Tallgrass Restoration Legacies, Summary

Beau Larkin

Last updated: 07 December, 2023

Contents

Packages

```r
packages_needed = c("tidyverse", "png", "knitr", "conflicted", "formatR")
packages_installed = packages_needed %in% rownames(installed.packages())
if (any(!packages_installed)) {
    install.packages(packages_needed[!packages_installed])
}
for (i in 1:length(packages_needed)) {
    library(packages_needed[i], character.only = T)
}
conflict_prefer("filter", "dplyr")
conflict_prefer("select", "dplyr")
```

Introduction

The tallgrass prairie biome in North America is nearly extinct, creating a tremendous desire for restoration or recreation of this critical habitat. In the upper Midwest, interested landowners have attempted to convert previous agricultural fields (corn or soybeans, usually) to prairie for decades. The consequences of

this practice for soil biota are largely unknown. Agricultural fields filter fungal and bacterial communities for species that can tolerate extreme nutrient loads, frequent tillage, and poor plant diversity. When agricultural practices cease, and native perennial plants are introduced, soil biotic communities are filtered by the new regime and rapidly change. What is the endpoint of these new communities? Do they begin to resemble communities in remnant prairie (the reference condition), or do they form novel communities, indicating a change of state in the system due to disturbance associated with agriculture? Are these communities' fates similar across distant regions, or is each context dependent?

Description

The goal is to present results and discuss whether a path forward exists. If so, we will determine the strategy that presents the best story, organization, and interpretation of these results.

Data

For this summary, I'll pull as many objects as possible from existing files to reduce the number of interspersed code chunks. A few quick new analyses will be necessary, though. Data are loaded here.

```r
sites <- read_csv(paste0(getwd(), "/clean_data/sites.csv"), show_col_types = FALSE) %>%
    mutate(field_type = factor(field_type, ordered = TRUE, levels = c("corn", "restored",
        "remnant")), yr_since = replace(yr_since, which(field_type %in% c("remnant",
        "corn")), NA)) %>%
    select(-lat, -long, -yr_restore, -yr_rank) %>%
    arrange(field_key)
```

Methods

Sites

We visited fields in southern Wisconsin and at Fermilab, just west of Chicago, Illinois. Site descriptions and photos are available in this field note. The survey followed an unbalanced complete block design. Corn, restored, and remnant fields are compared, with at least one of each field type in each block. I have called blocks "regions" so far. We collected samples and data from four regions (Fig 1, left).

Design

The design is unbalanced because there are more restored fields than corn or remnant. In all but one case, only single corn and remnant fields were available in each region. This means that we only have replication to separate field types when using the entire block design.
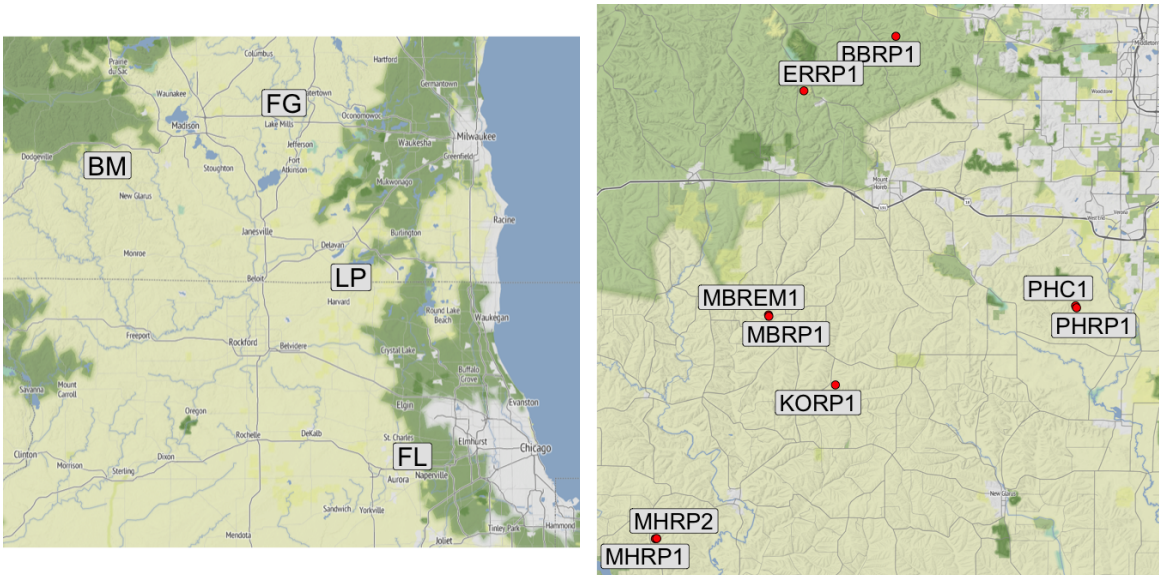
Figure 1: (Left) Labels show centroids of regions used for this work. BM = Blue Mounds, FG = Faville Grove, FL = Fermilab, LP = Lake Petite. (Right) Labels show individual fields in the Blue Mounds region.

