Potato Soil Health

Sudha GC Upadhaya

2023-01-21

```
library(ggplot2)
library(raster)
library(agricolae)
library(tidyr)
library(ggpubr)
library(dplyr)
library(tidyverse)
library(data.table)
library(vegan)
library(lme4)
library(lmerTest)
```

Plot sampling locations on Map

```
df = read_csv(file = 'C:/Users/theox/Desktop/Projects/SoilFingerPrints/Data/SamplingLocations/SamplingL
df = df[c('ID', 'lat', 'long', 'FieldType')]
head(df)
```

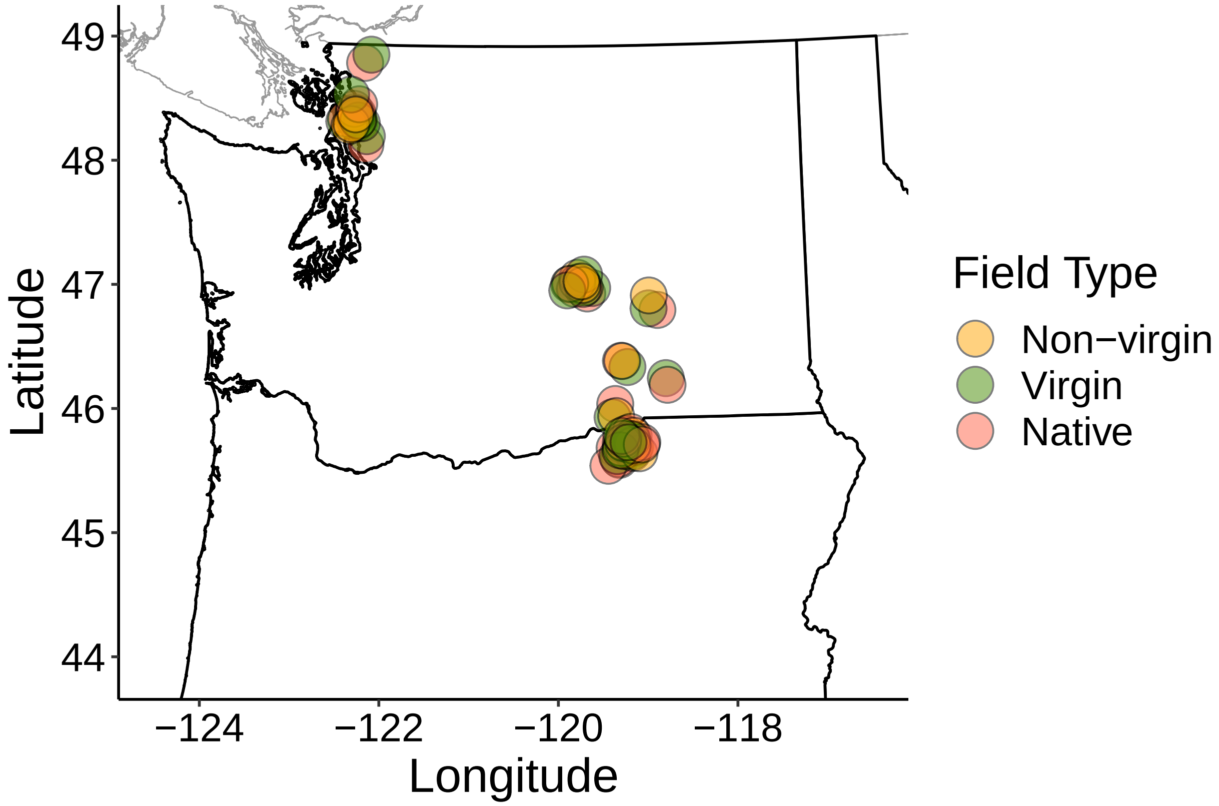
```
## # A tibble: 6 x 4
##      ID   lat  long FieldType
##   <dbl> <dbl> <dbl> <chr>
## 1     1  46.3 -119. Non-virgin
## 2     2  46.3 -119. Virgin
## 3     4  47.1 -120. Non-virgin
## 4     5  47.1 -120. Virgin
## 5     6  47.0 -120. Virgin
## 6     7  47.0 -120. Non-virgin
```

```
df$FieldType = gsub('Field', "", df$FieldType)
df$FieldType = gsub('field', "", df$FieldType)
df$FieldType = gsub('Non-virgin ', "Non-virgin", df$FieldType)
df$FieldType = gsub('Virgin ', "Virgin", df$FieldType)
unique(df$FieldType)
```

```
## [1] "Non-virgin" "Virgin"     "Native"
```

```r
state_list    <- c('California', 'Nevada', 'Utah', 'Wyoming',
                   'Montana', 'Idaho', 'Oregon', 'Washington')
province_list <- c("British Columbia", "Alberta")
us <- raster::getData("GADM",country="USA",level=1)
canada <- raster::getData("GADM",country="CAN",level=1)
us.states <- us[us$NAME_1 %in% state_list,]
ca.provinces <- canada[canada$NAME_1 %in% province_list,]
```

```r
options(repr.plot.width=10, repr.plot.height=8)
#jpeg('Map_soilprints.jpg',width=10,height=8,units='in',res=300)
Map = ggplot(us.states,aes(x=long,y=lat, group = group))+
  geom_path(size = 0.25)+
  geom_path(data=ca.provinces, size = 0.25, color = 'grey60')+
  geom_path(data=subset(us.states, NAME_1 %in% c('Montana', 'California','Nevada','Utah','Wyoming')),
            size=0.25,color="grey60")+
  geom_polygon(data=subset(us.states, NAME_1 %in% c('Idaho', 'Oregon', 'Washington')),
               aes(x=long, y=lat, group=group), fill="white", colour = "black", size=0.50)+
  theme(panel.background = element_rect(fill = "white", colour = "grey50"))+
  coord_map("bonne", lat0 = 50,
            xlim = c(-124.5, -116.5),ylim = c(44, 49))+
  xlab("Longitude") +
  ylab("Latitude")+
  geom_point(data = df, aes(x= long, y = lat,color = FieldType, fill = FieldType), position=position_ji
  scale_fill_manual(name = 'Field Type', labels = c('Non-virgin', 'Virgin', 'Native'), values = c('orang
  scale_colour_manual(name = 'Field Type', labels = c('Non-virgin', 'Virgin', 'Native'),values = c('bla
  guides(color = guide_legend('Field Type'))+
  theme_classic()+
  theme(axis.title.x= element_text(size =18, color ='black'),
        axis.title.y = element_text(size =18, color ='black'),
        axis.text.x = element_text(size =15,color = 'black'),
        axis.text.y = element_text(size =15, color = 'black'),
        legend.text = element_text(size =15),
        legend.title = element_text(size =17),
  )
Map
```



```
#dev.off()
```

2021 data visulatization and inference

```
df_21 = read.csv('C:/Users/theox/Desktop/Projects/SoilFingerPrints/Data/2021/SoilFingerPrintsData_21.cs
df_21[c("ID", "Pair", "FieldType","Block")] <- lapply(df_21[c("ID", "Pair", "FieldType", "Block")],
                                                       factor)
df_21 = df_21[, c(1, 3, 4:18, 20, 22, 23,24,26:32, 34)]
#lapply(df, unique)
head(df_21)
```

```
##   ID        Location Pair        FieldType Block Wt_Tuber.g. Weight_GT20
## 1  1 Columbia Basin    1 Non-virgin field     2     1586.65     1537.88
## 2  2 Columbia Basin    1     Virgin field     2      207.26      160.03
## 3  4 Columbia Basin    2 Non-virgin field     1     1096.27     1077.21
## 4  5 Columbia Basin    2     Virgin field     1      610.05      610.50
## 5  6 Columbia Basin    3     Virgin field     5      814.12      801.73
## 6  7 Columbia Basin    3 Non-virgin field     5     1038.14     1009.35
##   Count_Tuber Count_GT20 Sen_7.29.21 Sen_8.4.21 Sen_8.11.21 Sen_8.18.21
## 1          15         12          10          8          10          15
## 2          11          2          20         25          20          30
## 3           9          8           5         10          15          15
## 4           5          5           0          0           0           0
## 5          12         11           0          0           0           5
## 6          14         11           0          0           2          10
##   Sen_8.25.21 Sen_9.1.21 Sen_9.8.21 Sen_9.23.21 BD_Per_Tuber Vd_Per_Tuber
## 1          10         15         30          40            0            0
## 2          30         40         40          50            0            0
## 3           8         15         35          70            0            0
## 4           0          0          2           5            0            0
## 5           5         10          5           5            0            0
## 6           8         10         10          10            0           10
##   BD_Stem Vd_Stem Pyth_Avg_Soil ResPyth_Avg_Soil ResPyth_Per_Soil
## 1       0       0            11                0                0
## 2       0       0            56               45               80
## 3       0       0           109               22               20
## 4       0       0            60               33               55
## 5       0       0           157               29               19
## 6       0       0             0                0                0
##   ResPyth_binary Fus_Avg_Soil Vd_Tot_Soil BD_Tot_Soil     POXC
## 1              0          110           0           0 177.7824
## 2              1          613           0           0 281.7720
## 3              1          709           0           0 364.1400
## 4              1          110           0           0 189.1080
## 5              1         1668           2           0 315.7488
## 6              0          650           0           0 267.3576
```

Microplot study

Senescence

```
library('agricolae')
ts = gsub("(?:[^.]+\\.){1}([^.]+).*", "\\1",
    names(df_21)[(grepl("Sen", names(df_21)))])
ts = as.numeric(ts)
ts[2:5] = ts[2:5] +31
ts[6:8] = ts[6:8]+60#= ts[2]+30
ts = ts-28
ts
```

AUSPC calcuation

```
## [1]  1  7 14 21 28 33 40 55
```

```
names(df_21)[(grepl("Sen", names(df_21)))] <- ts
df_21$AUDPC <- audpc(df_21[,9:16],ts)
head(df_21$AUDPC)
```

```
## [1]  767.0 1551.0  797.0   30.0  240.5  340.0
```

```
names(df_21)
```

```
##  [1] "ID"               "Location"         "Pair"             "FieldType"
##  [5] "Block"            "Wt_Tuber.g."      "Weight_GT20"      "Count_Tuber"
##  [9] "Count_GT20"       "1"                "7"                "14"
## [13] "21"               "28"               "33"               "40"
## [17] "55"               "BD_Per_Tuber"     "Vd_Per_Tuber"     "BD_Stem"
## [21] "Vd_Stem"          "Pyth_Avg_Soil"    "ResPyth_Avg_Soil" "ResPyth_Per_Soil"
## [25] "ResPyth_binary"   "Fus_Avg_Soil"     "Vd_Tot_Soil"      "BD_Tot_Soil"
## [29] "POXC"             "AUDPC"
```

```
options(repr.plot.width = 5, repr.plot.height = 3)
#jpeg('Sen_21.jpg',width=5,height=3,units='in',res=300)
c1= ggplot(data = df_21, aes(x =FieldType, y = AUDPC, color = FieldType))+
    geom_boxplot(width = 0.4)+
    geom_jitter(alpha=0.8, pch =20,size =5, position=position_jitter(width=0.1))+
    theme_bw()+
    labs(x = 'Field Type', y="Senescence (AUSPC)") +
    theme(legend.position = "none")+
    theme(axis.title.x= element_blank(),
        axis.title.y = element_text(size =10, color ='black'),
        axis.text.x = element_text(size =10, color ='black'),
        axis.text.y = element_text(size =8, color = 'black'))+
scale_color_manual(name = 'Field Type', labels = c('Non-virgin', 'Virgin'), values = c('orange', 'chart
  facet_wrap(~ Location)
#c1
#dev.off()
```

