T7 warmx 16S Analysis 2021

Moriah Young

2023-01-31

```
##  [1] "2022 REX Peak Drought Post-processing copy.gsheet"
##  [2] "animal"
##  [3] "crop_yields"
##  [4] "Data Workshop Pre-Survey.gform"
##  [5] "expected"
##  [6] "Falvo"
##  [7] "GHG"
##  [8] "Irrigation"
##  [9] "lookup_tables"
## [10] "microbes"
## [11] "Post drought post processing copy.gsheet"
## [12] "REX_Data_Management.gdoc"
## [13] "REX_Data_Template.gsheet"
## [14] "REX_Microbial_sampling_IDs_complete.gsheet"
## [15] "REX_SampleArchive.gsheet"
## [16] "REX_stats_basics.Rmd"
## [17] "REX_T7_metadata.csv"
## [18] "REX_T7_metadata.gsheet"
## [19] "REX_template.csv"
## [20] "REX_warmx_metadata.csv"
## [21] "REX_warmx_metadata.xlsx"
## [22] "REX_warmx_Soca_ID_metadata_2021.csv"
## [23] "REX_warmx_Soca_ID_metadata_2021.gsheet"
## [24] "REX_warmx_taxon.csv"
## [25] "REX_warmx_taxon.gsheet"
## [26] "sampling-notes"
## [27] "SamplingDemands_Plants.gsheet"
## [28] "SamplingDemands_Soil.xlsx"
## [29] "SamplingSchedule_T7Plants.gsheet"
## [30] "sensors"
## [31] "soil"
## [32] "T2"
## [33] "T7_ANPP"
## [34] "T7_plant_comp"
## [35] "T7_plant_phenology"
## [36] "T7_warmx_insect"
## [37] "T7_warmx_plant_traits"
## [38] "T7_warmx_VOC"
## [39] "weather"
```

Read QZA files into dataframe, re-format taxonomic tables, and re-upload them as .csv files

```
# Code below you only need to do once so it's # out as to not have to run every time
#SVs16S <- read_qza("/Users/moriahyoung/Downloads/16S-2021-merged-dada2table.qza")
#SVs16Stable <- SVs16S$data
#write.csv(SVs16Stable, file = "/Users/moriahyoung/Downloads/16S-merged-dada2table.csv")

#taxonomy16S <- read_qza("/Users/moriahyoung/Downloads/16S-taxonomy.qza")
#tax16S <- taxonomy16S$data %>% as_tibble() %>%
#  mutate(Taxon=gsub("[a-z]__", "", Taxon)) %>%
#  separate(Taxon, sep=";", c("Kingdom","Phylum","Class","Order","Family","Genus","Species"))%>%
#  mutate(Phylum=replace_na(Phylum,"empty"))
#write.csv(tax16S, file = "/Users/moriahyoung/Downloads/16S-taxonomy.csv", row.names =F)

physeq_tree <- qza_to_phyloseq(
        tree = "/Users/moriahyoung/Downloads/16S-2021-fasttree-rooted-tree.qza"
        )
physeq_tree
```

```
##
## Phylogenetic tree with 58895 tips and 58640 internal nodes.
##
## Tip labels:
##   fcd87e7de7bc3786b4dae4854f0167d4, fe4b0b291d258dac0ce14c06183a0024, 38572562126e672bddb2cc4cdce7b5
## Node labels:
##   root, 0.991, 0.193, 0.893, 0.925, 0.983, ...
##
## Rooted; includes branch lengths.
```

```
# create phyloseq object
#data_16S_unfiltered <- read_csv2phyloseq(otu.file = "/Users/moriahyoung/Downloads/16S-merged-dada2tabl
#                                         taxonomy.file = "/Users/moriahyoung/Downloads/16S-taxonomy.cs
#                                         metadata.file = "/Users/moriahyoung/Desktop/16S_2021_metadata
#summarize_phyloseq(data_16S_unfiltered)
```

```
# filter data

# filter out non bacteria
#data_16S_uf1 <- subset_taxa(data_16S_unfiltered, Kingdom == "Bacteria" | Kingdom == "Archaea")
#data_16S_uf2 <- subset_taxa(data_16S_uf1, Kingdom != "Eukaryota")
#data_16S_uf3 <- subset_taxa(data_16S_uf2, Order != "Chloroplast")
#data_16S_uf4 <- subset_taxa(data_16S_uf3, Family != "Mitochondria")

#summarize_phyloseq(data_16S_uf4)

# Remove samples with extremely low read depth
# data_16S_uf5 <- prune_samples(sample_sums(data_16S_uf4)>=1000, data_16S_uf4)

# Export filtered data
# write.csv(data_16S_uf5@otu_table, "/Users/moriahyoung/Desktop/16S-merged-table-filtered.csv")
# write.csv(data_16S_uf5@tax_table, "/Users/moriahyoung/Desktop/16S-taxonomy-filtered.csv")
# upload these to REX google shared drive

# Create phyloseq objects of filtered data (16S and ITS)
data_16S_filtered <- read_csv2phyloseq(otu.file = "/Users/moriahyoung/Desktop/16S-merged-table-filtered
                                       taxonomy.file = "/Users/moriahyoung/Desktop/16S-taxonomy-filtere
                                       metadata.file = "/Users/moriahyoung/Desktop/16S_2021_metadata.cs
summarize_phyloseq(data_16S_filtered)
```

```