Table 1: Count of fields by type and region: BM = Blue Mounds, FG = Faville Grove, FL = Fermilab, LP = Lake Petite

	corn	restored	remnant
BM	1	7	1
FG	1	1	1
FL	2	6	1
LP	1	2	1

A feature of our design is that restored fields vary in time since restoration, setting up the potential for a chronosequence in addition to the contrasts with corn or remnant fields. The rules for establishing a chronosequence are strict. We cannot call fields from all regions a chronosequence. With seven restored fields in a small geographic area, the Blue Mounds fields are our best bet for this (Fig 1, right), but we will likely have to call this a "pseudochronosequence" and avoid some inferences. Mantel tests (not shown) failed to find correlations between soil variables and pairwise distance, giving us a little confidence that we're avoiding systemic bias.

Soil Fungal Communities

We collected soil cores from 10 haphazardly-selected locations in each field. Genomic DNA was extracted, and the lab and bioinformatics pipeline delivered community data from ITS or 18S regions clustered as 97% similar OTUs or 99% similar SVs. In preliminary work, inferences made with SVs were weaker but not qualitatively different. I continued with OTUs only.

Originally, I had planned to average the sequence abundance per OTU in each field. In part, I did this for stastistical design reasons. Since fields are replicates, the samples in fields aren't independent. Also, explanatory and response data exist as one data point per field, meaning that they should not be expanded to apply singly to each sample. Finally, I didn't have the tools to account for this design in

permutation tests. With available tools, I could only choose one "nesting" factor, but to include samples, I'd have to account for regions and fields as blocking elements when testing the design element of field type. Luckily, I've recently found a new tool, namely function `how()` from package `permute`, that allows for more complicated permutation designs. I decided to proceed along parallel analysis paths: one with sample-level data, and one with field-level data.

Creating the sample-level data took some work. A few samples had failed to amplify, leaving some fields represented with 9 samples instead of 10. This unbalance, which is normally not a problem, became unacceptable for a couple of reasons. First, the permutations I had planned using function `how()` require balance at the plot (i.e., field) level. Also, some samples contained very few sequences compared with the others. Rarefying the entire dataset to these tiny depths would result in the loss of many admittedly low-abundance OTUs. To acheive a balanced count of samples per field and retain a higher rarefaction threshold, I used an iterative process to remove samples with few sequences, choosing a rarefaction cutoff near the plateau of OTU recovery at depth (Fig 2). This ended up being 8 samples for ITS and 7 for 18S datasets. These sample sequence values were then rarefied and used as-is for sample-level data, or, when analyzing at the field level, sample sequence values were summed and rarefied. The resulting data files recovered many more OTUs, but didn't change any major interpretation of the data. The situation with 18S data is similar, but at a much lower total sequencing depth (not shown).
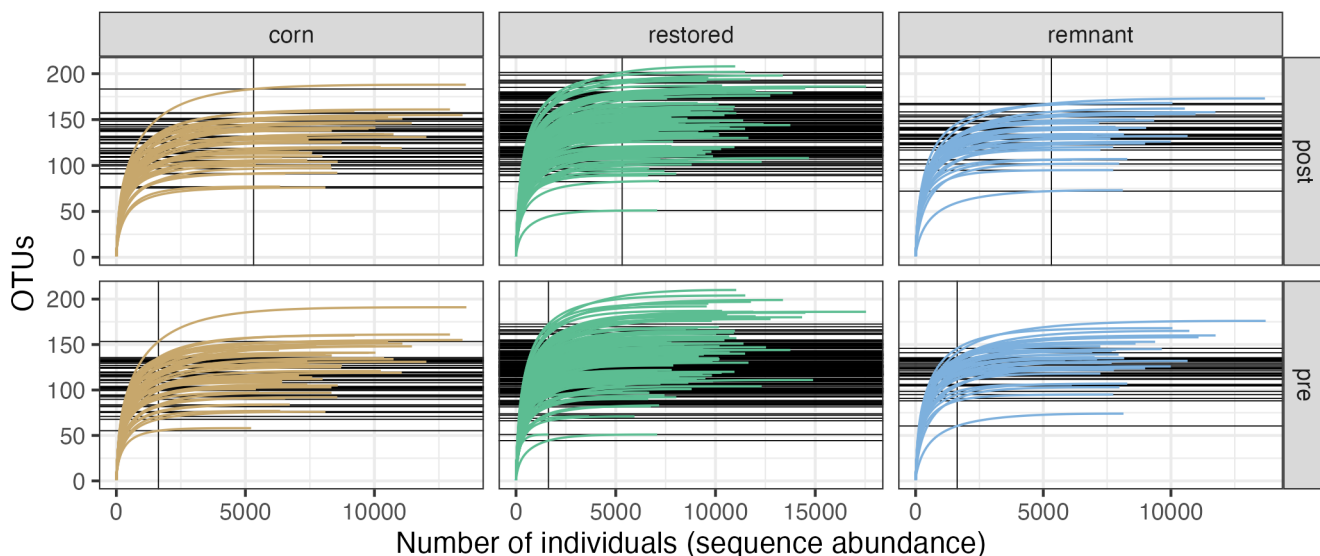