Plot Senescence

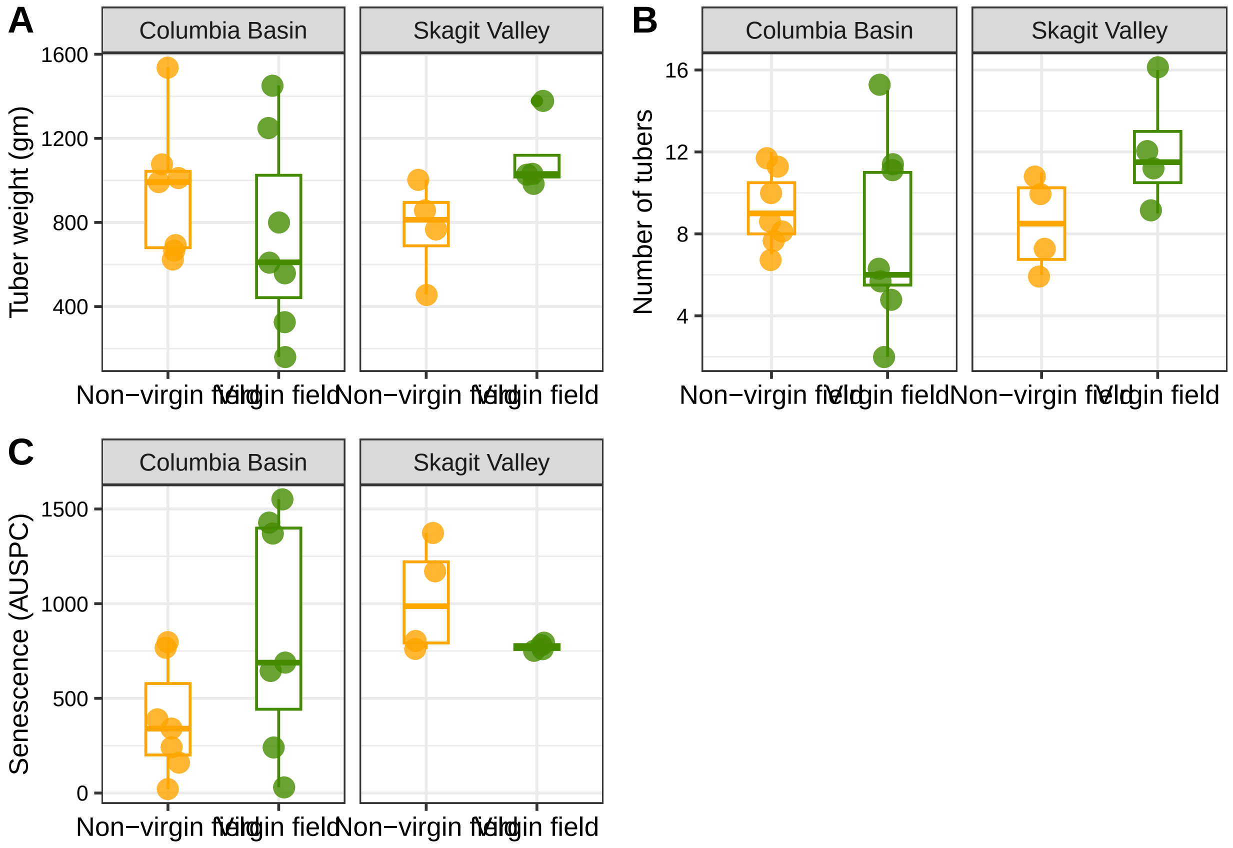
Tuber weight (more than 20 gm)

```
options(repr.plot.width = 5, repr.plot.height = 3)
#jpeg('TuberwtAll_soilprints.jpg',width=5,height=3,units='in',res=300)
a1= ggplot(data = df_21, aes(x =FieldType, y = Weight_GT20, color = FieldType))+
    geom_boxplot(width = 0.4)+
    geom_jitter(alpha=0.8, pch =20,size =5, position=position_jitter(width=0.1))+
    theme_bw()+
    labs(x = 'Field Type', y="Tuber weight (gm)") +
    theme(legend.position = "none")+
    theme(axis.title.x= element_blank(),
        axis.title.y = element_text(size =10, color ='black'),
        axis.text.x = element_text(size =10, color ='black'),
        axis.text.y = element_text(size =8, color = 'black'))+
scale_color_manual(name = 'Field Type', labels = c('Non-virgin', 'Virgin'), values = c('orange', 'chart
  facet_wrap(~ Location)
#a1
#dev.off()
```

Tuber count (more than 20 gm)

```
options(repr.plot.width = 5, repr.plot.height = 3)
#jpeg('TuberwtAll_soilprints.jpg',width=5,height=3,units='in',res=300)
b1= ggplot(data = df_21, aes(x =FieldType, y = Count_GT20, color = FieldType))+
    geom_boxplot(width = 0.4)+
    geom_jitter(alpha=0.8, pch =20,size =5, position=position_jitter(width=0.1))+
    theme_bw()+
    labs(x = 'Field Type', y="Number of tubers") +
    theme(legend.position = "none")+
    theme(axis.title.x= element_blank(),
        axis.title.y = element_text(size =10, color ='black'),
        axis.text.x = element_text(size =10, color ='black'),
        axis.text.y = element_text(size =8, color = 'black'))+
scale_color_manual(name = 'Field Type', labels = c('Non-virgin', 'Virgin'), values = c('orange', 'chart
  facet_wrap(~ Location)
#b1
#dev.off()
```

```
options(repr.plot.width = 10, repr.plot.height = 5)
#jpeg('microplot_21.jpg',width=10,height=5,units='in',res=300)
microplot_21 = ggarrange(a1,b1,c1, labels = c("A", "B", "C"), ncol = 2, nrow = 2)
microplot_21
```



```
#dev.off()
```

```
df21_wt_diff = df_21 %>% dplyr::select(FieldType, Pair, Wt_Tuber.g., Location) %>%
  spread(FieldType, Wt_Tuber.g.)
df21_wt_diff = rename(df21_wt_diff, non_virgin = `Non-virgin field`, virgin = `Virgin field`)
df21_wt_diff = mutate(df21_wt_diff, diff = virgin - non_virgin)
df21_wt_diff
```

Calculate yield Difference

```
##    Pair       Location non_virgin  virgin     diff
## 1     1 Columbia Basin    1586.65  207.26 -1379.39
## 2     2 Columbia Basin    1096.27  610.05  -486.22
## 3     3 Columbia Basin    1038.14  814.12  -224.02
## 4     4  Skagit Valley     859.18 1378.06   518.88
## 5     5 Columbia Basin     929.24  394.53  -534.71
## 6     6 Columbia Basin     617.91 1546.06   928.15
## 7     7 Columbia Basin     662.00 1300.45   638.45
## 8     8 Columbia Basin     698.16  566.92  -131.24
## 9     9  Skagit Valley     510.10 1162.44   652.34
## 10   10  Skagit Valley    1042.63 1100.35    57.72
## 11   11  Skagit Valley     934.47 1034.51   100.04
```

plot yield difference

```
fit_yield21 = lmer(Wt_Tuber.g. ~ FieldType + (1|Location),
                  data = df_21)
summary(fit_yield21)
```

Inferece on weight

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Wt_Tuber.g. ~ FieldType + (1 | Location)
##    Data: df_21
##
## REML criterion at convergence: 298.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -1.9213 -0.7495  0.0676  0.5053  1.8339
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  Location (Intercept)      0     0.0
##  Residual             137430   370.7
## Number of obs: 22, groups:  Location, 2
##
## Fixed effects:
##                        Estimate Std. Error     df t value Pr(>|t|)
## (Intercept)             906.80     111.78  20.00   8.113 9.38e-08 ***
## FieldTypeVirgin field    12.73     158.07  20.00   0.081    0.937
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr)
## FldTypVrgnf -0.707
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
anova(fit_yield21)
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##           Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## FieldType 890.91  890.91     1    20  0.0065 0.9366
```

```
## for columbia basin
df_21_CB = df_21[df_21$Location == 'Columbia Basin', ]
anova(lmer(Wt_Tuber.g. ~ FieldType + (1|Pair),
                  data = df_21_CB))
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##           Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## FieldType 100977  100977     1    12  0.5767 0.4623
```

```
anova(lm(Wt_Tuber.g. ~ FieldType, data = df_21_CB))
```

```
## Analysis of Variance Table
##
## Response: Wt_Tuber.g.
##           Df  Sum Sq Mean Sq F value Pr(>F)
## FieldType  1  100977  100977  0.5767 0.4623
## Residuals 12 2101226  175102
```

```
# conduct paired t test
t.test(Wt_Tuber.g.~ FieldType,
       #paired=T,
       data=df_21_CB)
```

```
##
##  Welch Two Sample t-test
##
## data:  Wt_Tuber.g. by FieldType
## t = 0.75939, df = 10.745, p-value = 0.464
## alternative hypothesis: true difference in means between group Non-virgin field and group Virgin fiel
## 95 percent confidence interval:
```

```
##  -323.8758  663.5844
## sample estimates:
## mean in group Non-virgin field     mean in group Virgin field
##                       946.9100                       777.0557
```

```
#for skagit valley
df_21_SV = df_21[df_21$Location == 'Skagit Valley', ]
anova(lmer(Wt_Tuber.g. ~ FieldType + (1|Pair),
                  data = df_21_SV))
```

```
## Type III Analysis of Variance Table with Satterthwaite's method
##           Sum Sq Mean Sq NumDF DenDF F value  Pr(>F)
## FieldType 220773  220773     1     6  5.8693 0.05167 .
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
```

```
anova(lm(Wt_Tuber.g. ~ FieldType, data = df_21_SV))
```

```
## Analysis of Variance Table
##
## Response: Wt_Tuber.g.
##           Df Sum Sq Mean Sq F value  Pr(>F)
## FieldType  1 220773  220773  5.8693 0.05167 .
## Residuals  6 225688   37615
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
```

```
# conduct paired t test
t.test(Wt_Tuber.g.~ FieldType,
       paired=T,
       data=df_21_SV)
```

```
##
##  Paired t-test
##
## data:  Wt_Tuber.g. by FieldType
## t = -2.2291, df = 3, p-value = 0.1121
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  -806.5759  142.0859
## sample estimates:
## mean difference
##        -332.245
```

```
# Test if there is difference between locations
t.test(diff~ Location,
       data=df21_wt_diff)
```

```
##
##  Welch Two Sample t-test
##
```

```
## data:  diff by Location
## t = -1.5352, df = 8.403, p-value = 0.1615
## alternative hypothesis: true difference in means between group Columbia Basin and group Skagit Valley
## 95 percent confidence interval:
##  -1250.0458   245.8472
## sample estimates:
## mean in group Columbia Basin  mean in group Skagit Valley
##                    -169.8543                     332.2450
```

```
inf = function(x ,df = df_21){
  fit = lmer(x ~ FieldType + (1|Location),
                  data = df)
return(anova(fit))}

variables = names(df_21)[-c(1,2,3,4,5,10,11,12,13,14,15,16,17,18,19, 20,21,25)]

df_var = df_21[variables]

for (i in 1:length(df_var)){
  print(names(df_var)[i])
  f = inf(df_var[[i]], df_21)
  print(f)
}
```

Inference on all variables