## Compositional = NO2

## 1] Min. number of reads = 24882] Max. number of reads = 2029793] Total number of reads = 78936004] A
##         (i.e. exactly one read detected across all samples)0.016915067566297710] Number of sample va

## [[1]]
## [1] "1] Min. number of reads = 2488"
##
## [[2]]
## [1] "2] Max. number of reads = 202979"
##
## [[3]]
## [1] "3] Total number of reads = 7893600"
##
## [[4]]
## [1] "4] Average number of reads = 37950"
##
## [[5]]
## [1] "5] Median number of reads = 27683.5"
##
## [[6]]
## [1] "7] Sparsity = 0.981261141174311"
##
## [[7]]
## [1] "6] Any OTU sum to 1 or less? YES"
##
## [[8]]
## [1] "8] Number of singletons = 551"
##
## [[9]]
## [1] "9] Percent of OTUs that are singletons \n        (i.e. exactly one read detected across all sam
##
## [[10]]
## [1] "10] Number of sample variables are: 17"
##
## [[11]]
##  [1] "RSTF_SampleID"            "Barcode"
##  [3] "LinkerPrimerSequence"     "Reverse_Primer"
##  [5] "MiSeqRun"                 "Treatment"
##  [7] "Replicate"                "Rep"
##  [9] "Footprint_Treatment_full" "Footprint"
## [11] "Footprint_Location"       "Subplot"
## [13] "Subplot_Location"         "Subplot_Descriptions"
## [15] "Unique_ID"                "Drought"
## [17] "Datetime_UTC"
```

```
# merge the phyloseq object with the phylogenetic tree with the other phyloseq object
data_16S_filtered <- merge_phyloseq(data_16S_filtered, physeq_tree)
summarize_phyloseq(data_16S_filtered)
```

```
## Compositional = NO2
## 1] Min. number of reads = 24882] Max. number of reads = 2029793] Total number of reads = 78936004] Av
##         (i.e. exactly one read detected across all samples)0.016915067566297710] Number of sample va
```

```
## [[1]]
## [1] "1] Min. number of reads = 2488"
##
## [[2]]
## [1] "2] Max. number of reads = 202979"
##
## [[3]]
## [1] "3] Total number of reads = 7893600"
##
## [[4]]
## [1] "4] Average number of reads = 37950"
##
## [[5]]
## [1] "5] Median number of reads = 27683.5"
##
## [[6]]
## [1] "7] Sparsity = 0.981261141174311"
##
## [[7]]
## [1] "6] Any OTU sum to 1 or less? YES"
##
## [[8]]
## [1] "8] Number of singletons = 551"
##
## [[9]]
## [1] "9] Percent of OTUs that are singletons \n        (i.e. exactly one read detected across all samp
##
## [[10]]
## [1] "10] Number of sample variables are: 17"
##
## [[11]]
##  [1] "RSTF_SampleID"            "Barcode"
##  [3] "LinkerPrimerSequence"     "Reverse_Primer"
##  [5] "MiSeqRun"                 "Treatment"
##  [7] "Replicate"                "Rep"
##  [9] "Footprint_Treatment_full" "Footprint"
## [11] "Footprint_Location"       "Subplot"
## [13] "Subplot_Location"         "Subplot_Descriptions"
## [15] "Unique_ID"                "Drought"
## [17] "Datetime_UTC"
```

```r
# order sample data - "Drought" and "Subplot_Descriptions"
sample_data(data_16S_filtered)$Drought <- ordered(sample_data(data_16S_filtered)$Drought, c("Pre-Drought
sample_data(data_16S_filtered)$Subplot_Descriptions <- ordered(sample_data(data_16S_filtered)$Subplot_De
```

Rarefaction Curves

```
data_16S_counts <- data_16S_filtered
# remove samples that have less than 1000 reads
data_16S_counts <- prune_samples(sample_sums(data_16S_counts)>=1000, data_16S_counts) # this is already
#data_16S_filtered <- transform_sample_counts(data_16S_filtered, function(x) x/sum(x))
# Prune SVs that are not present 5 times in at least 2 samples --> remove taxa not seen more than 5 tim
#data_16S_counts <- filter_taxa(data_16S_counts, function(x) sum(x >5) > (0.01058201*length(x)), TRUE)
summarize_phyloseq(data_16S_counts)
```

```
## Compositional = NO2
```

```
## 1] Min. number of reads = 24882] Max. number of reads = 2029793] Total number of reads = 78936004] A
##         (i.e. exactly one read detected across all samples)0.016915067566297710] Number of sample va
```

```