Figure 2: Individual-based rarefaction of ITS OTUs in subsamples when nine subsamples per field are retained (pre, top row) or when eight subsamples are retained (post, bottom row). Vertical lines show the minimum sequencing depth available across samples; this is the depth to which the entire set would be rarefied. Horizontal lines show the number of OTUs that would be recovered from each sample at that rarefaction depth.

Species (OTU) accumulation was performed on field-level sums of sequences. Most fields were undersampled (Fig 3), particularly remnant and restored fields (ITS) and corn and remnant fields (18S).

Plant data

Plant community data resulted from two independent surveys. In the Wisconsin regions, haphazard transects were established and 10 meter frames placed, with percent cover estimated for all species, resulting in a dataset with plant composition. In Fermi, relevé methods were used, resulting in presence/absence data only. Plant metadata were assembled from multiple sources and include plant traits and natural
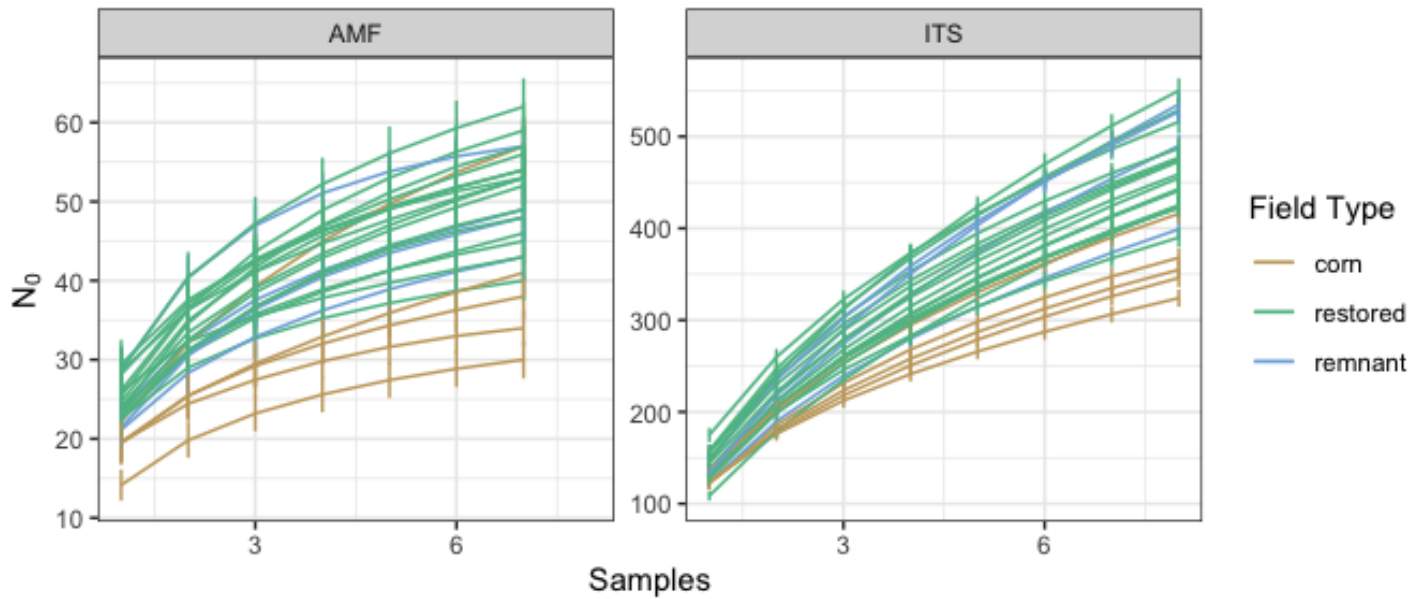
Figure 3: Species accumulations for rarefied sums of field sequences for ITS and 18S datasets. Species accumulation by the exact method; standard deviation (vertical lines) conditioned by the empirical dataset.