```
## [1] "Wt_Tuber.g."
## Type III Analysis of Variance Table with Satterthwaite's method
##           Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## FieldType 890.91  890.91     1    20  0.0065 0.9366
## [1] "Weight_GT20"
## Type III Analysis of Variance Table with Satterthwaite's method
##           Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## FieldType  548.1   548.1     1    20  0.0042 0.9491
## [1] "Count_Tuber"
## Type III Analysis of Variance Table with Satterthwaite's method
##           Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## FieldType 3.6818  3.6818     1    19  0.1004 0.7548
## [1] "Count_GT20"
## Type III Analysis of Variance Table with Satterthwaite's method
##           Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## FieldType 1.1364  1.1364     1    19   0.102 0.7529
## [1] "Pyth_Avg_Soil"
## Type III Analysis of Variance Table with Satterthwaite's method
##           Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## FieldType 9327.7  9327.7     1    19  2.4829 0.1316
## [1] "ResPyth_Avg_Soil"
## Type III Analysis of Variance Table with Satterthwaite's method
##           Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## FieldType 248.91  248.91     1    19    1.84 0.1909
## [1] "ResPyth_Per_Soil"
## Type III Analysis of Variance Table with Satterthwaite’s method
##           Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## FieldType 1693.1  1693.1     1    19  2.7803 0.1118
## [1] "Fus_Avg_Soil"
## Type III Analysis of Variance Table with Satterthwaite’s method
##           Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## FieldType 149903  149903     1    19  0.3926 0.5384
## [1] "Vd_Tot_Soil"
## Type III Analysis of Variance Table with Satterthwaite’s method
##            Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## FieldType 0.18182 0.18182     1    19  0.0625 0.8052
## [1] "BD_Tot_Soil"
## Type III Analysis of Variance Table with Satterthwaite’s method
##            Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## FieldType 0.18182 0.18182     1    20       1 0.3293
## [1] "POXC"
## Type III Analysis of Variance Table with Satterthwaite’s method
##           Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## FieldType  143.8   143.8     1    19  0.0156  0.902
## [1] "AUDPC"
## Type III Analysis of Variance Table with Satterthwaite’s method
##           Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## FieldType 222910  222910     1    19   1.161 0.2948
```

```r
df_21_CB = df_21[df_21$Location == 'Columbia Basin', ]
inf = function(x ,df = df_21_CB){
  fit = lmer(x ~ FieldType + (1|Pair),
               data = df)
return(anova(fit))}

variables = names(df_21_CB)[-c(1,2,3,4,5,10,11,12,13,14,15,16,17,18,19, 20,21,25)]
variables
```

Columbia Basin

```
##  [1] "Wt_Tuber.g."      "Weight_GT20"      "Count_Tuber"      "Count_GT20"
##  [5] "Pyth_Avg_Soil"    "ResPyth_Avg_Soil" "ResPyth_Per_Soil" "Fus_Avg_Soil"
##  [9] "Vd_Tot_Soil"      "BD_Tot_Soil"      "POXC"             "AUDPC"
```

```r
df_var = df_21_CB[variables]

for (i in 1:length(df_var)){
  print(names(df_var)[i])
  f = inf(df_var[[i]], df_21_CB)
  print(f)
}
```

```
## [1] "Wt_Tuber.g."
## Type III Analysis of Variance Table with Satterthwaite’s method
##           Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## FieldType 100977  100977     1    12  0.5767 0.4623
## [1] "Weight_GT20"
## Type III Analysis of Variance Table with Satterthwaite’s method
##           Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## FieldType 149101  149101     1    12  0.9209 0.3562
## [1] "Count_Tuber"
## Type III Analysis of Variance Table with Satterthwaite’s method
##            Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## FieldType 0.64286 0.64286     1     6  0.0789 0.7882
## [1] "Count_GT20"
## Type III Analysis of Variance Table with Satterthwaite’s method
##           Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## FieldType 5.7857  5.7857     1    12   0.498 0.4939
## [1] "Pyth_Avg_Soil"
## Type III Analysis of Variance Table with Satterthwaite’s method
##           Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## FieldType    350     350     1    12  0.1512 0.7042
## [1] "ResPyth_Avg_Soil"
## Type III Analysis of Variance Table with Satterthwaite’s method
##           Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## FieldType 457.14  457.14     1     6  2.4603 0.1678
## [1] "ResPyth_Per_Soil"
## Type III Analysis of Variance Table with Satterthwaite’s method
##           Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## FieldType   2744    2744     1     6  3.8413 0.0977 .
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
## [1] "Fus_Avg_Soil"
## Type III Analysis of Variance Table with Satterthwaite’s method
##           Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## FieldType  81016   81016     1     6  0.3312 0.5859
## [1] "Vd_Tot_Soil"
## Type III Analysis of Variance Table with Satterthwaite’s method
##            Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## FieldType 0.28571 0.28571     1    12     0.2 0.6627
## [1] "BD_Tot_Soil"
## Type III Analysis of Variance Table with Satterthwaite’s method
##            Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## FieldType 0.28571 0.28571     1    12       1  0.337
## [1] "POXC"
## Type III Analysis of Variance Table with Satterthwaite’s method
##           Sum Sq Mean Sq NumDF DenDF F value Pr(>F)
## FieldType 6589.5  6589.5     1    12  1.4626 0.2498
## [1] "AUDPC"
## Type III Analysis of Variance Table with Satterthwaite’s method
##           Sum Sq Mean Sq NumDF DenDF F value  Pr(>F)
## FieldType 748209  748209     1    12   3.289 0.09481 .
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
```

```
df_21_SV = df_21[df_21$Location == 'Skagit Valley', ]
inf = function(x ,df = df_21_SV){
  fit = lmer(x ~ FieldType + (1|Pair),
                  data = df)
return(anova(fit))}

variables = names(df_21_SV)[-c(1,2,3,4,5,10,11,12,13,14,15,16,17,18,19, 20,21,25, 28)]
variables
df_var = df_21_SV[variables]

for (i in 1:length(df_var)){
  print(names(df_var)[i])
  f = inf(df_var[[i]], df_21_SV)
  print(f)
}
```

Skagit valley

Soil biological properties

Pythium

```
options(repr.plot.width = 5, repr.plot.height = 3)
#jpeg('TuberwtAll_soilprints.jpg',width=5,height=3,units='in',res=300)
sb1= ggplot(data = df_21, aes(x =FieldType, y = Pyth_Avg_Soil, color = FieldType))+
    geom_boxplot(width = 0.4)+
    geom_jitter(alpha=0.8, pch =20,size =5, position=position_jitter(width=0.1))+
    theme_bw()+
    labs(x = 'Field Type', y=expression(paste("Average ",italic("Pythium")))) +
    theme(legend.position = "none")+
    theme(axis.title.x= element_blank(),
        axis.title.y = element_text(size =10, color ='black'),
        axis.text.x = element_text(size =10, color ='black'),
        axis.text.y = element_text(size =8, color = 'black'))+
scale_color_manual(name = 'Field Type', labels = c('Non-virgin', 'Virgin'), values = c('orange', 'chart
  facet_wrap(~ Location)
#sb1
#dev.off()
```

Resistant Pythium

```
options(repr.plot.width = 5, repr.plot.height = 3)
#jpeg('TuberwtAll_soilprints.jpg',width=5,height=3,units='in',res=300)
sb2= ggplot(data = df_21, aes(x =FieldType, y = ResPyth_Avg_Soil, color = FieldType))+
    geom_boxplot(width = 0.4)+
    geom_jitter(alpha=0.8, pch =20,size =5, position=position_jitter(width=0.1))+
    theme_bw()+
    labs(x = 'Field Type', y=expression(paste("Average Res.  ",italic("Pythium")))) +
    theme(legend.position = "none")+
    theme(axis.title.x= element_blank(),
        axis.title.y = element_text(size =10, color ='black'),
        axis.text.x = element_text(size =10, color ='black'),
        axis.text.y = element_text(size =8, color = 'black'))+
scale_color_manual(name = 'Field Type', labels = c('Non-virgin', 'Virgin'), values = c('orange', 'chart
  facet_wrap(~ Location)
#sb2
#dev.off()
```

Fusarium

```
options(repr.plot.width = 5, repr.plot.height = 3)
#jpeg('TuberwtAll_soilprints.jpg',width=5,height=3,units='in',res=300)
sb3= ggplot(data = df_21, aes(x =FieldType, y = Fus_Avg_Soil, color = FieldType))+
    geom_boxplot(width = 0.4)+
    geom_jitter(alpha=0.8, pch =20,size =5, position=position_jitter(width=0.1))+
    theme_bw()+
    labs(x = 'Field Type', y=expression(paste("Average ",italic("Fusarium")))) +
    theme(legend.position = "none")+
    theme(axis.title.x= element_blank(),
        axis.title.y = element_text(size =10, color ='black'),
        axis.text.x = element_text(size =10, color ='black'),
        axis.text.y = element_text(size =8, color = 'black'))+
scale_color_manual(name = 'Field Type', labels = c('Non-virgin', 'Virgin'), values = c('orange', 'chart
  facet_wrap(~ Location)
#sb3
#dev.off()
```

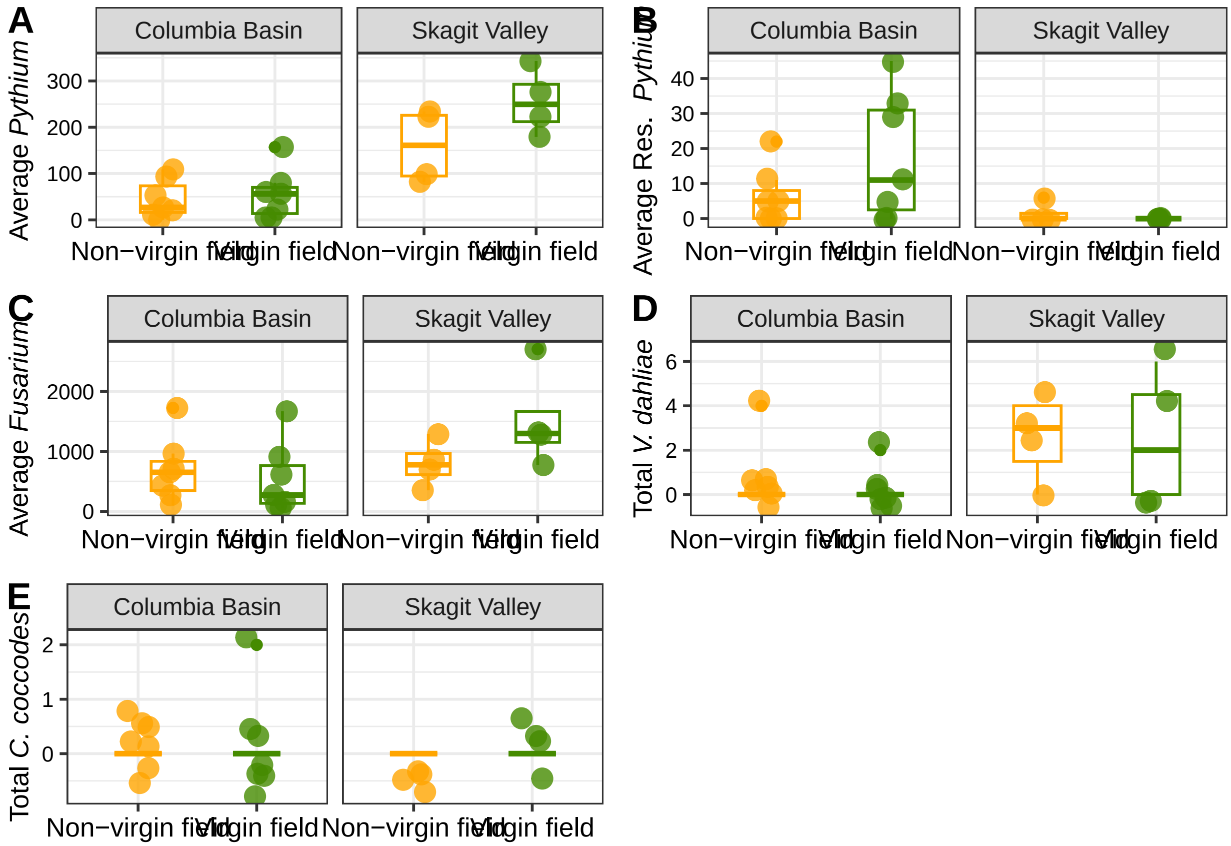
Verticillium