## [[1]]
## [1] "1] Min. number of reads = 2488"
##
## [[2]]
## [1] "2] Max. number of reads = 202979"
##
## [[3]]
## [1] "3] Total number of reads = 7893600"
##
## [[4]]
## [1] "4] Average number of reads = 37950"
##
## [[5]]
## [1] "5] Median number of reads = 27683.5"
##
## [[6]]
## [1] "7] Sparsity = 0.981261141174311"
##
## [[7]]
## [1] "6] Any OTU sum to 1 or less? YES"
##
## [[8]]
## [1] "8] Number of singletons = 551"
##
## [[9]]
## [1] "9] Percent of OTUs that are singletons \n        (i.e. exactly one read detected across all samp
##
## [[10]]
## [1] "10] Number of sample variables are: 17"
##
## [[11]]
##  [1] "RSTF_SampleID"            "Barcode"
##  [3] "LinkerPrimerSequence"     "Reverse_Primer"
##  [5] "MiSeqRun"                 "Treatment"
##  [7] "Replicate"                "Rep"
##  [9] "Footprint_Treatment_full" "Footprint"
## [11] "Footprint_Location"       "Subplot"
## [13] "Subplot_Location"         "Subplot_Descriptions"
## [15] "Unique_ID"                "Drought"
## [17] "Datetime_UTC"
```

```
sum(colSums(otu_table(data_16S_counts)))
```

```
## [1] 7893600
```

```
sort(colSums(otu_table(data_16S_counts)))
```

```
## REX_729 REX_611      38 REX_657 REX_828 REX_771      46 REX_604 REX_293 REX_272
##    2488    2600    3083    3418    4201    4811    5477    5593    5689    6124
## REX_832 REX_601 REX_707 REX_330      40 REX_383 REX_512      45 REX_727 REX_728
##    6263    7017    7283    7485    7680    7718    8632    8646    8927    9617
## REX_506 REX_824 REX_557 REX_720 REX_674      24 REX_830 REX_615 REX_380 REX_786
##    9884    9934   10959   10986   11063   11273   11509   11612   11623   11680
## REX_444 REX_498      22 REX_505 REX_845 REX_295      18       8 REX_600 REX_614
##   12374   13098   13234   13471   13747   13845   14210   14442   14612   14620
## REX_228 REX_783 REX_455      31 REX_292 REX_387 REX_719 REX_281 REX_709      10
##   14853   14916   15008   15241   15401   15588   15966   16068   16308   16317
##      14 REX_399      27 REX_297 REX_736 REX_400 REX_829      41 REX_386 REX_844
##   16590   16612   16870   16912   17508   17569   17691   17701   17771   17861
##       1 REX_507 REX_284      43 REX_388 REX_610 REX_675      42 REX_394 REX_329
##   18084   18461   18609   18612   18713   18986   19003   19057   19422   19700
## REX_283      37 REX_332 REX_344 REX_718 REX_238       4      12 REX_404 REX_784
##   19781   19822   20125   20136   20178   20235   20387   20504   20821   21090
## REX_509 REX_560 REX_237 REX_841 REX_616 REX_673 REX_499      19 REX_497 REX_500
##   21238   21777   21817   22116   22187   22343   23011   23126   23716   23861
## REX_508 REX_384 REX_773 REX_381      44 REX_734 REX_510 REX_850 REX_271       9
##   23873   23957   24100   24791   25272   26293   26334   26490   26531   26544
## REX_785 REX_735 REX_331       2 REX_223 REX_453 REX_296      13      28 REX_333
##   27520   27646   27663   27681   27686   27741   27949   27956   28511   28817
## REX_437 REX_342 REX_398 REX_558      29 REX_547 REX_609 REX_495      25 REX_596
##   28846   29269   29445   29711   30510   30796   30882   31051   31266   31282
## REX_269 REX_239      20       6 REX_598      17      39 REX_839 REX_240 REX_774
##   31810   32175   32324   32567   32747   33738   33983   34652   35759   35886
## REX_289 REX_602 REX_599 REX_548 REX_620 REX_672      48 REX_270      15 REX_546
##   35960   36748   37832   38983   39021   39314   39467   39609   39940   39991
## REX_733 REX_385 REX_612 REX_842 REX_772 REX_487 REX_279 REX_613 REX_706 REX_559
##   40450   40594   40860   41017   41126   41379   41482   41997   42903   43116
##      23      34 REX_716 REX_488 REX_597 REX_294 REX_224 REX_549 REX_396 REX_838
##   43153   43597   44367   45735   46805   47385   48121   48380   49512   50239
## REX_658 REX_343 REX_486 REX_456 REX_438 REX_840 REX_770 REX_656 REX_454 REX_290
##   50444   50792   50951   52710   53036   53036   55832   55886   57511   59061
## REX_222 REX_511 REX_708 REX_843 REX_439 REX_732 REX_382       5      21 REX_662
##   60177   61137   62014   63893   64846   65220   65612   66800   68385   68801
## REX_655      47 REX_721      35 REX_291 REX_393 REX_485      32 REX_440 REX_827
##   72087   72442   73567   74166   74553   75174   77403   78118   81627   86534
##      26       3 REX_603       7 REX_395 REX_513 REX_825 REX_826 REX_282 REX_221
##   87883   88787   90066   91303   93695   94294   97560   99357   99625  115718
##      16 REX_730      33      11 REX_545 REX_397 REX_341      30
##  122571  123822  142973  147011  159575  162334  164498  202979
```