history details. These metadata aren't used with fields from Fermi because it doesn't make sense to think of the presence/absence of traits only. More information about the plant data, including preliminary analysis, is available in this report

Soil data

In the field, soil was pooled from 10 haphazardly-selected locations, mixed, and sampled once for soil chemical analysis. Soil data includes abiotic macro and micronutrients, organic carbon, and properties like pH. Average precipitation was determined for each field using PRISM climate data and is included with the soil data for constrained analyses. More information is available here

Design/site data

- Field type: corn, restored, remnant
- Field age: years since restoration (NA for corn and remnant)
- Region: blocks

Response data

Also taken from the pooled soil in each field, one sample was taken for analysis of Water Stable Aggregates (I don't know who did this), and one for quantification of microbial biomass.

Data sources, summary

- Fungal genomic data, ITS and 18S, OTU clusters
- Plant community data, composition in Wisconsin and presence/absence in Fermi

- Plant traits and natural history

    - Native status, photosystem, functional guild, lifecycle...

- Soil properties
- Fungal biomass (PLFA, NLFA)
- Water stable aggregates
- Site metadata and design

Results

Fungal communities

Fungal communities in cornfields stand apart from remnants and restored fields restored fields (Fig 4). As the age of restored fields increases, their distance from corn grows, and remnants lie in an intermediate position among restored fields. The high dimensionality here keeps the explanatory power of axes 1 & 2 low. Clusters based on field type were significant based on a permanova (with regions as blocks and subsamples held constant within fields, the replicate). However, a pairwise post-hoc permutation failed to separate remnant fields from any other type. This is due partially to a low number of replicates, and it's also a problem with dispersion. The remnants are as distant from each other as they are from corn. I also attempted this using field-level data (not shown), and the p values were somewhat improved, but remnants still failed to separate from corn or restored fields.

Table 2: Permutation test under reduced model. Blocks: region, samples not permuted (not independent), number of permutations: 1999

Df	SumOfSqs	R2	F	Pr(>F)
2	5.758178	0.080954	8.676357	5e-04
197	65.370815	0.919046	NA	NA
199	71.128992	1.000000	NA	NA

Table 3: Pairwise tests under reduced model. Blocks: region, samples not permuted (not independent), number of permutations: 1999, p values corrected after multiple tests using the Benjamini & Hochberg correction.

group1	group2	R2	F_value	df1	df2	p_value	p_value_adj
restored	corn	0.066	11.692	1	166	0.0015000	0.0045000
restored	remnant	0.018	2.827	1	158	0.2305000	0.2305000
corn	remnant	0.134	10.854	1	70	0.1428571	0.2142857