```
options(repr.plot.width = 5, repr.plot.height = 3)
#jpeg('TuberwtAll_soilprints.jpg',width=5,height=3,units='in',res=300)
sb4= ggplot(data = df_21, aes(x =FieldType, y = Vd_Tot_Soil, color = FieldType))+
    geom_boxplot(width = 0.4)+
    geom_jitter(alpha=0.8, pch =20,size =5, position=position_jitter(width=0.1))+
    theme_bw()+
    labs(x = 'Field Type', y=expression(paste("Total ",italic("V. dahliae")))) +
    theme(legend.position = "none")+
    theme(axis.title.x= element_blank(),
        axis.title.y = element_text(size =10, color ='black'),
        axis.text.x = element_text(size =10, color ='black'),
        axis.text.y = element_text(size =8, color = 'black'))+
scale_color_manual(name = 'Field Type', labels = c('Non-virgin', 'Virgin'), values = c('orange', 'chart
  facet_wrap(~ Location)
#sb4
#dev.off()
```

Colletotrichum

```
options(repr.plot.width = 5, repr.plot.height = 3)
#jpeg('TuberwtAll_soilprints.jpg',width=5,height=3,units='in',res=300)
sb5= ggplot(data = df_21, aes(x =FieldType, y = BD_Tot_Soil, color = FieldType))+
    geom_boxplot(width = 0.4)+
    geom_jitter(alpha=0.8, pch =20,size =5, position=position_jitter(width=0.1))+
    theme_bw()+
    labs(x = 'Field Type', y=expression(paste("Total ",italic("C. coccodes")))) +
    theme(legend.position = "none")+
    theme(axis.title.x= element_blank(),
        axis.title.y = element_text(size =10, color ='black'),
        axis.text.x = element_text(size =10, color ='black'),
        axis.text.y = element_text(size =8, color = 'black'))+
scale_color_manual(name = 'Field Type', labels = c('Non-virgin', 'Virgin'), values = c('orange', 'chart
  facet_wrap(~ Location)
#sb5
#dev.off()
```

```
options(repr.plot.width = 10, repr.plot.height = 6)
#jpeg('soil_pathogens_21.jpg',width=10,height=7,units='in',res=300)
plot = ggarrange(sb1,sb2,sb3,sb4,sb5, labels = c("A", "B", "C", "D","E"), ncol = 2, nrow = 3)
plot
```

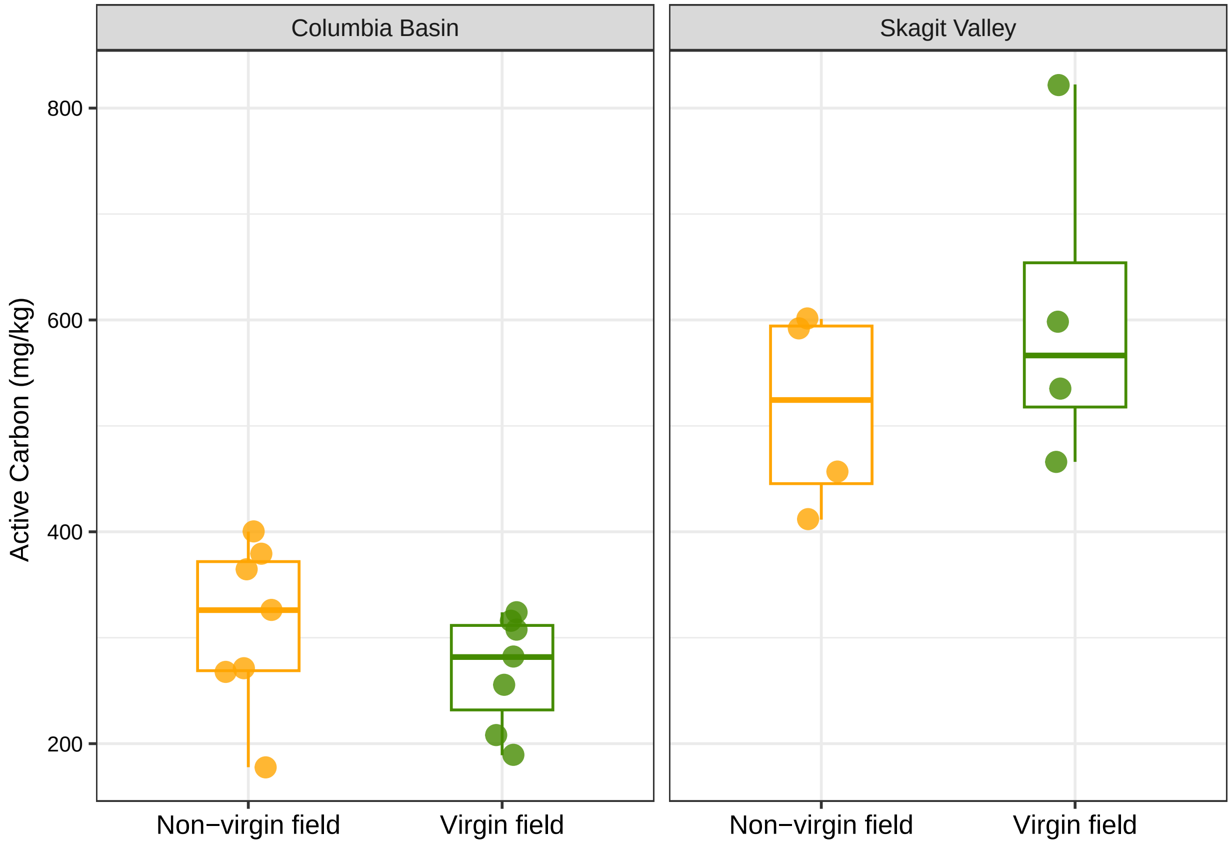


```
#dev.off()
```

Soil phyiscal and chemical properties

Active Carbon

```
options(repr.plot.width = 5, repr.plot.height = 3)
#jpeg('ActiveCarbon_21.jpg',width=5,height=3,units='in',res=300)
sp1= ggplot(data = df_21, aes(x =FieldType, y = POXC, color = FieldType))+
    geom_boxplot(width = 0.4)+
    geom_jitter(alpha=0.8, pch =20,size =5, position=position_jitter(width=0.1))+
    theme_bw()+
    labs(x = 'Field Type', y=expression(paste("Active Carbon (mg/kg) "))) +
    theme(legend.position = "none")+
    theme(axis.title.x= element_blank(),
        axis.title.y = element_text(size =10, color ='black'),
        axis.text.x = element_text(size =10, color ='black'),
        axis.text.y = element_text(size =8, color = 'black'))+
scale_color_manual(name = 'Field Type', labels = c('Non-virgin', 'Virgin'), values = c('orange', 'chart
  facet_wrap(~ Location)
sp1
```



```
#dev.off()
```

2022 data visualization and inference

Microplot Study

```
yield = read_csv(file = 'C:/Users/theox/Desktop/Projects/SoilFingerPrints/Data/2022/Microplot/microplot_
yield = yield[c('ID', 'location', 'pair', 'FieldType', 'TW_gt20', 'TW_All', 'TC_gt20', 'TC_All')]

yield$location = gsub('columbia Basin', 'Columbia Basin', yield$location)

yield$FieldType = gsub('Virgin field', "Virgin", yield$FieldType)
yield$FieldType = gsub('Non-virginfield', "Non-virgin", yield$FieldType)
yield$FieldType = gsub('Non-virgin Field', "Non-virgin", yield$FieldType)
yield$FieldType = gsub('Non-virgin field', "Non-virgin", yield$FieldType)
yield$FieldType = gsub('Virgin Field', "Virgin", yield$FieldType)

yield[c("ID", "pair", "FieldType","location")]  = lapply(yield[c("ID", "pair", "FieldType","location")]
head(yield)
```

```
## # A tibble: 6 x 8
##   ID    location       pair  FieldType  TW_gt20 TW_All TC_gt20 TC_All
##   <fct> <fct>          <fct> <fct>        <dbl>  <dbl>   <dbl>  <dbl>
## 1 24    Columbia Basin 1     Virgin       1502.  1502.      17     17
## 2 25    Columbia Basin 1     Non-virgin   1622.  1622.      15     15
## 3 26    Columbia Basin 1     Native       1584.  1584.      11     11
## 4 27    Columbia Basin 2     Virgin       1286.  1297.      11     12
## 5 28    Columbia Basin 2     Non-virgin   1636.  1636.      13     13
## 6 29    Columbia Basin 2     Native       1042.  1083.      12     15
```

```
#unique(yield$FieldType)
#str(yield)
#lapply(yield, unique)
```

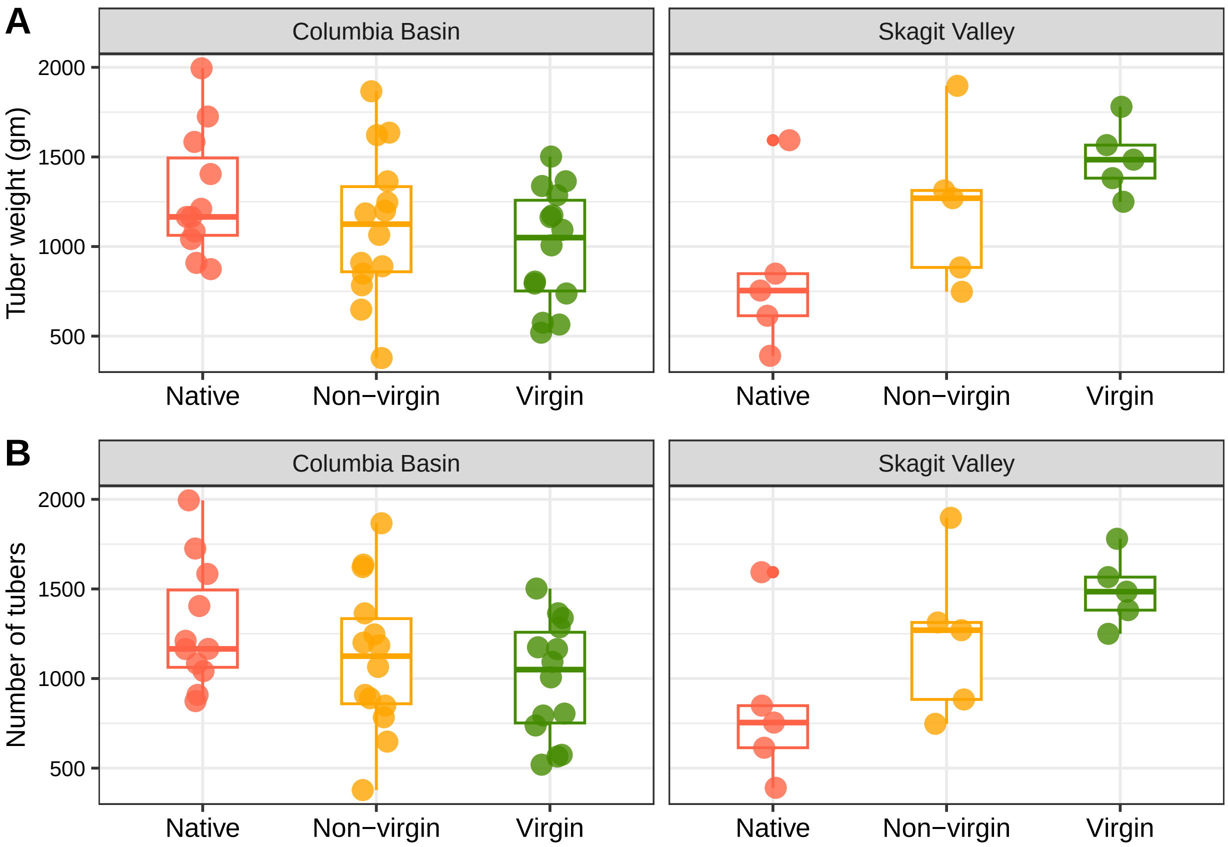
Tuber weigt more than 20gm