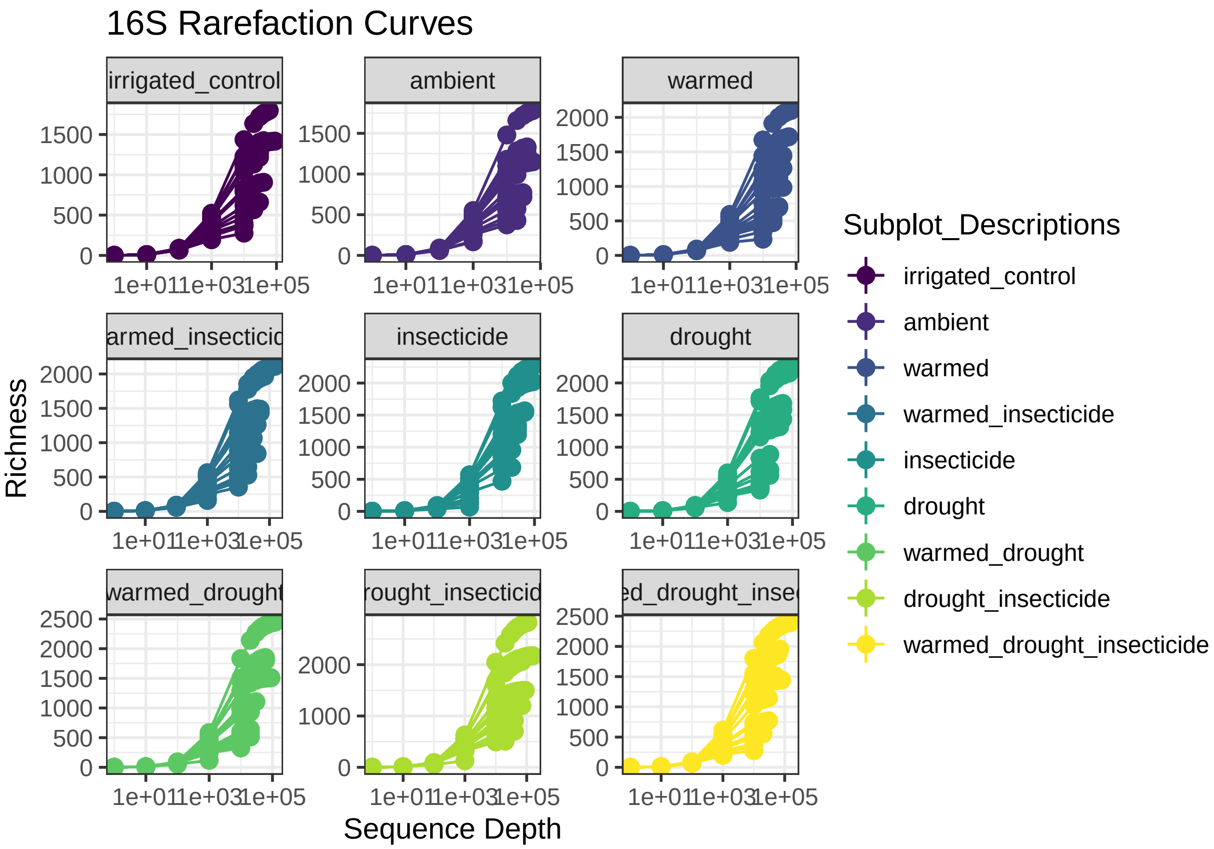
```
# min = 2488
# max = 202,979

# https://github.com/joey711/phyloseq/issues/143
# Rarefaction Curve Function
calculate_rarefaction_curves <- function(psdata, measures, depths, parallel=T) {
        require('plyr') # ldply
        require('reshape2') # melt
        require('doParallel')

# set parallel options if required
if (parallel) {
        paropts  <- list(.packages=c("phyloseq", "reshape2"))
        } else {
                paropts  <- NULL
                }

estimate_rarified_richness <- function(psdata, measures, depth) {
    if(max(sample_sums(psdata)) < depth) return()
    psdata <- prune_samples(sample_sums(psdata) >= depth, psdata)

    rarified_psdata <- rarefy_even_depth(psdata, depth, verbose = FALSE)

    alpha_diversity <- estimate_richness(rarified_psdata, measures = measures)

    # as.matrix forces the use of melt.array, which includes the Sample names (rownames)
    molten_alpha_diversity <- melt(as.matrix(alpha_diversity), varnames = c('Sample', 'Measure'), value

    molten_alpha_diversity
  }

  names(depths) <- depths # this enables automatic addition of the Depth to the output by ldply
  rarefaction_curve_data <- ldply(depths, estimate_rarified_richness, psdata = psdata, measures = measu

  # convert Depth from factor to numeric
  rarefaction_curve_data$Depth <- as.numeric(levels(rarefaction_curve_data$Depth))[rarefaction_curve_da

  rarefaction_curve_data
}

# Summarize alpha diversity
rarefaction_curve_data <- calculate_rarefaction_curves(data_16S_counts, c('Observed', 'Shannon'),
                                                       rep(c(1, 10, 100, 1000, 1:100 * 10000), each = 1
```

```
## Loading required package: reshape2

##
## Attaching package: 'reshape2'

## The following object is masked from 'package:tidyr':
##
##     smiths
## The following objects are masked from 'package:data.table':
##
##     dcast, melt
```

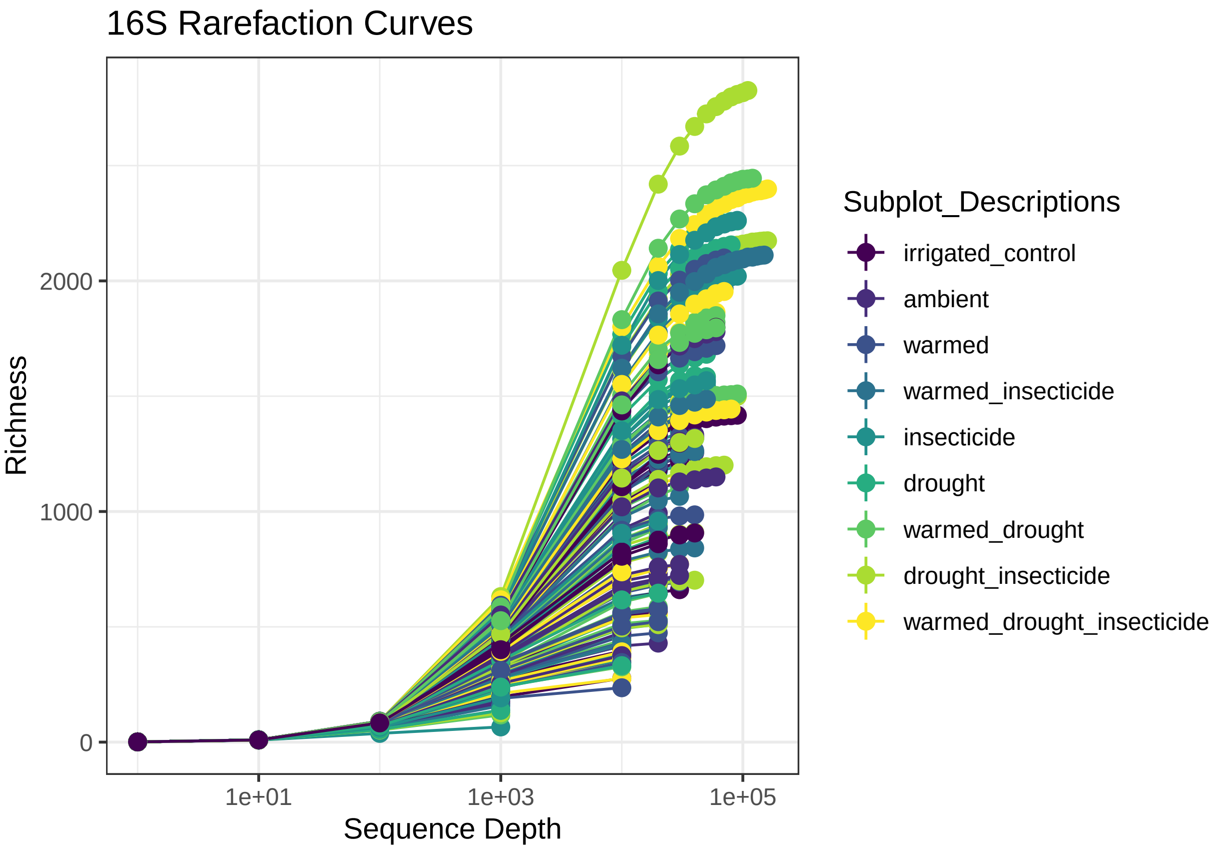
```
## Warning in setup_parallel(): No parallel backend registered
```

```
## Warning: executing %dopar% sequentially: no parallel backend registered
```

```
rarefaction_curve_data_summary <- ddply(rarefaction_curve_data, c('Depth', 'Sample', 'Measure'), summari

# Add sample data
rarefaction_curve_data_summary_verbose <-  merge(rarefaction_curve_data_summary %>% mutate(Sample = gsul
                                                 data.frame(sample_data(data_16S_counts)) %>%
                                                         rownames_to_column(var = "rowname"),
                                                 by.x = 'Sample', by.y = 'rowname')

discrete_palettes <- list(RColorBrewer::brewer.pal(3, "Set2"))

rarefaction_curve_data_summary_verbose$Subplot_Descriptions <- factor(rarefaction_curve_data_summary_ve

# plot
curve_16S_facet <-
        ggplot(rarefaction_curve_data_summary_verbose %>% filter(Measure == "Observed"), aes(x = Depth,
        geom_line(alpha=1) +
        geom_pointrange() +
        scale_x_continuous(trans = "log10", name = "Sequence Depth") +
        ylab("Richness") +
        facet_wrap(~Subplot_Descriptions, scales = 'free') +
        theme(legend.text = element_text(hjust = 0)) +
        labs(title = "16S Rarefaction Curves") +
        theme_bw()

curve_16S_facet
```



```
curve_16S <- ggplot(rarefaction_curve_data_summary_verbose %>% filter(Measure == "Observed"), aes(x = De
        geom_line(alpha=1) +
        geom_pointrange() +
        scale_x_continuous(trans = "log10", name = "Sequence Depth") +
        ylab("Richness") +
        #facet_wrap(~Subplot_Descriptions, scales = 'free') +
        theme(legend.text = element_text(hjust = 0)) +
        labs(title = "16S Rarefaction Curves") +
        theme_bw()

curve_16S
```



#Stacked Bar Plots - Phylum/Family/Genus levels

#Taxonomy - Phylum Level

```
# 16S Habitat Overview Plot
# Rarefy data to 1000 reads
set.seed(01221990)
rare_16S <- rarefy_even_depth(data_16S_counts, sample.size = 1000)
```

