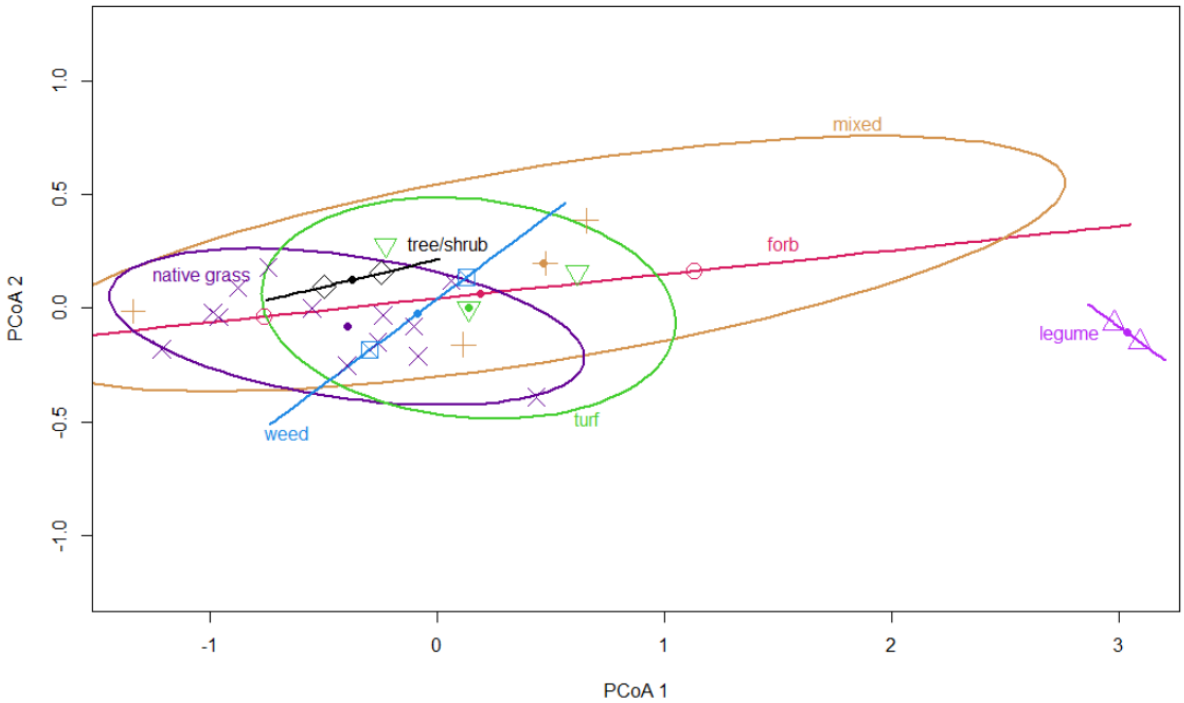

Figure S.15: Principal Coordinate Analysis (PCoA) plot for the PERMDISP analysis of the dominant planting categorical variable. Visually, apart from the legume grouping, there is overlap of the confidence intervals, which is an expected result based on the *p* value greater than 0.05. We fail to reject the null hypothesis and can safely assume that dispersions for these groupings are homogeneous.