```
options(repr.plot.width = 5, repr.plot.height = 3)
#jpeg('Tuberwt_gt20_222_combinedlocations.jpg',width=5,height=3,units='in',res=300)
a= ggplot(data = yield, aes(x =FieldType, y = TW_gt20, color = FieldType))+
    geom_boxplot(width = 0.4)+
    geom_jitter(alpha=0.8, pch =20,size =5, position=position_jitter(width=0.1))+
    theme_bw()+
    labs(x = 'Field Type', y="Tuber weight (gm)") +
    theme(legend.position = "none")+
    theme(axis.title.x= element_blank(),
        axis.title.y = element_text(size =10, color ='black'),
        axis.text.x = element_text(size =10, color ='black'),
        axis.text.y = element_text(size =8, color = 'black'))+facet_wrap(~ location)+
scale_color_manual(name = 'Field Type', labels = c('Non-virgin', 'Virgin', 'Native'),
                  values = c('tomato', 'orange', 'chartreuse4'))
#a
#dev.off()
```

Tuber count more than 20gm

```
options(repr.plot.width = 5, repr.plot.height = 20)
#jpeg('TuberCount_gt20_22_locationwise.jpg',width=5,height=3,units='in',res=300)
b= ggplot(data = yield, aes(x =FieldType, y = TW_gt20, color = FieldType))+
    geom_boxplot(width = 0.4)+
    geom_jitter(alpha=0.8, pch =20,size =5, position=position_jitter(width=0.1))+
    theme_bw()+
    labs(x = 'Field Type', y="Number of tubers") +
    theme(legend.position = "none")+
    theme(axis.title.x= element_blank(),
        axis.title.y = element_text(size =10, color ='black'),
        axis.text.x = element_text(size =10, color ='black'),
        axis.text.y = element_text(size =8, color = 'black'))+
scale_color_manual(name = 'Field Type', labels = c('Non-virgin', 'Virgin', 'Native'),
                  values = c('tomato', 'orange', 'chartreuse4'))+facet_grid(~ location)
#b
#dev.off()
```

```
options(repr.plot.width = 7, repr.plot.height = 6)
#jpeg('yield 2022_combinedLocations.jpg',width=7,height=6,units='in',res=300)
combined_plot = ggarrange(a,b, labels = c("A", "B"), ncol = 1, nrow = 2)
combined_plot
```



```r
#dev.off()
```

Tuber weigt more than 20gm

Tuber weigt more than 20gm

Inference (linear mixed model)

Here, locations: columbia basin and skagit valley are considered random effect and field type is considered fixed effect.

Inference on tuber weight

```r
fit_yield = lmer(TC_gt20 ~ FieldType + (1|location),
                 data = yield)
#summary(fit_yield)
anova(fit_yield)
```

```
## Type III Analysis of Variance Table with Satterthwaite’s method
##           Sum Sq Mean Sq NumDF  DenDF F value Pr(>F)
## FieldType 8.2981  4.1491     2 50.044  0.3283 0.7217
```

```r
library(emmeans)
emmeans(fit_yield, list(pairwise ~ FieldType), adjust = "tukey")
```

```
## $‘emmeans of FieldType‘
##  FieldType  emmean   SE   df lower.CL upper.CL
##  Native       12.5 1.19 2.56     8.36     16.7
##  Non-virgin   13.2 1.16 2.13     8.52     17.9
##  Virgin       13.5 1.16 2.13     8.79     18.2
##
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95
##
## $‘pairwise differences of FieldType‘
##  1                      estimate   SE   df t.ratio p.value
##  Native - (Non-virgin)    -0.693 1.21 50.1  -0.573  0.8349
##  Native - Virgin          -0.956 1.21 50.1  -0.791  0.7103
##  (Non-virgin) - Virgin    -0.263 1.15 50.0  -0.228  0.9717
##
## Degrees-of-freedom method: kenward-roger
## P value adjustment: tukey method for comparing a family of 3 estimates
```

```r
library(lmerTest)
#difflsmeans(fit_yield, test.effs = "FieldType", df="Kenward-Roger")

#Multiple comparison
library(multcomp)
#summary(glht(fit_yield, linfct = mcp(FieldType = "Tukey")), test = adjusted("holm"))
```

```
## for columbia basin
df_22_CB =yield[yield$location == 'Columbia Basin', ]
anova(lmer(TW_gt20 ~ FieldType + (1|pair),
                  data = df_22_CB))
```

```
## Type III Analysis of Variance Table with Satterthwaite’s method
##           Sum Sq Mean Sq NumDF  DenDF F value Pr(>F)
## FieldType 491904  245952     2 24.085  2.1942 0.1332
```

```
anova(lm(TW_gt20 ~ FieldType, data = df_22_CB))
```

```
## Analysis of Variance Table
##
## Response: TW_gt20
##           Df  Sum Sq Mean Sq F value Pr(>F)
## FieldType  2  527182  263591  1.9504  0.157
## Residuals 36 4865370  135149
```

```
#for skagit valley
df_22_SV = yield[yield$location == 'Skagit Valley', ]
fit_yield_22_SV = lmer(TW_gt20 ~ FieldType + (1|pair),
                  data = df_22_SV)

#summary(fit_yield)
anova(fit_yield_22_SV)
```

```
## Type III Analysis of Variance Table with Satterthwaite’s method
##            Sum Sq Mean Sq NumDF DenDF F value  Pr(>F)
## FieldType 1074105  537052     2    12  3.5926 0.05988 .
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
```

```
library(lmerTest)
#difflsmeans(fit_yield_22_SV, test.effs = "FieldType", df="Kenward-Roger")

#Multiple comparison
library(multcomp)
summary(glht(fit_yield_22_SV, linfct = mcp(FieldType = "Tukey")), test = adjusted("holm"))
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Multiple Comparisons of Means: Tukey Contrasts
##
##
## Fit: lmer(formula = TW_gt20 ~ FieldType + (1 | pair), data = df_22_SV)
##
## Linear Hypotheses:
##                          Estimate Std. Error z value Pr(>|z|)
## Non-virgin - Native == 0    381.9      244.5   1.562   0.2366
## Virgin - Native == 0        652.3      244.5   2.668   0.0229 *
## Virgin - Non-virgin == 0    270.4      244.5   1.106   0.2689
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
## (Adjusted p values reported -- holm method)
```

Inference on tuber count

```
fit_count = lmer(TC_gt20 ~ FieldType + (1|location),
                  data = yield)
#summary(fit_count)
anova(fit_count)
```

```
## Type III Analysis of Variance Table with Satterthwaite’s method
##           Sum Sq Mean Sq NumDF  DenDF F value Pr(>F)
## FieldType 8.2981  4.1491     2 50.044  0.3283 0.7217
```

```
#plot(fit_count)

library(emmeans)
emmeans(fit_count, list(pairwise ~ FieldType), adjust = "tukey")
```

```
## $‘emmeans of FieldType‘
##  FieldType   emmean   SE   df lower.CL upper.CL
##  Native        12.5 1.19 2.56     8.36     16.7
##  Non-virgin    13.2 1.16 2.13     8.52     17.9
##  Virgin        13.5 1.16 2.13     8.79     18.2
##
## Degrees-of-freedom method: kenward-roger
## Confidence level used: 0.95
##
## $‘pairwise differences of FieldType‘
##  1                     estimate   SE   df t.ratio p.value
##  Native - (Non-virgin)   -0.693 1.21 50.1  -0.573  0.8349
##  Native - Virgin         -0.956 1.21 50.1  -0.791  0.7103
##  (Non-virgin) - Virgin   -0.263 1.15 50.0  -0.228  0.9717
##
## Degrees-of-freedom method: kenward-roger
## P value adjustment: tukey method for comparing a family of 3 estimates
```

2022 Nematode Abundance

Data preparation

```
df_NC = read_csv('C:/Users/theox/Desktop/Projects/SoilFingerPrints/Data/2022/Nematode_morpho/Nematode_ra

df_NC[c("Sites", "pair", "location", "FieldType")] <- lapply(df_NC[c("Sites", "pair", "location", "Fiel
                                                factor)

df_NC$FieldType = gsub('Field', "", df_NC$FieldType)
```

```r
df_NC$FieldType = gsub('field', "", df_NC$FieldType)
df_NC$FieldType = gsub('Non-virgin ', "Non-virgin", df_NC$FieldType)
df_NC$FieldType = gsub('Virgin ', "Virgin", df_NC$FieldType)
df_NC$location = gsub('columbia Basin', 'Columbia Basin', df_NC$location)
unique(df_NC$FieldType)
```

```
## [1] "Virgin"     "Non-virgin" "Native"
```

```r
unique(df_NC$location)
```

```
## [1] "Columbia Basin" "Skagit Valley"
```

```r
df_taxa = df_NC[-c(1,2,3,4)]
labels = df_NC[c(1,2,3,4)]
dim(df_NC)
```

```
## [1] 54 47
```

Visualization

```r
# check total abundance in each sample
rowSums(df_taxa)
```

```
##  [1]  39 124  49  25  75  55 252  80  70  37  34 161 230  66 104 145  81  63  51
## [20]  51  77 182 151 219  91 192 289 133  39  95 220 235 258  72  50 295  99 131
## [39] 205  83 271  76 126 460 144 246 515 291 457  86  26  85 279 130
```

```r
# Turn percent cover to relative abundance by dividing each value by sample
# total abundance
comm <- decostand(df_taxa, method = "total")
# check total abundance in each sample
colSums(comm)
```

```
##          Acrobeles       Acrobeloides         Anguinidae     Aphelenchoides
##        3.586755680       11.541159829        0.314471962        8.469110266
##        Aphelenchus     Aporcelaimidae    Aporcelaimellus       Aporcelaimus
##        3.120362496        0.013888889        0.005825243        0.071805978
##          Axonchium     Carcharolaimus        Cervidellus         Cephalobus
##        0.019379845        0.178430426        0.072727273        0.786993583
##       Cephalenchus        Chiloplacus           Cruznema            Clarkus
##        0.079584775        0.586700818        1.790555810        0.097915687
##          Criconema        Ditylenchus       Dorylaimidae        Discolaimus
##        0.350457604        1.158071154        0.317360108        0.014946833
##      Dorylaimellus         Dorylaimus     Dolichodoridae       Eudorylaimus
##        0.006622517        0.042529667        0.000000000        0.101698718
##      Ecphyadophora       Eucephalobus          Filenchus         Geocenamus
##        0.019512195        0.248492787        0.642433845        0.104545455
##         Gracilacus   Heterocephalobus         Longidorus        Meloidogyne
##        0.199630304        1.536414696        0.017301038        0.429553265
## Metateratocephalus          Mononchus     Nothotylenchus      Paratylenchus
##        0.072598584        0.021287976        0.063492063        1.016415753
##            Plectus       Pratylenchus         Psilenchus        Rhabditidae
##        0.241336530        1.978526756        0.031141869       12.096376262
##          Rhabditis       Quinisulcius        Tylenchidae
##        0.799553551        0.245592737        1.508439173
```