```
## You set `rngseed` to FALSE. Make sure you've set & recorded
##  the random seed of your session for reproducibility.
## See `?set.seed`

## ...

## 38775OTUs were removed because they are no longer
## present in any sample after random subsampling

## ...
```

```
data_16S_rel <- transform_sample_counts(rare_16S, function(x) x/sum(x))

phylum_sum_16S <- tapply(taxa_sums(data_16S_rel),
                         tax_table(data_16S_rel)[, "Phylum"], sum, na.rm=TRUE)
top10Phylum_16S <- names(sort(phylum_sum_16S, TRUE))[1:10]
```

```
abundant_16S_phylum <- subset_taxa(data_16S_rel, Phylum %in% top10Phylum_16S)
abundant_16S_phylum_glom <- tax_glom(abundant_16S_phylum, taxrank = "Phylum")
abundant_16S_phylum_melt <- psmelt(abundant_16S_phylum_glom)
```

```
## Warning in psmelt(abundant_16S_phylum_glom): The rank names:
## OTU
##  have been renamed to:
## taxa_OTU
## to avoid conflicts with special phyloseq plot attribute names.
```

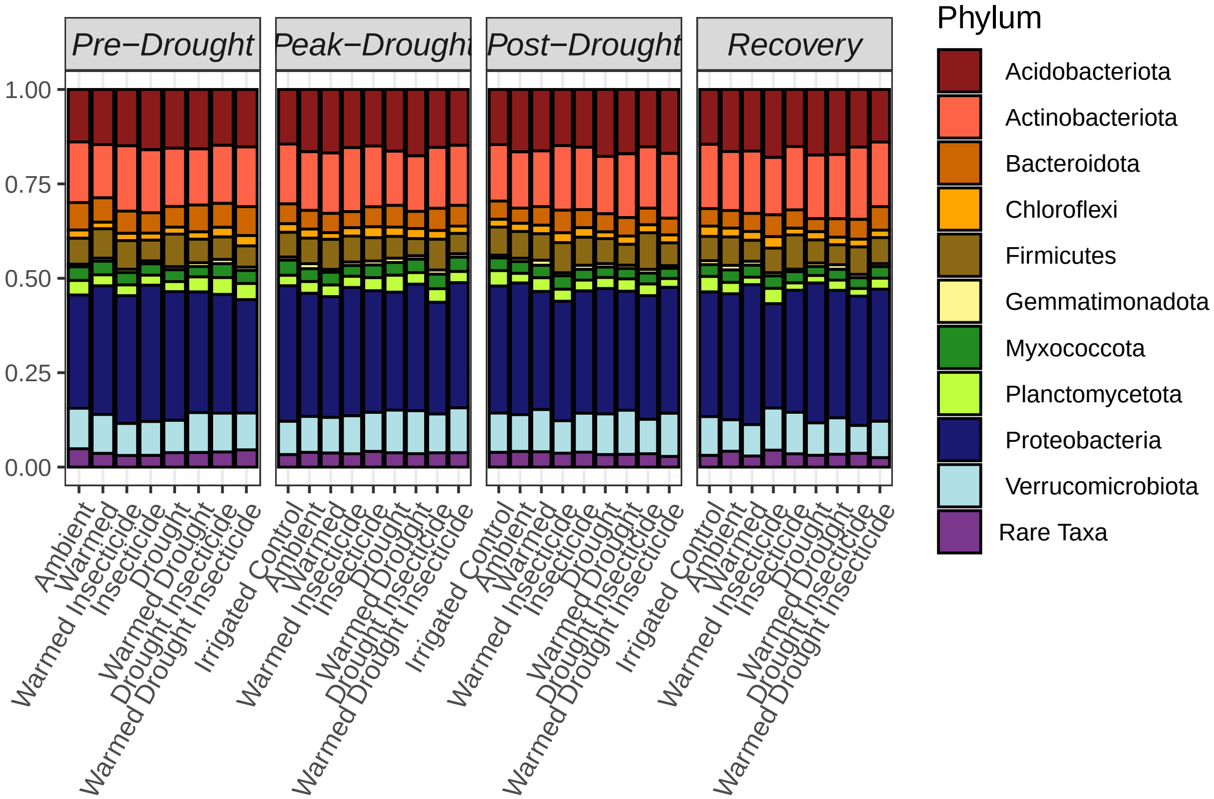
```
sum2 <- abundant_16S_phylum_melt %>%
  group_by(Subplot_Descriptions, Drought, Phylum) %>%
  summarise(means = mean(Abundance))
```