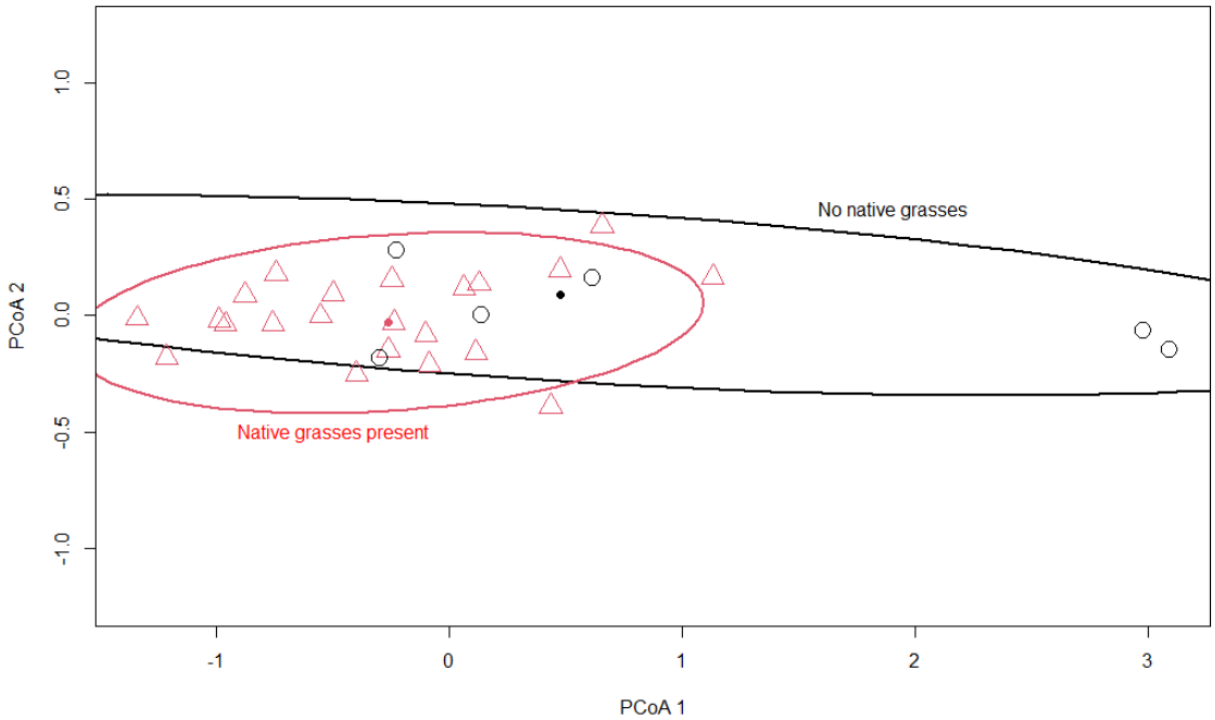

Figure S.16: PCoA plot for the PERMDISP analysis of the categorical Boolean variable for presence of native grasses. TRUE corresponds to red triangles and indicates presence of native grasses. FALSE corresponds to black circles and no presence of native grasses. Though there is overlap of the confidence intervals, the *p* value indicates a statistically significant PERMDISP result, which leads us to reject the null hypothesis that there is no significant difference in the dispersions between groups.