```
rowSums(comm)
```

```
##  [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
## [39] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
```

```
# select major taxa using threshold 0.3

taxa_sel = comm[colSums(comm)>0.3]
names(taxa_sel)
```

```
##  [1] "Acrobeles"        "Acrobeloides"     "Anguinidae"       "Aphelenchoides"
##  [5] "Aphelenchus"      "Cephalobus"       "Chiloplacus"      "Cruznema"
##  [9] "Criconema"        "Ditylenchus"      "Dorylaimidae"     "Filenchus"
## [13] "Heterocephalobus" "Meloidogyne"      "Paratylenchus"    "Pratylenchus"
## [17] "Rhabditidae"      "Rhabditis"        "Tylenchidae"
```

```
taxa_other = comm[!colSums(comm)>0.3]
names(taxa_other)
```

```
##  [1] "Aporcelaimidae"     "Aporcelaimellus"    "Aporcelaimus"
##  [4] "Axonchium"          "Carcharolaimus"     "Cervidellus"
##  [7] "Cephalenchus"       "Clarkus"            "Discolaimus"
## [10] "Dorylaimellus"      "Dorylaimus"         "Dolichodoridae"
## [13] "Eudorylaimus"       "Ecphyadophora"      "Eucephalobus"
## [16] "Geocenamus"         "Gracilacus"         "Longidorus"
## [19] "Metateratocephalus" "Mononchus"          "Nothotylenchus"
## [22] "Plectus"            "Psilenchus"         "Quinisulcius"
```

```
names(taxa_other) = rep('Other',length(taxa_other))
dim(taxa_other); dim(taxa_sel)
```

```
## [1] 54 24
```

```
## [1] 54 19
```

```
#taxa_other = comm[!unlist(names(comm)) %in% sel_taxa]
```

- Melt to longer format

```
df_melt = cbind(labels, taxa_sel, taxa_other)

df_melt = melt(as.data.table(df_melt), c("Sites", "location", "pair", "FieldType"))
dim(df_melt)
```

```
## [1] 2322    6
```

```
head(df_melt)
```

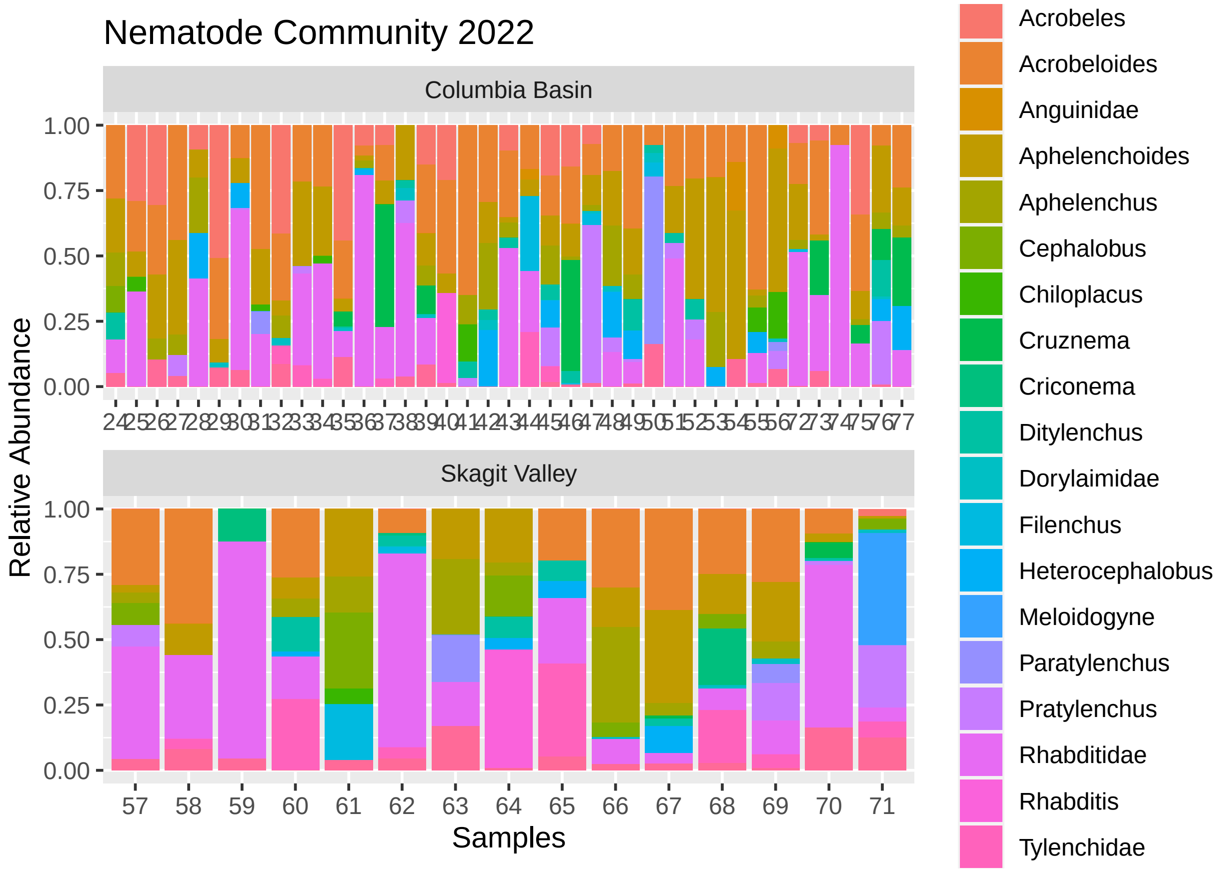
```
##    Sites       location pair  FieldType  variable      value
## 1:    24 Columbia Basin    1     Virgin Acrobeles 0.00000000
## 2:    25 Columbia Basin    1 Non-virgin Acrobeles 0.29032258
## 3:    26 Columbia Basin    1     Native Acrobeles 0.30612245
## 4:    27 Columbia Basin    2     Virgin Acrobeles 0.00000000
## 5:    28 Columbia Basin    2 Non-virgin Acrobeles 0.09333333
## 6:    29 Columbia Basin    2     Native Acrobeles 0.50909091
```

```
unique(df_melt$variable)
```

```
##  [1] Acrobeles         Acrobeloides      Anguinidae        Aphelenchoides
##  [5] Aphelenchus       Cephalobus        Chiloplacus       Cruznema
##  [9] Criconema         Ditylenchus       Dorylaimidae      Filenchus
## [13] Heterocephalobus Meloidogyne       Paratylenchus     Pratylenchus
## [17] Rhabditidae       Rhabditis         Tylenchidae       Other
## 20 Levels: Acrobeles Acrobeloides Anguinidae Aphelenchoides ... Other
```

plot Nematode Abundance all samples

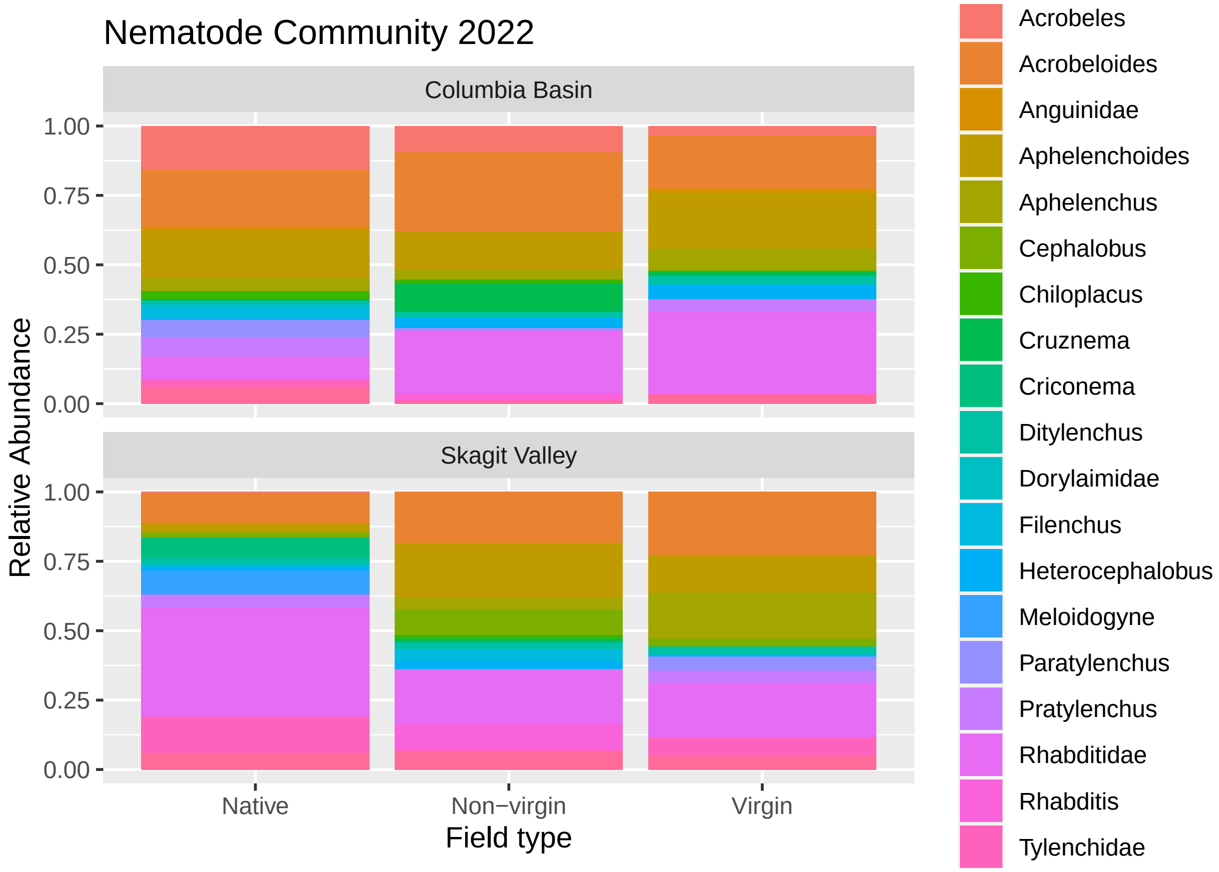
```
options(repr.plot.width = 15, repr.plot.height = 6)
#jpeg('Abundance_samples.jpg',width=12,height=6,units='in',res=300)
bar = ggplot(df_melt, aes(fill = variable, y=value, x= Sites)) +
    geom_bar(position="fill", stat="identity", alpha = 1)+
  labs(x = 'Samples', y="Relative Abundance", title="Nematode Community 2022")+
    facet_wrap(~location, scales = 'free', ncol = 1)+theme(legend.title=element_blank())
bar
```