AMF communities (18S data) showed qualitatively similar patterns to the ITS results (Fig 5), but with weaker support.
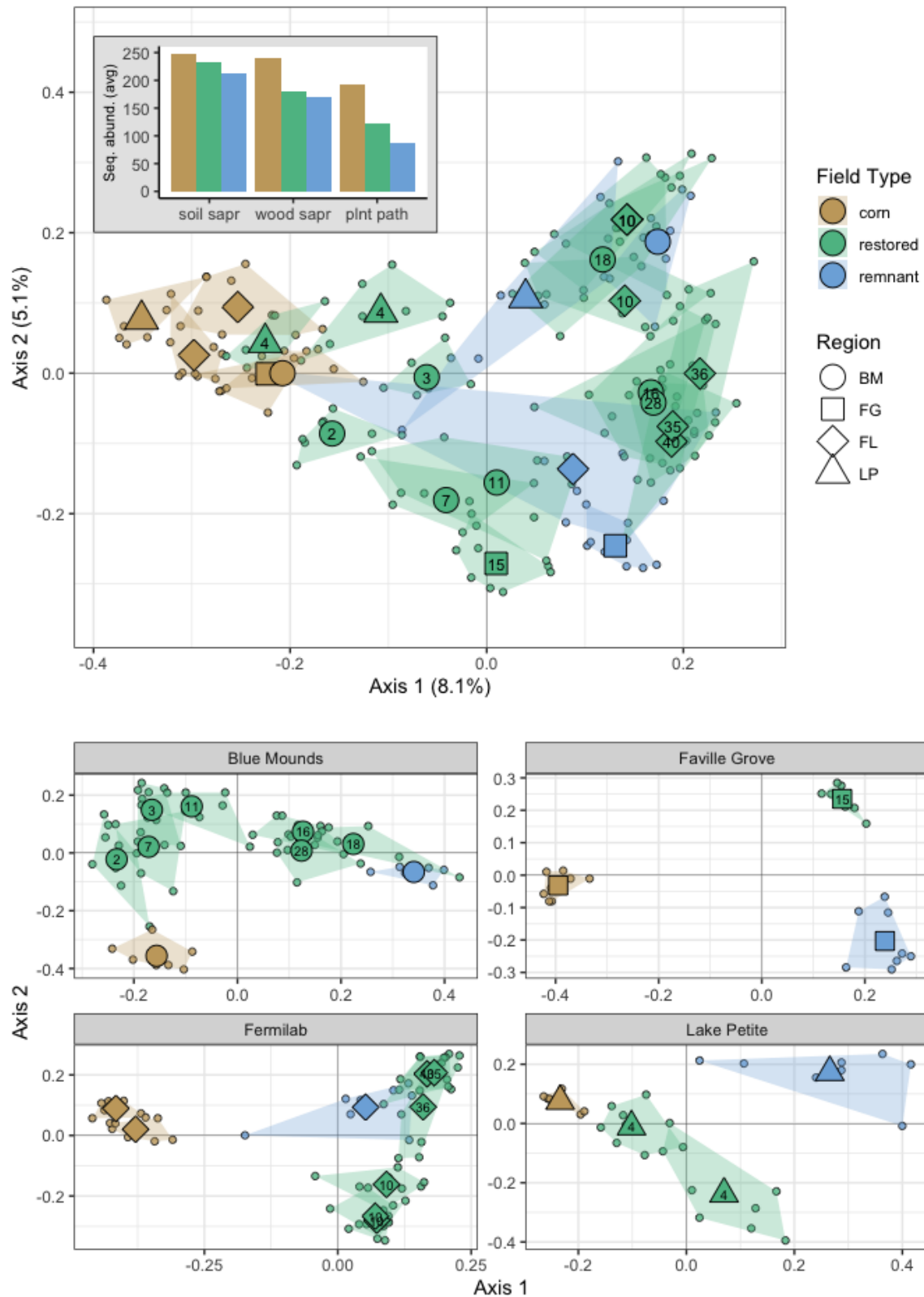
Figure 4: The top panel is a db-PCoA ordination of samples in fields, ITS dataset. Small circular icons show locations of samples and hulls enclose fields. Each field is represented with a large icon on the field centroid, where icon shapes identify the field region and color identifies the field type. Text on the restored field icons shows years since restoration. The inset column chart displays average sequence abundance across all fields for the three most abundant fungal guilds. The lower four panels show independent db-PCoAs for each region.
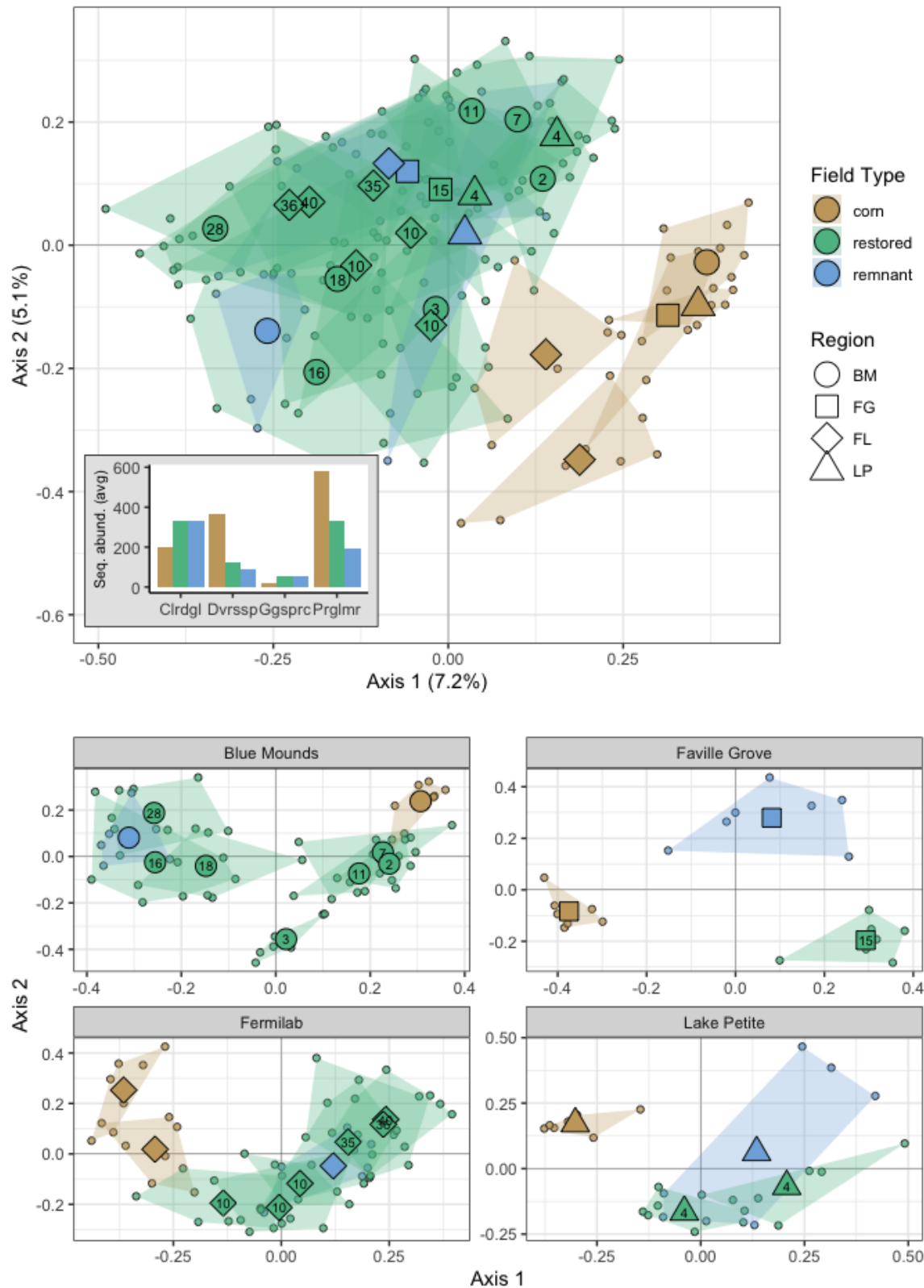
Figure 5: The top panel is a db-PCoA ordination of samples in fields, 18S dataset. Small circular icons show locations of samples and hulls enclose fields. Each field is represented with a large icon on the field centroid, where icon shapes identify the field region and color identifies the field type. Text on the restored field icons shows years since restoration. The inset column chart displays average sequence abundance across all fields for the three most abundant fungal guilds. The lower four panels show independent db-PCoAs for each region.