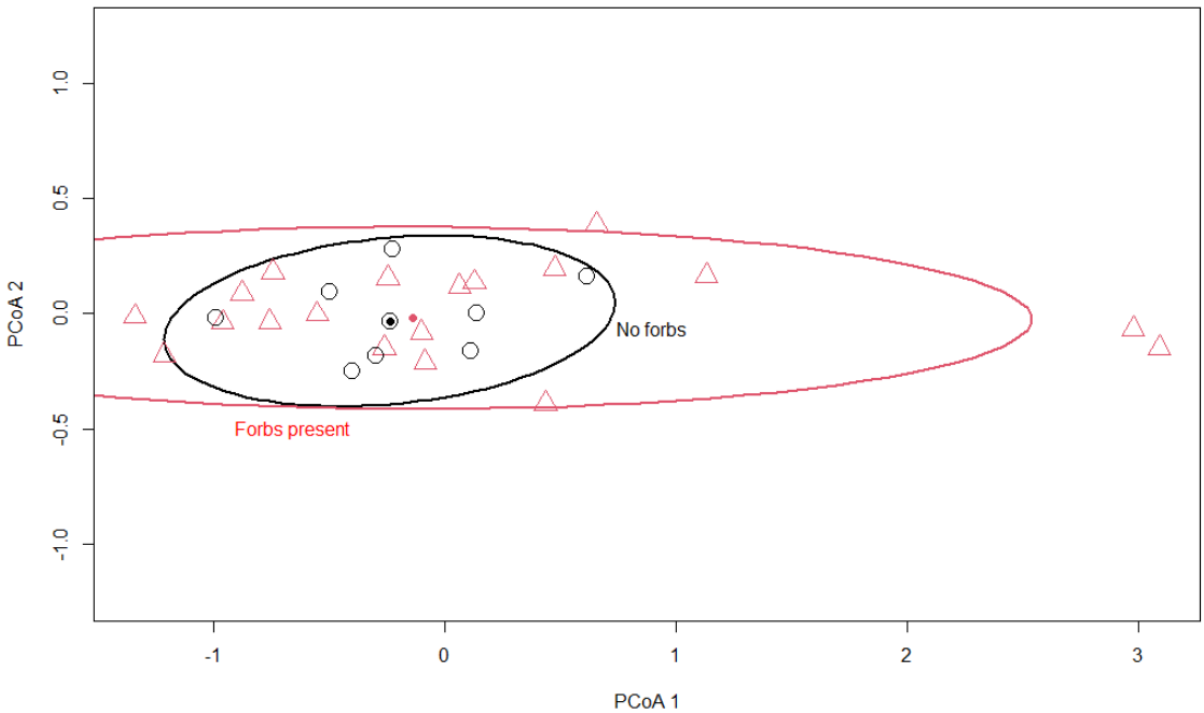

Figure S.17: PCoA plot for the PERMDISP analysis of the categorical Boolean variable for presence of forbs. Red triangles indicate presence of forbs. Black circles indicate no forbs present. Visually, there is a

slight overlap of the confidence intervals, which is an expected result based on the *p* value greater than 0.05. We fail to reject the null hypothesis and can safely assume that dispersions for these groupings are homogeneous.

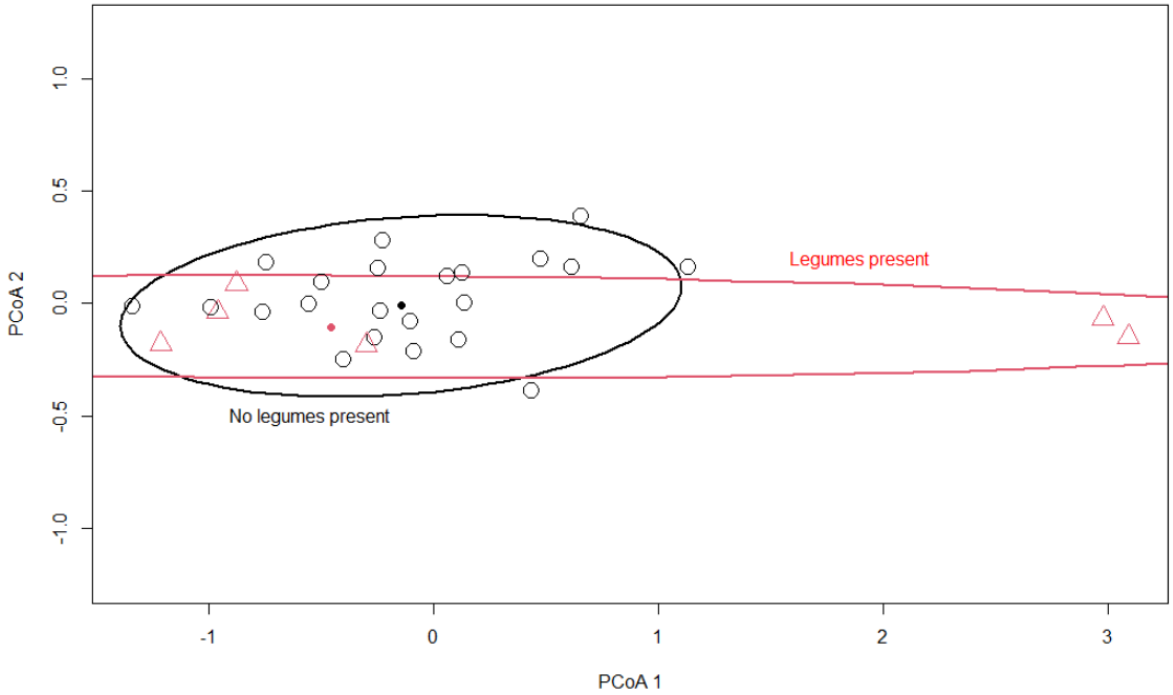

Figure S.18: PCoA plot for the PERMDISP analysis of the categorical Boolean variable for presence of legumes. TRUE corresponds to red triangles and indicates presence of legumes. FALSE corresponds to black circles and no presence of legumes. Though there is overlap of the confidence intervals, the *p* value indicates a statistically significant PERMDISP result, which leads us to reject the null hypothesis that there is no significant difference in the dispersions between groups.