```
## `summarise()` has grouped output by 'Subplot_Descriptions', 'Drought'. You can
## override using the `.groups` argument.
```

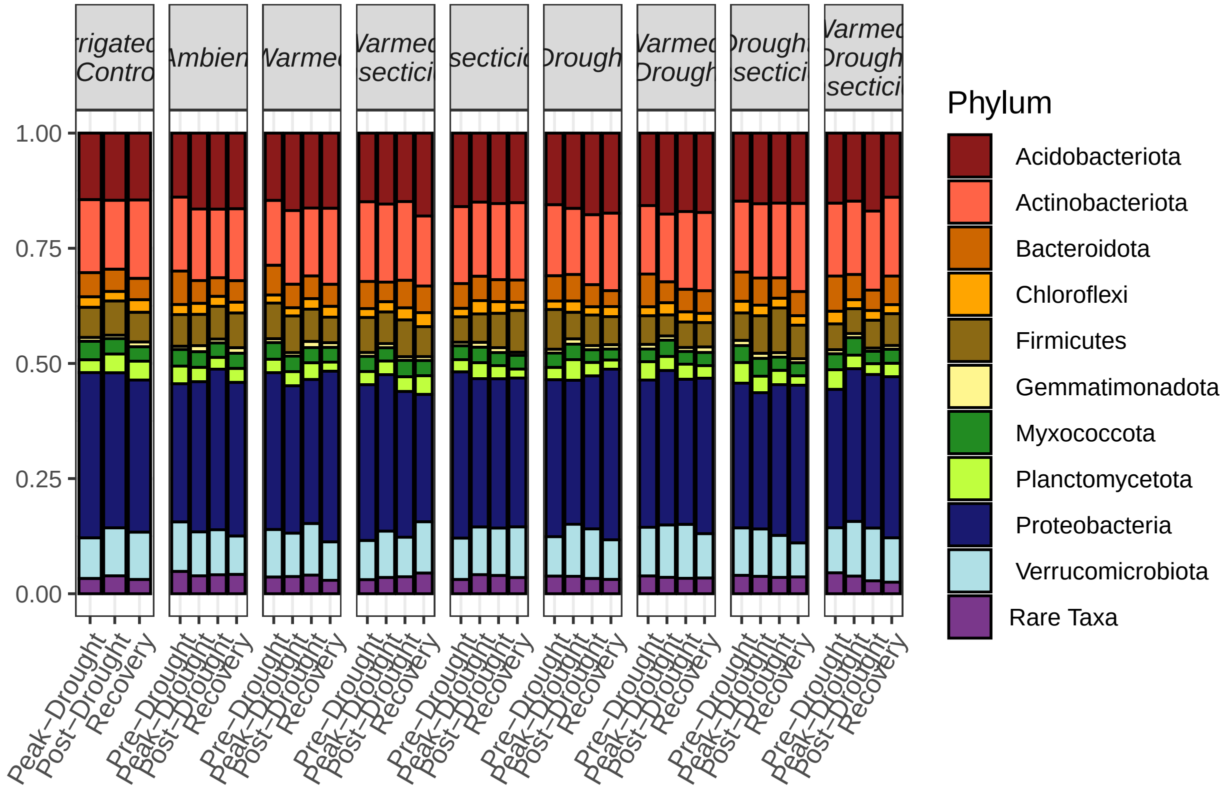
```
#Everything that's left is considered rare
rare <- sum2 %>%
  group_by(Subplot_Descriptions, Drought) %>%
  summarise(means = 1- sum(means)) %>%
  mutate(Phylum = "Rare Taxa")
```

```
## `summarise()` has grouped output by 'Subplot_Descriptions'. You can override
## using the `.groups` argument.
```

```
#concatenate the datasets
sum2 = rbind(sum2, rare)

#order groups
sum2$Phylum <- forcats::fct_relevel(sum2$Phylum, "Rare Taxa", after = Inf)

# Stacked bar plot for top phylas
# facet by timing of soil collection (aka "Drought")
#png("T7_warmx_16S_phylum_stacked_bar_plot_by_drought.png", units="in", width=10, height=6, res=300)
ggplot(sum2, aes(x = Subplot_Descriptions, y = means, fill = Phylum)) +
        geom_bar(position = "stack", stat = "identity", col = "black") +
        facet_grid(~Drought, scale="free") +
        theme(legend.position = "right", legend.title = element_blank(), axis.line = element_blank()) +
        scale_fill_manual(values = c("firebrick4","tomato", "darkorange3", "orange1" ,"goldenrod4", "kha
                                     "forestgreen", "olivedrab1", "midnightblue", "powderblue", "medium
        scale_x_discrete(labels=c("ambient" = "Ambient",
                                  "drought" = "Drought",
                                  "irrigated_control" = "Irrigated Control",
                                  "warmed" = "Warmed",
                                  "warmed_drought" = "Warmed Drought",
                                  "drought_insecticide" = "Drought Insecticide",
                                  "insecticide" = "Insecticide",
                                  'warmed_drought_insecticide' = "Warmed Drought Insecticide",
                                  'warmed_insecticide' = "Warmed Insecticide")) +
        #scale_y_continuous(name = NULL, breaks = NULL) +
        panel_border() +
        theme_bw() +
        ggtitle("") +
        theme(axis.text.x = element_text(angle = 60, hjust = 1)) +
        theme(strip.text = element_text(size = 12, face = "italic"), axis.text.x = element_text(size = 
              axis.title = element_blank(),
              title = element_text(size = 12))
```



```
#dev.off()

treatment_names <- c("ambient" = "Ambient",
                     "drought" = "Drought",
                     "irrigated_control" = "Irrigated \n Control",
                     "warmed" = "Warmed",
                     "warmed_drought" = "Warmed \n Drought",
                     "drought_insecticide" = "Drought \n Insecticide",
                     "insecticide" = "Insecticide",