Table 4: Permutation test under reduced model. Blocks: region, samples not permuted (not independent), number of permutations: 1999

Df	SumOfSqs	R2	F	Pr(>F)
2	5.758178	0.080954	8.676357	5e-04
197	65.370815	0.919046	NA	NA
199	71.128992	1.000000	NA	NA

Table 5: Pairwise tests under reduced model. Blocks: region, samples not permuted (not independent), number of permutations: 1999, p values corrected after multiple tests using the Benjamini & Hochberg correction.

group1	group2	R2	F_value	df1	df2	p_value	p_value_adj
restored	corn	0.066	11.692	1	166	0.0015000	0.0045000
restored	remnant	0.018	2.827	1	158	0.2305000	0.2305000
corn	remnant	0.134	10.854	1	70	0.1428571	0.2142857

Fungal diversity

Fungal guilds

By sequence abundance across the entire dataset, the top guilds are:

1. Unidentified (not shown on column charts): 92,534 sequences
2. plant pathogens: 25,471 sequences
3. soil saprotrophs: 24,538 sequences
4. wood saprotrophs: 8,353 sequences
5. dung saprotrophs: 6,543 sequences
6. litter saprotrophs: 5,574 sequences

Compared with the sequence abundance in the unidentified group, plant pathogens and soil saprotrophs are abundant enough to feel somewhat confident about in terms of coverage. At the level of taxonomic order, composition of soil saprotrophs in corn fields looks different from restored and remnant fields (Fig 6, left). Among well-represented orders, Agarics increase from corn to remnant, Cystofilobasidiales are prevalent only in corn, and Filobasidiales, often human pathogens, decline from corn to remnant. Geoglossales, which are associated with woodlands, moss, grasslands, probably don't like tillage or lack of substrate in cornfields and aren't found there. The order Hypocreales only occurs in restored and remnant fields. OTUs identified as Mortierellales are abundant in corn and restored fields, less so in remnants, with different OTUs associated with different field types.