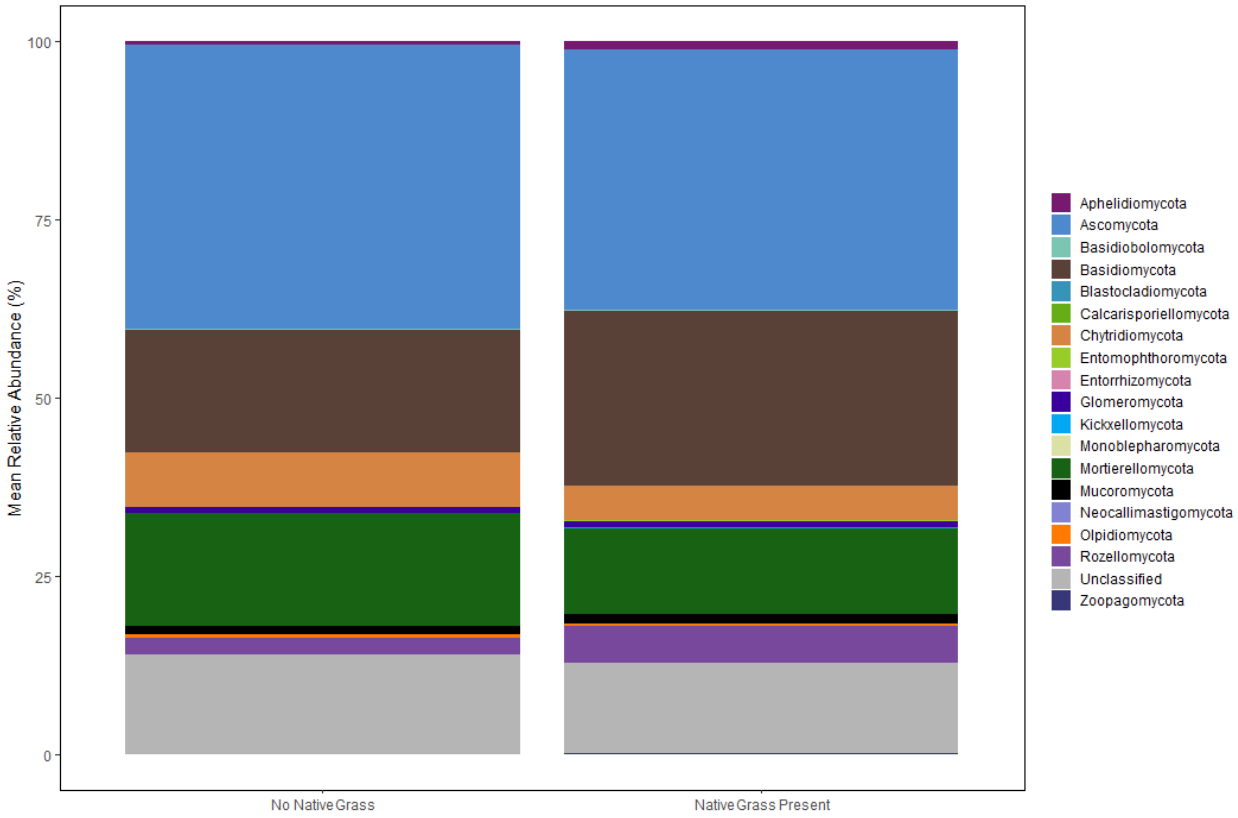

Figure S.19. Relative abundance of fungal phyla for bioretention cells, grouped by the presence or absence of native grasses in the cell. The number of sites with no native grass was six, and the number of sites with native grass present was 22.