                     'warmed_drought_insecticide' = "Warmed \n Drought \n Insecticide",
                     'warmed_insecticide' = "Warmed \n Insecticide")

# facet by treatment
#png("T7_warmx_16S_phylum_stacked_bar_plot_by_treatment.png", units="in", width=10, height=6, res=300)
ggplot(sum2, aes(x = Drought, y = means, fill = Phylum)) +
        geom_bar(position = "stack", stat = "identity", col = "black") +
        facet_grid(~Subplot_Descriptions, scale="free", labeller = as_labeller(treatment_names)) +
        theme(legend.position = "right", legend.title = element_blank(), axis.line = element_blank()) +
        scale_fill_manual(values = c("firebrick4","tomato", "darkorange3", "orange1" ,"goldenrod4", "kh
                                     "forestgreen", "olivedrab1", "midnightblue", "powderblue", "medium
         scale_x_discrete(labels=c("ambient" = "Ambient",
                                  "drought" = "Drought",
                                  "irrigated_control" = "Irrigated Control",
                                  "warmed" = "Warmed",
                                  "warmed_drought" = "Warmed Drought",
                                  "drought_insecticide" = "Drought Insecticide",
                                  "insecticide" = "Insecticide",
                                  'warmed_drought_insecticide' = "Warmed Drought Insecticide",
                                  'warmed_insecticide' = "Warmed Insecticide")) +
        #scale_y_continuous(name = NULL, breaks = NULL) +
        panel_border() +
        theme_bw() +
        ggtitle("") +
        theme(axis.text.x = element_text(angle = 60, hjust = 1)) +
        theme(strip.text = element_text(size = 10, face = "italic"), axis.text.x = element_text(size = 
              axis.title = element_blank(),
              title = element_text(size = 12))
```



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

#Taxonomy - Family Level

```
# 16S Habitat Overview Plot
# Rarefy data to 1000 reads
set.seed(01221990)
```

```r
family_sum_16S <- tapply(taxa_sums(data_16S_rel),
                         tax_table(data_16S_rel)[, "Family"], sum, na.rm=TRUE)
top10Family_16S <- names(sort(family_sum_16S, TRUE))[1:10]

abundant_16S_family <- subset_taxa(data_16S_rel, Family %in% top10Family_16S)
abundant_16S_family_glom <- tax_glom(abundant_16S_family, taxrank = "Family")
abundant_16S_family_melt <- psmelt(abundant_16S_family_glom)
```

```
## Warning in psmelt(abundant_16S_family_glom): The rank names:
## OTU
##  have been renamed to:
## taxa_OTU
## to avoid conflicts with special phyloseq plot attribute names.
```

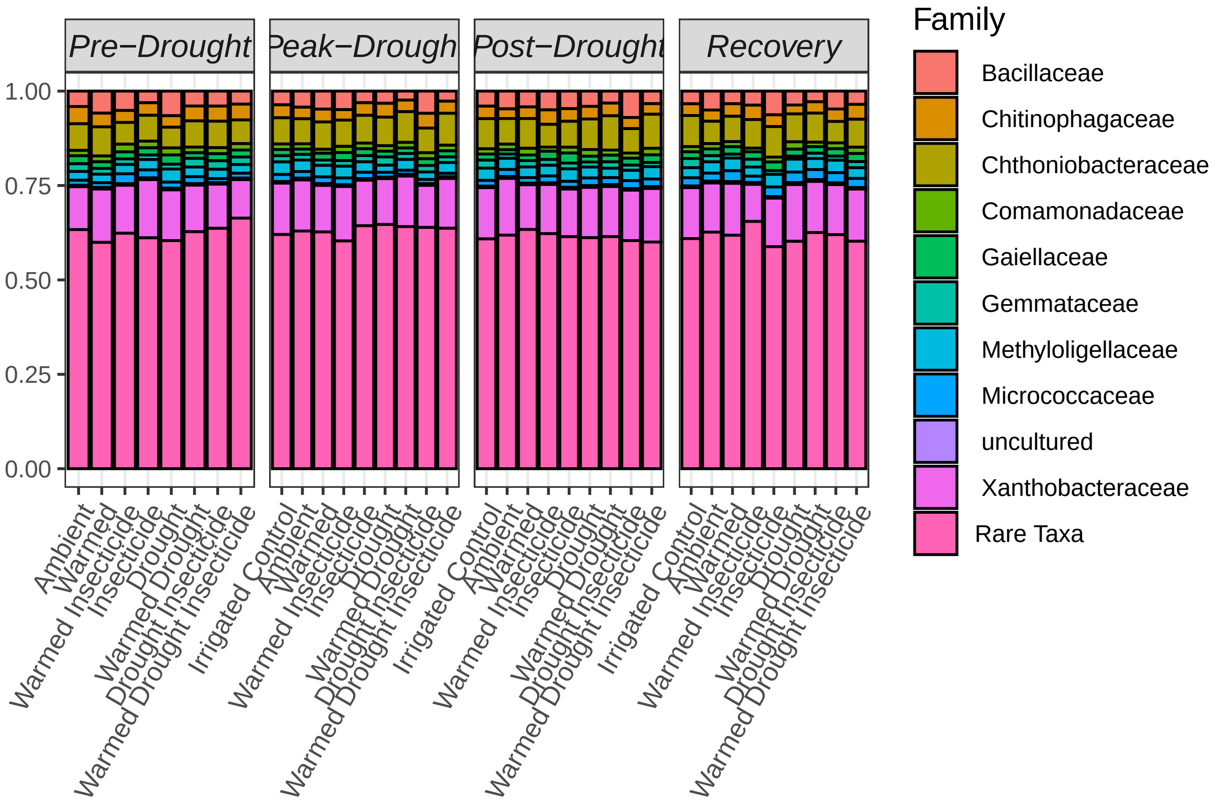
```r
sum3 <- abundant_16S_family_melt %>%
  group_by(Subplot_Descriptions, Drought, Family) %>%
  summarise(means = mean(Abundance))
```