OTUs from these guilds were also heavily represented in an analysis to find indicator species for each field type. Note that the indicator species analysis is flawed because of the unbalanced number of fields for each field type. It's easier for a species to be detected in five cornfields than in sixteen restored fields, for example. I'd argue that as a basic filter to find interesting species, this analysis is acceptable if we don't try to call these species "indicators" of a particular habitat.

I filtered indicator species to those which were identified to at least the genus level. Saprotrophs associated with cornfields tended to be disturbance-tolerant extremophiles, in particular, tolerating very cold environments. Some are known to produce suites of extracellular enzymes, exploting various nutrient sources (*Tausonia pullulans*) and making recalcitrant nutrients available to plants (*Mortierella minutissima*, which is used as an agricultural inoculant). *Pseudaleuria sp.* was also associated with cornfields. Species in this genus are suspected to increase crop yields and protect plants from disease. The only saprotroph associated with restored fields was another species in the genus *Mortierella*. Remnants hosted few indicator species. OTUs identified as *Gymnostellatospora sp.* and *Striaticonidium cinctum* appear to have little natural history information. An OTU in the genus *Clavaria* is a coral fungus, associated with woodland, savanna, or old-field habitats worldwide.

The contrast among field types with plant pathogens isn't obvious (Fig 6, right). Not surprisingly, most pathogenic taxa associated with cornfields are known agricultural pathogens, including *Corynespora cassiicola*, *Curvularia sp.*, *Plectosphaerella cucumerina*, and *Setophoma terrestris*. *Microdochium colombiense*, a snow mold of rye and other cereal grains, is also an extreme cold tolerant species. In restored fields, an indicator pathogen was a *Pseudopithomyces sp.*, which is often found in anthropogenic environments and is a known wheat pathogen.

Natural history details about indicator fungi could make this paper more compelling, but the best value for such an analysis here is to show whether restored fields look more like cornfields, more like remnants, or more like something else entirely. But only 36% of these OTUs were assigned to a guild and 46% were assigned to a genus. It's likely that many saprotrophs and pathogens exist in these data that are opaque to our analysis. We'd have to assume that these unassigned taxa are distributed equally across field types, but this assumption is likely invalid because agricultural systems have received much more attention. Recovery of crop-associated fungi in restored fields probably reflects this bias as much as it shows a similarity with cornfield fungi. With few taxa identified in remnant fields, there's not much to compare with restored fields. Finally, very abundant OTUs tended to be commonly distributed, but all of these indicator OTUs occur in low abundance. Are they relevant to ecological processes in these fields? It's hard to say.

If natural history details are included, I think the most important case studies will come from these taxa:

1. Mortierella
    1. Cornfields mostly, lots known about them, interesting
2. Stagnospora
    1. Aligns restored and remnant
3. Pseudopithomyces
    1. Aligns corn and restored
4. Clavaria
    1. Aligns restored and remnant
5. Microscypha/DSE
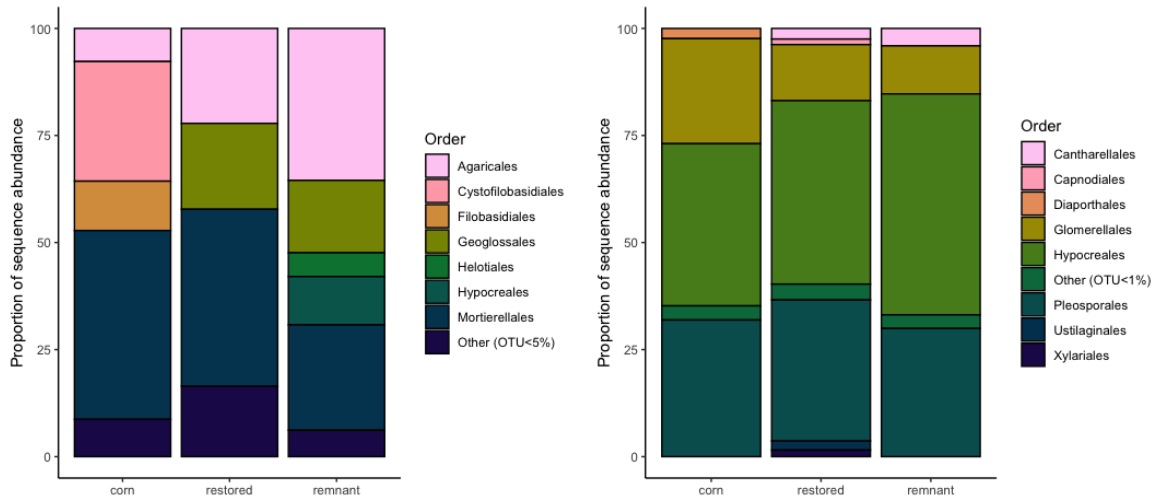   1. proposed mycorrhizae, in all remnants and 4 restored