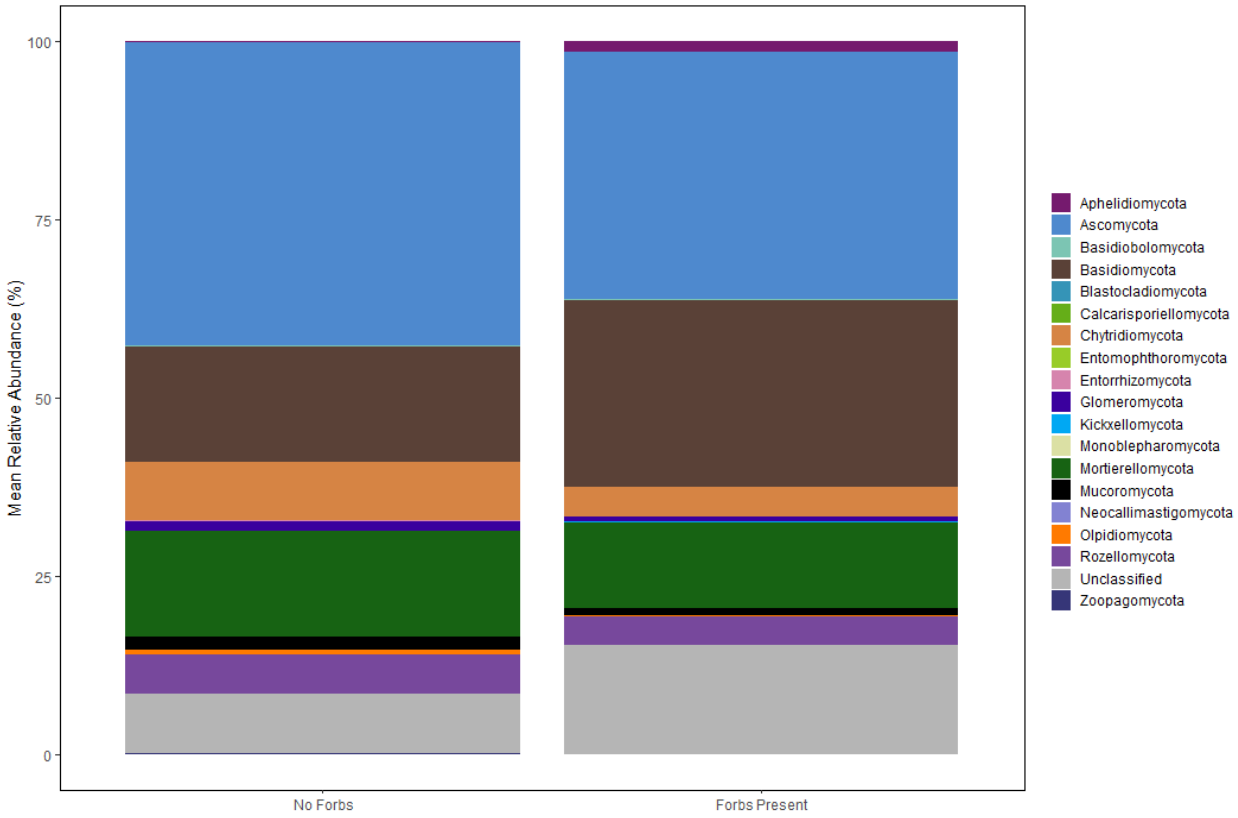

Figure S.20. Relative abundance of fungal phyla for bioretention cells, grouped by the presence or absence of forbs in the cell. The number of sites with no forbs was 9, and the number of sites with forbs present was 19.

Functional Gene Quantification Analysis Results

Table S.12: Descriptive statistics for functional gene abundance (in gene copies/g dry soil) for each functional gene tested.

Gene Target	Minimum	Mean	Maximum	Standard Error
nirKf	8.7×10^{4}	7.4×10^{5}	1.6×10^{6}	7.8×10^{4}
Cu1	3.4×10^{5}	3.2×10^{6}	8.4×10^{6}	3.5×10^{5}
Cu1A	3.2×10^{7}	4.0×10^{8}	1.1×10^{9}	5.2×10^{7}