```
#dev.off()
```

Nematode abundance based on field type

```
options(repr.plot.width = 10, repr.plot.height = 6)
#jpeg('Abundance_fieldType.jpg',width=10,height=6,units='in',res=300)
bar = ggplot(df_melt, aes(fill = variable, y=value, x= FieldType)) +
    geom_bar(position="fill", stat="identity", alpha = 1)+
  labs(x="Field type", y="Relative Abundance", title="Nematode Community 2022")+
  facet_wrap(~location, ncol = 1)+theme(legend.title=element_blank())
bar
```

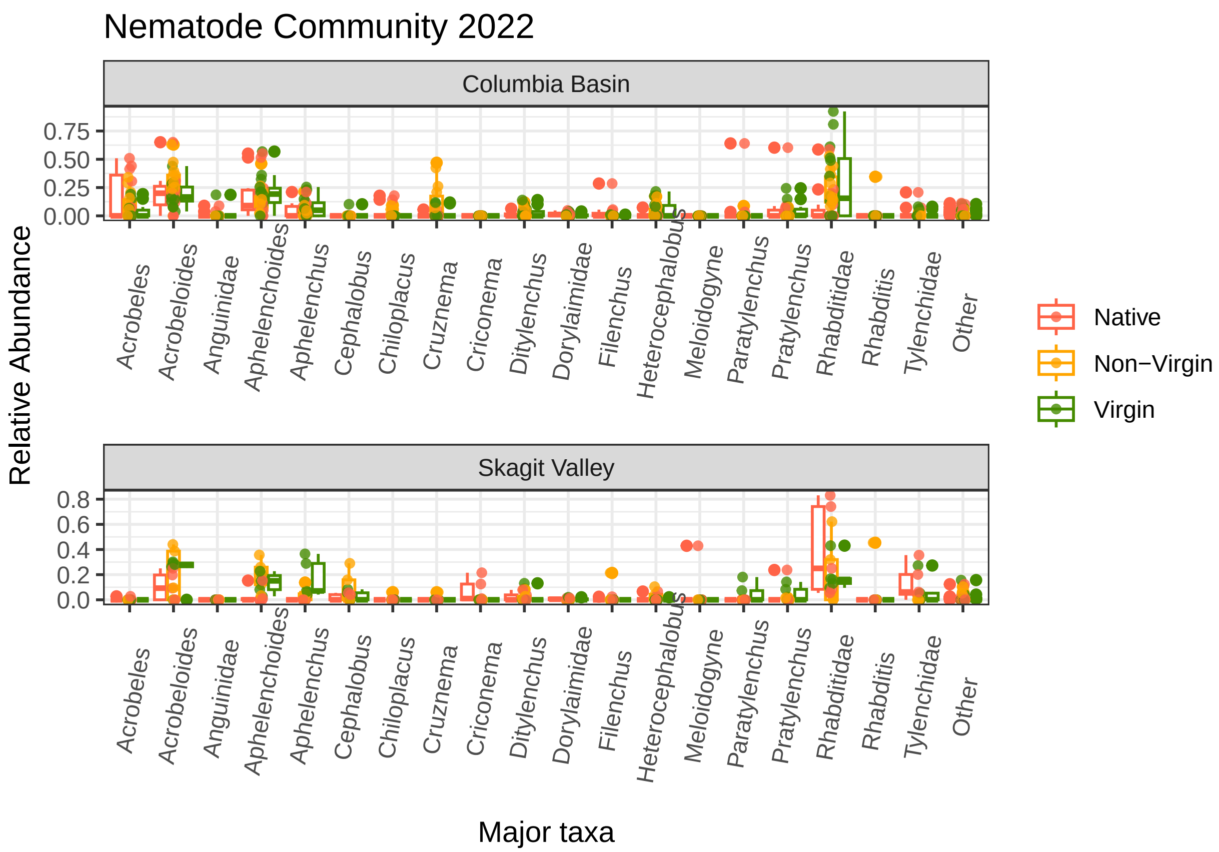


```
#dev.off()
```

Box plot

```
options(repr.plot.width = 13, repr.plot.height = 7)
#jpeg('Boxplot_nematode abundance_fieldType.jpg',width=13,height=7,units='in',res=300)
boxplot = ggplot(df_melt, aes(y=value, x= variable, color = FieldType)) +
    geom_boxplot(width = 0.9)+
    geom_jitter(alpha=0.8, pch =20,size =2, position=position_jitter(width=0.05))+
  labs(x="Major taxa", y="Relative Abundance", title="Nematode Community 2022")+ theme_bw()+
  theme(axis.text.x=element_text(angle =80, vjust=0.5, hjust=0.35), legend.title=element_blank())+
  facet_wrap(~location, ncol = 1, scales = 'free')+
  scale_color_manual(name = 'Field Type', labels = c('Native', 'Non-Virgin', 'Virgin'),
                 values = c('tomato', 'orange', 'chartreuse4'))

boxplot
```



```
#dev.off()
```

Shannon's diversity

```
library(phyloseq)
library('picante')
```

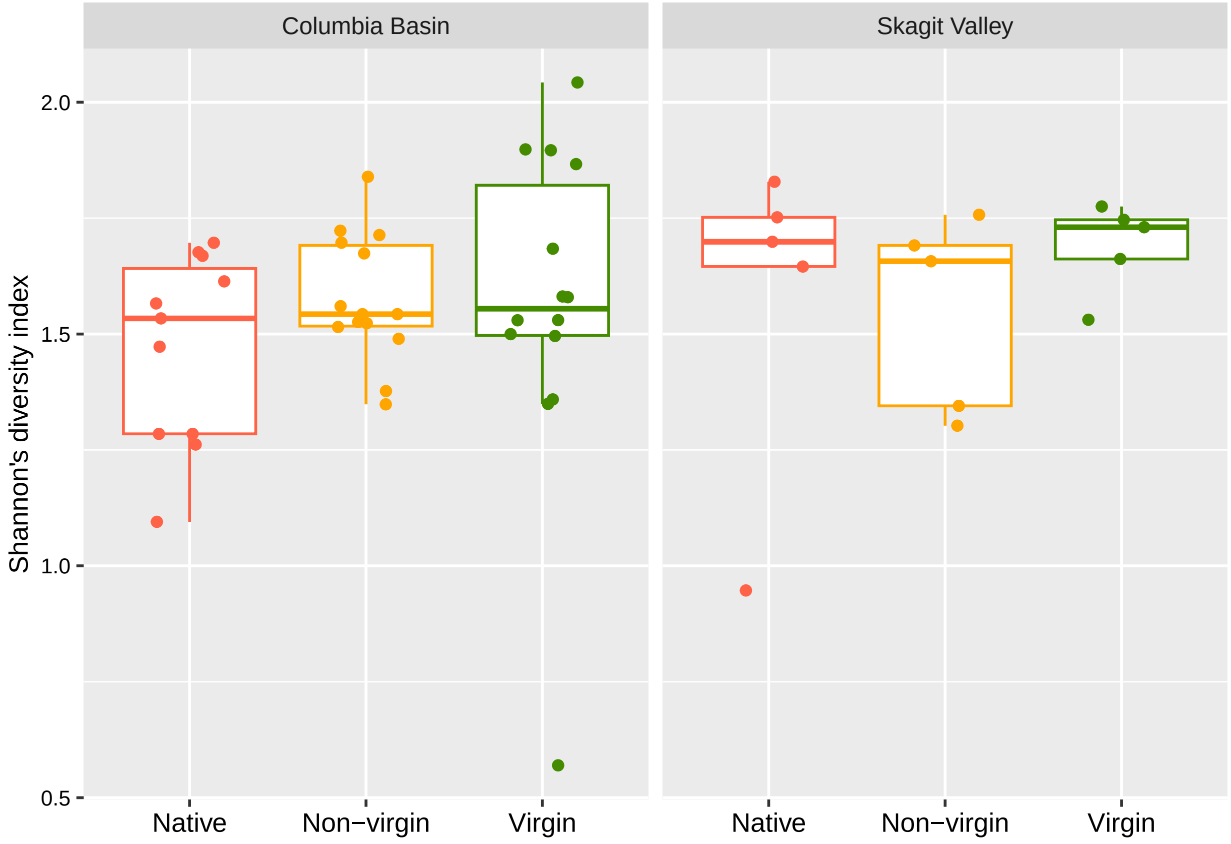
```
#Subsample reads
count_tab_phy <- otu_table(df_taxa, taxa_are_rows=F)
ps_rare <- rarefy_even_depth(count_tab_phy, rngseed = 123, replace = FALSE)
dim(ps_rare)
```

```
## [1] 54 33
```

```
log_rare= log2(ps_rare +1)
shannon = phyloseq::estimate_richness(log_rare, measures = "Shannon")
#observed = phyloseq::estimate_richness(count_tab_phy, measures = "Observed")
diversity = cbind(shannon, labels)
head(diversity)
```

```
##      Shannon Sites       location pair  FieldType
## sa1 1.898206    24 Columbia Basin    1     Virgin
## sa2 1.559994    25 Columbia Basin    1 Non-virgin
## sa3 1.565953    26 Columbia Basin    1     Native
## sa4 1.499796    27 Columbia Basin    2     Virgin
## sa5 1.542766    28 Columbia Basin    2 Non-virgin
## sa6 1.284638    29 Columbia Basin    2     Native
```

```
options(repr.plot.width = 5, repr.plot.height = 20)
#jpeg('Alpha diveristy.jpg',width=5,height=3,units='in',res=300)
b= ggplot(data = diversity, aes(x =FieldType, y = Shannon, color = FieldType))+
    geom_boxplot(outlier.color = NA)+
    geom_jitter(aes(color = FieldType), height = 0, width = .2)+
    labs(x = 'Field Type', y="Shannon's diversity index") +
    theme(legend.position = "none")+
    theme(axis.title.x= element_blank(),
        axis.title.y = element_text(size =10, color ='black'),
        axis.text.x = element_text(size =10, color ='black'),
        axis.text.y = element_text(size =8, color = 'black'))+
scale_color_manual(name = 'Field Type', labels = c('Native', 'Non-Virgin', 'Virgin'),
                  values = c('tomato', 'orange', 'chartreuse4'))+facet_wrap(~ location)
b
```



```
#dev.off()
```

Inference

```
fit_shannon = lmer(Shannon ~ FieldType + (1|location), data = diversity)
summary(fit_shannon)
```

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: Shannon ~ FieldType + (1 | location)
##    Data: diversity
##
## REML criterion at convergence: 11.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.1394 -0.3719  0.0452  0.6150  1.8008
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  location (Intercept) 0.00000  0.0000
##  Residual             0.06147  0.2479
## Number of obs: 54, groups:  location, 2
##
## Fixed effects:
##                     Estimate Std. Error       df t value Pr(>|t|)
## (Intercept)          1.50158    0.06199 51.00000  24.225   <2e-16 ***
## FieldTypeNon-virgin  0.06809    0.08413 51.00000   0.809    0.422
## FieldTypeVirgin      0.09452    0.08413 51.00000   1.123    0.266
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##             (Intr) FldTN-
## FldTypNn-vr -0.737
## FildTypVrgn -0.737  0.543
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

```
wilcox.observed = pairwise.wilcox.test(diversity$Shannon,
                     diversity$FieldType,
                     p.adjust.method = "BH")
wilcox.observed$p.value
```