Figure 6: Composition of soil saprotroph (left) and plant pathogen (right) orders are shown across field types. Unidentified taxa were not included, and low abundance OTUs were filtered, see legends for the cutoff percentage.
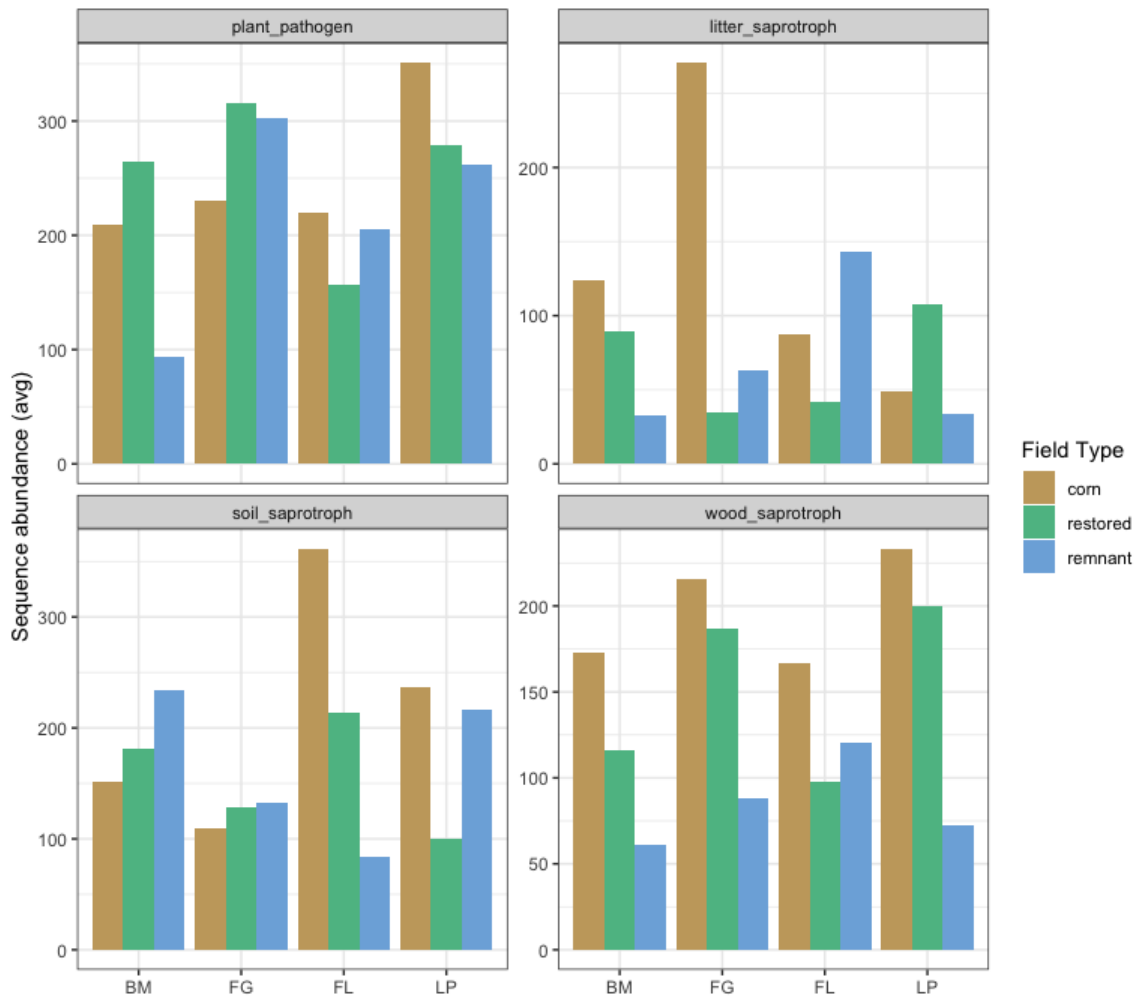

Figure 7: Abundant or dynamic guilds across field types in regions, ITS dataset.

AMF diversity

AMF taxa

Responses

WSA Biomass