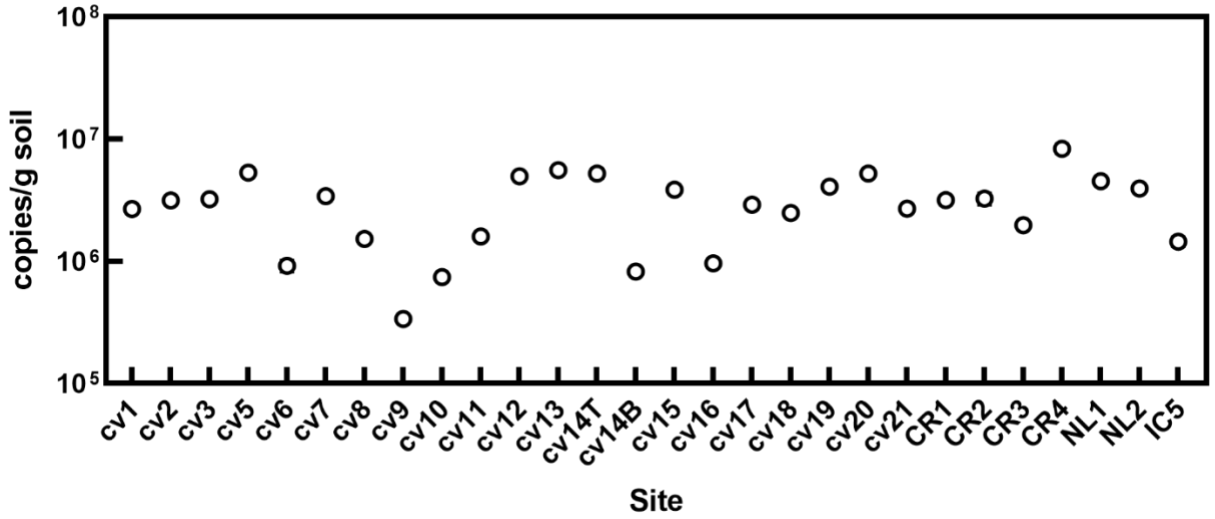

Figure S.21: Mean functional gene abundance normalized to grams of dry soil for the Cu1 laccase gene at all sampling locations. Error bars represent the standard error about the mean; some error bars fall within the markers and are not visible.

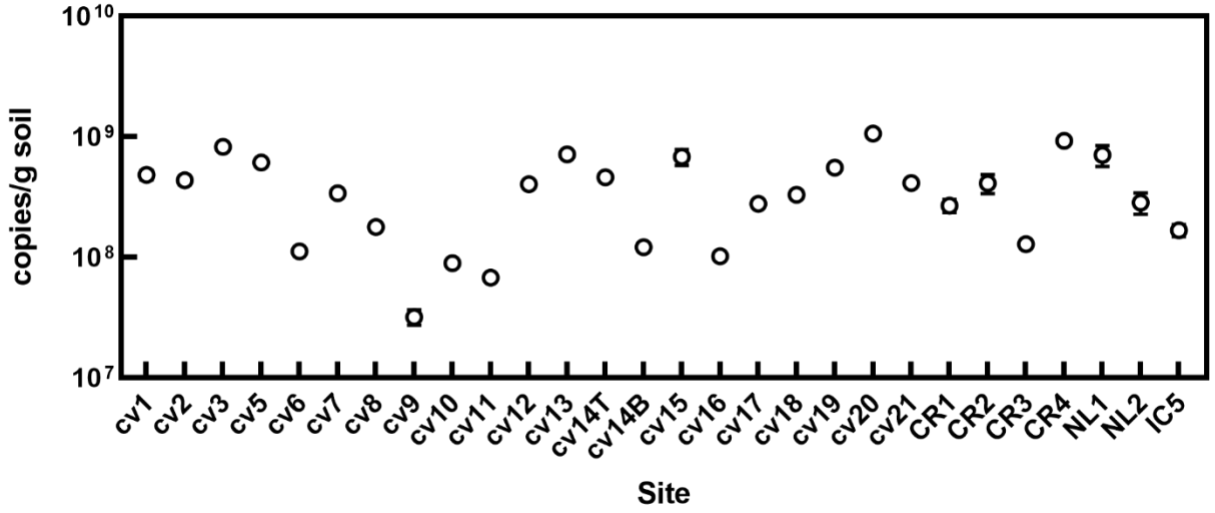

Figure S.22: Mean functional gene abundance normalized to grams of dry soil for the Cu1A laccase gene at all sampling locations. Error bars represent the standard error about the mean; some error bars fall within the markers and are not visible.