```
##            Native Non-virgin
## Non-virgin 0.5622086         NA
## Virgin     0.4813020   0.481302
```

Beta diversity

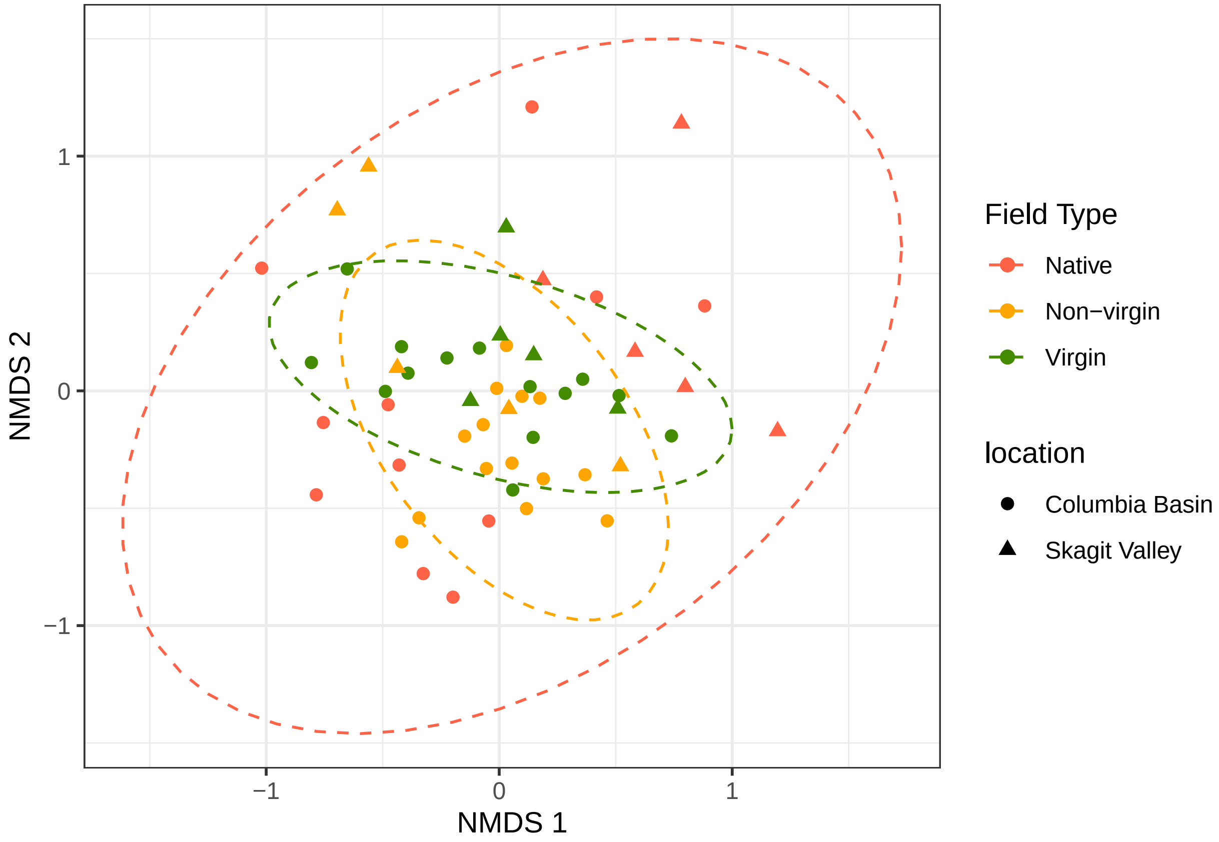
Ordination

```
ord = metaMDS(log_rare, distance = 'bray')
```

```
## Run 0 stress 0.2146699
## Run 1 stress 0.226462
## Run 2 stress 0.2283338
## Run 3 stress 0.2380314
## Run 4 stress 0.2173237
## Run 5 stress 0.2152331
## Run 6 stress 0.2360211
## Run 7 stress 0.2263862
## Run 8 stress 0.2468313
## Run 9 stress 0.2332766
## Run 10 stress 0.2338334
## Run 11 stress 0.2170365
## Run 12 stress 0.2359992
## Run 13 stress 0.2504821
## Run 14 stress 0.2178842
## Run 15 stress 0.2478469
## Run 16 stress 0.2317669
## Run 17 stress 0.2164967
## Run 18 stress 0.218306
## Run 19 stress 0.2268402
## Run 20 stress 0.2397831
## *** Best solution was not repeated -- monoMDS stopping criteria:
##     20: stress ratio > sratmax
```

```
options(repr.plot.width = 5, repr.plot.height = 3)
#jpeg('Beta_biplot.jpg',width=5,height=3,units='in',res=300)
beta = ggplot(labels, aes(x = ord$points[,1], y = ord$points[,2],
                          color = FieldType, shape = location))+
  geom_point(size = 2)+
  labs(x = 'NMDS 1', y="NMDS 2") +
  theme(axis.title.x= element_text(size =10, color ='black'),
        axis.title.y = element_text(size =10, color ='black'),
        axis.text.x = element_text(size =10, color ='black'),
        axis.text.y = element_text(size =8, color = 'black'))+
  scale_color_manual(name = 'Field Type', labels = c('Native', 'Non-virgin', 'Virgin'),
                     values = c('tomato', 'orange', 'chartreuse4'))+theme_bw()+
  stat_ellipse(aes(group = FieldType), linetype = 2)

beta
```



```
#dev.off()
```

Inference

```
dist = phyloseq::distance(log_rare, method="bray")
permanova_pairwise<-adonis(dist ~ labels$FieldType, strata = labels$location)
permanova_pairwise$aov.tab
```

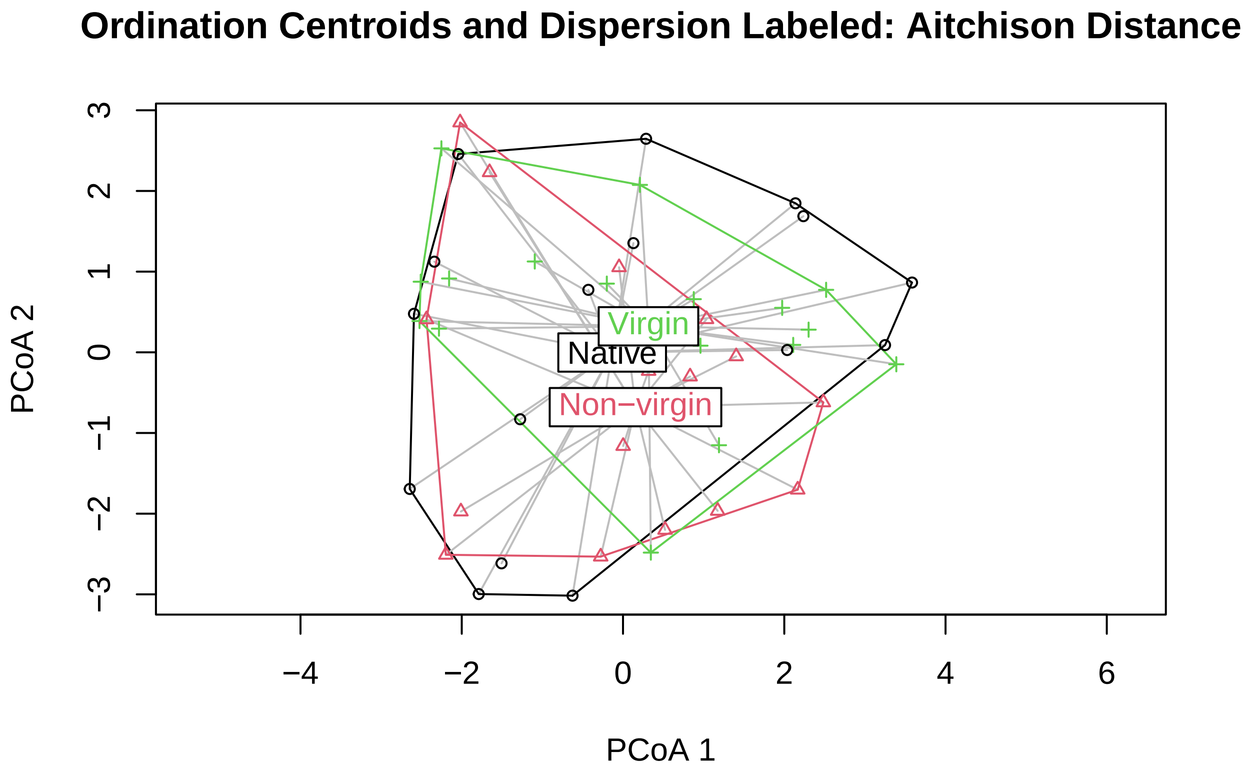
```
## Blocks:  strata
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##                  Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## labels$FieldType  2    0.6267 0.31334  1.5962 0.05891  0.063 .
## Residuals        51   10.0116 0.19631         0.94109
## Total            53   10.6383                 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Generate distance matrix
clr_dist_matrix <- phyloseq::distance(log_rare, method = "euclidean")

#ADONIS test
adon.results<-adonis(clr_dist_matrix ~ labels$FieldType, strata = labels$location)
adon.results$aov.tab
```

```
## Blocks:  strata
## Permutation: free
## Number of permutations: 999
##
## Terms added sequentially (first to last)
##
##                  Df SumsOfSqs MeanSqs F.Model      R2 Pr(>F)
## labels$FieldType  2     54.90  27.448  1.7652 0.06474  0.016 *
## Residuals        51    793.05  15.550         0.93526
## Total            53    847.95                 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
#Dispersion test and plot
dispr <- vegan::betadisper(clr_dist_matrix, labels$FieldType)
plot(dispr, main = "Ordination Centroids and Dispersion Labeled: Aitchison Distance", sub = "")
```



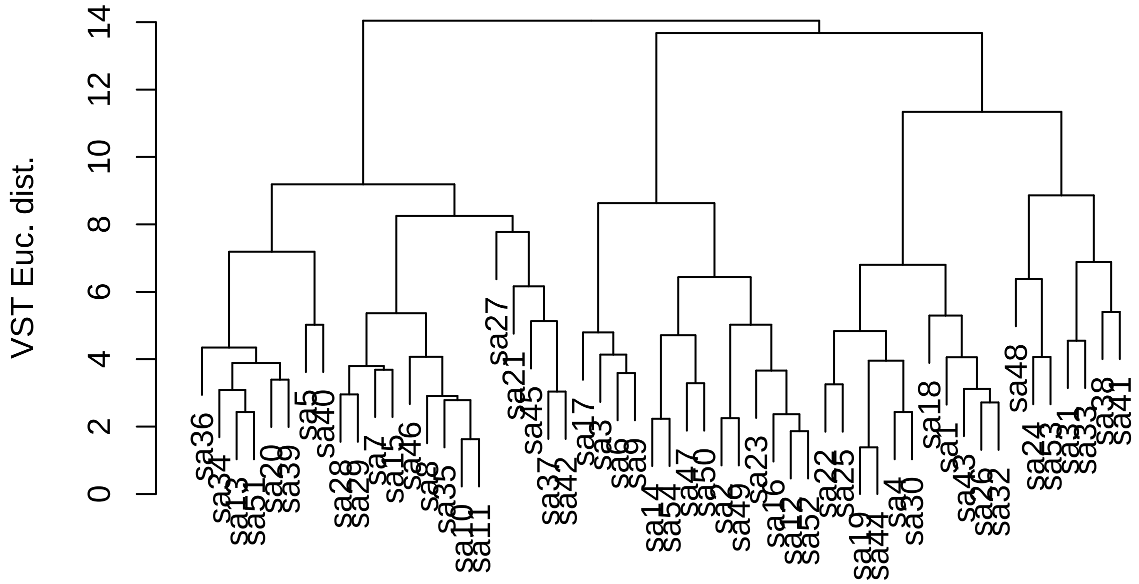
```
# and calculating our Euclidean distance matrix
euc_dist <- dist(log_rare)
length(euc_dist)
```

```
## [1] 1431
```

```
euc_clust <- hclust(euc_dist, method="ward.D2")

#plot(euc_clust) # hclust objects like this can be plotted as is, but i like to change them to dendrogr
  # 1) it's easier to color the dendrogram plot by groups
  # 2) if you want you can rotate clusters with the rotate() function of the dendextend package

euc_dend <- as.dendrogram(euc_clust, hang=0.1)
plot(euc_dend, ylab="VST Euc. dist.")
```


VST Euc. dist.
sa36
sa34
sa13
sa51
sa20
sa39
sa5
sa40
sa28
sa29
sa7
sa15
sa46
sa8
sa35
sa10
sa11
sa27
sa21
sa45
sa37
sa42
sa17
sa3
sa6
sa9
sa14
sa54
sa47
sa50
sa2
sa49
sa23
sa16
sa12
sa52
sa22
sa25
sa19
sa44
sa4
sa30
sa18
sa1
sa43
sa26
sa32
sa48
sa24
sa53
sa31
sa33
sa38
sa41