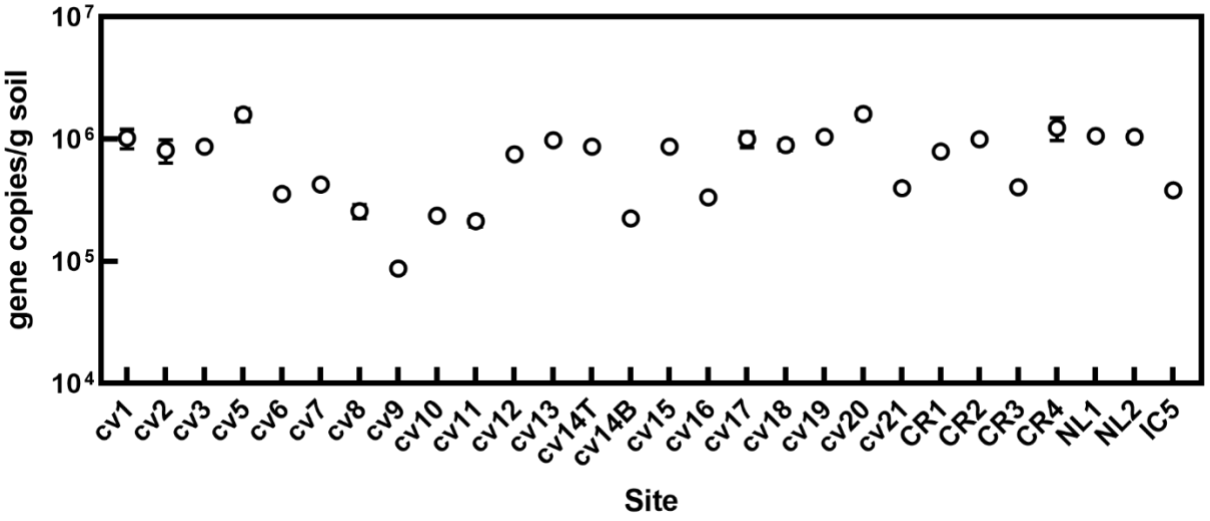

Figure S.23: Mean functional gene abundance normalized to grams of dry soil for the fungal *nirK* nitrite reductase gene at all sampling locations. Error bars represent the standard error about the mean; some error bars fall within the markers and are not visible.

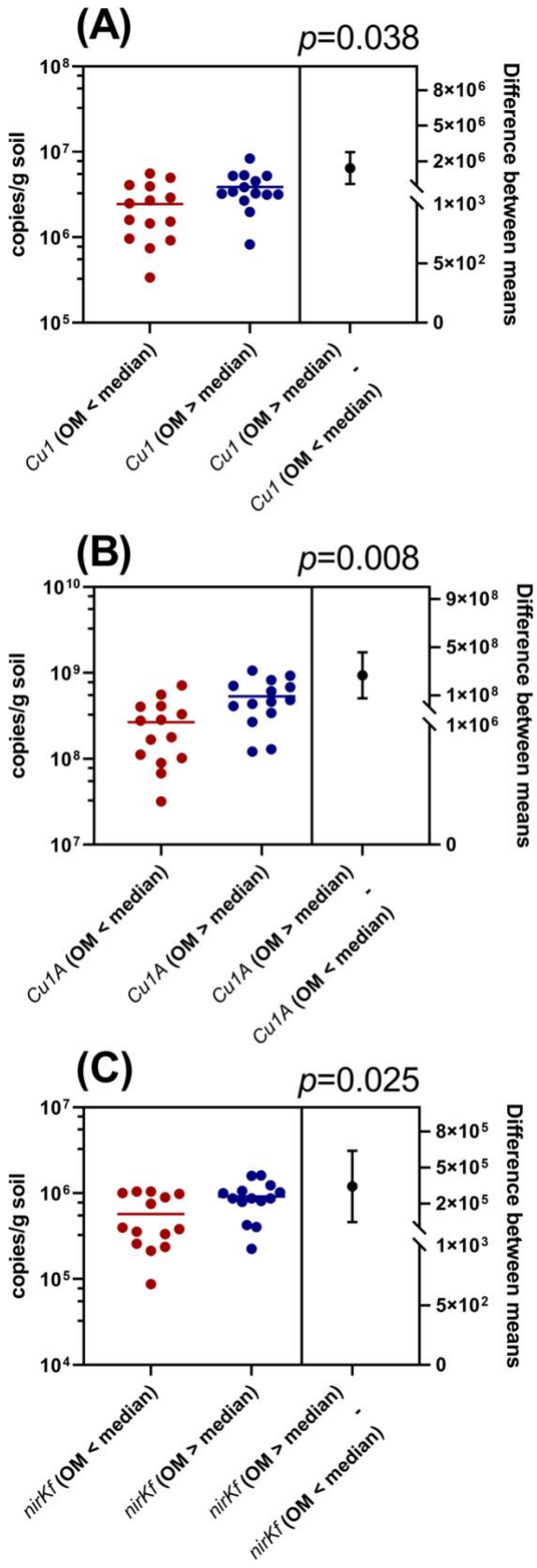

Figure S.24: Estimation plot for a two-tailed t test comparing **(A)** *Cu1*, **(B)** *Cu1A*, and **(C)** *nirKf* abundance grouped by percent OM less than the median (red circles) and greater than the median (blue circles). On the right side of the estimation plot is the 95% confidence interval for the difference between the means (*p*-values shown on respective plots). The 95% confidence interval does not include 0, indicating a statistically significant result. The distance to 0 provides orientation for the level of significance (confidence intervals further away from 0 indicate lower *p* values.)

Results for Categorical Variable Effects on Functional Gene Abundance

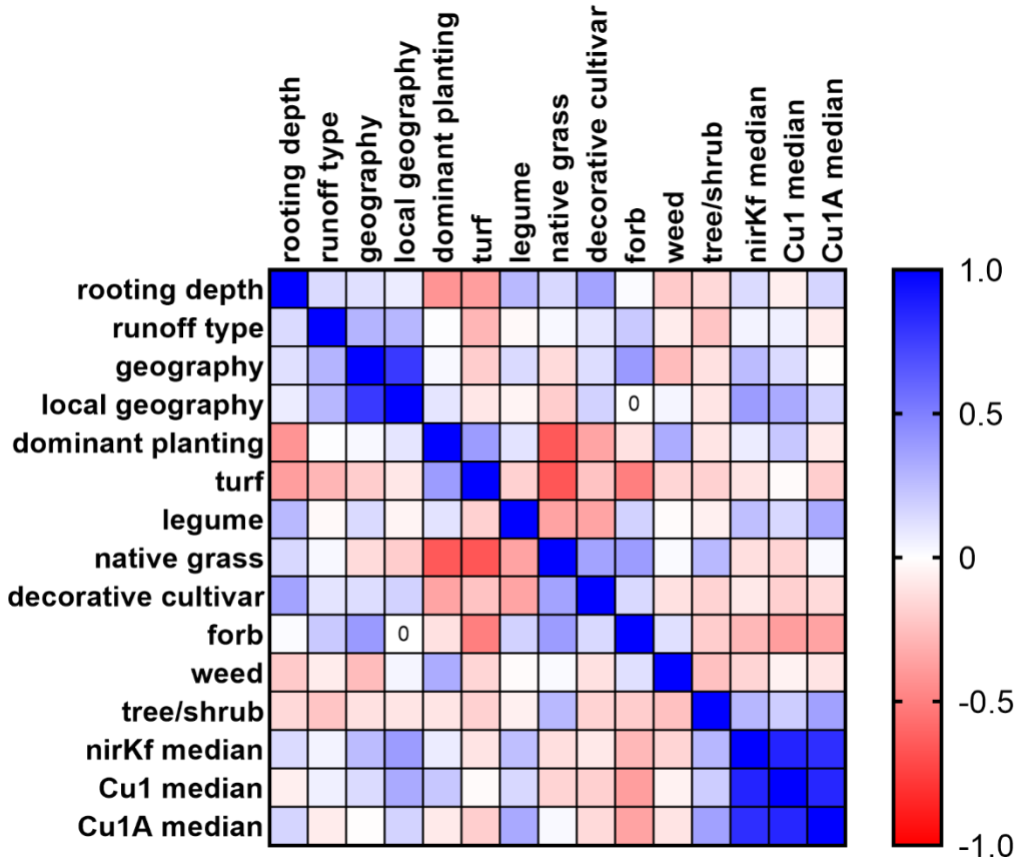

Figure S.25: Spearman's correlation chart for metadata categories (re-coded to numerical values).