```
## `summarise()` has grouped output by 'Subplot_Descriptions', 'Drought'. You can
## override using the `.groups` argument.
```

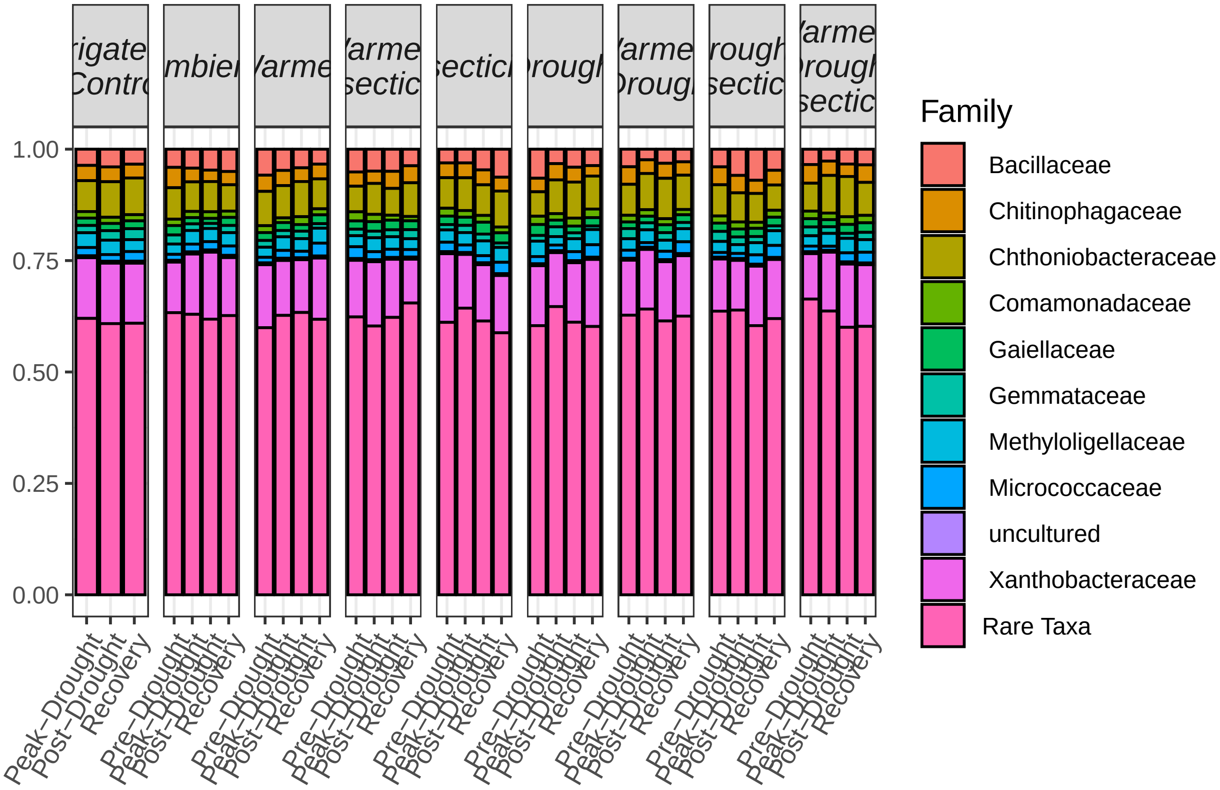
```r
#Everything that's left is considered rare
rare <- sum3 %>%
  group_by(Subplot_Descriptions, Drought) %>%
  summarise(means = 1- sum(means)) %>%
  mutate(Family = "Rare Taxa")
```

```
## `summarise()` has grouped output by 'Subplot_Descriptions'. You can override
## using the `.groups` argument.
```

```r
#concatenate the datasets
sum3 = rbind(sum3, rare)

#order groups
sum3$Family <- forcats::fct_relevel(sum3$Family, "Rare Taxa", after = Inf)

# Stacked bar plot for top phylas
#png("T7_warmx_16S_family_stacked_bar_plot.png", units="in", width=10, height=6, res=300)
ggplot(sum3, aes(x = Subplot_Descriptions, y = means, fill = Family)) +
        geom_bar(position = "stack", stat = "identity", col = "black") +
        facet_grid(~Drought, scale="free") +
        theme(legend.position = "right", legend.title = element_blank(), axis.line = element_blank()) +
        #scale_fill_manual(values = c("firebrick4","tomato", "darkorange3", "orange1" ,"goldenrod4", "k
        #                              "forestgreen", "olivedrab1", "midnightblue", "powderblue", "mediu
        scale_x_discrete(labels=c("ambient" = "Ambient",
                                  "drought" = "Drought",
                                  "irrigated_control" = "Irrigated Control",
                                  "warmed" = "Warmed",
                                  "warmed_drought" = "Warmed Drought",
                                  "drought_insecticide" = "Drought Insecticide",
                                  "insecticide" = "Insecticide",
                                  'warmed_drought_insecticide' = "Warmed Drought Insecticide",
                                  'warmed_insecticide' = "Warmed Insecticide")) +
        #scale_y_continuous(name = NULL, breaks = NULL) +
        panel_border() +
        theme_bw() +
        ggtitle("") +
        theme(axis.text.x = element_text(angle = 60, hjust = 1)) +
        theme(strip.text = element_text(size = 12, face = "italic"), axis.text.x = element_text(size = 
              axis.title = element_blank(),
              title = element_text(size = 12))
```



```
#dev.off()

# facet_grid by treatment aka "Subplot_Descriptions"
#png("T7_warmx_16S_family_stacked_bar_plot_by_treatment.png", units="in", width=10, height=6, res=300)
ggplot(sum3, aes(x = Drought, y = means, fill = Family)) +
        geom_bar(position = "stack", stat = "identity", col = "black") +
        facet_grid(~Subplot_Descriptions, scale="free", labeller = as_labeller(treatment_names)) +
        theme(legend.position = "right", legend.title = element_blank(), axis.line = element_blank()) +
        #scale_fill_manual(values = c("firebrick4","tomato", "darkorange3", "orange1" ,"goldenrod4", "k
        #                             "forestgreen", "olivedrab1", "midnightblue", "powderblue", "mediu
        scale_x_discrete(labels=c("ambient" = "Ambient",
                                  "drought" = "Drought",
                                  "irrigated_control" = "Irrigated Control",
                                  "warmed" = "Warmed",
                                  "warmed_drought" = "Warmed Drought",
                                  "drought_insecticide" = "Drought Insecticide",
                                    "insecticide" = "Insecticide",
                                    'warmed_drought_insecticide' = "Warmed Drought Insecticide",
                                    'warmed_insecticide' = "Warmed Insecticide")) +
        #scale_y_continuous(name = NULL, breaks = NULL) +
        panel_border() +
        theme_bw() +
        ggtitle("") +
        theme(axis.text.x = element_text(angle = 60, hjust = 1)) +
        theme(strip.text = element_text(size = 12, face = "italic"), axis.text.x = element_text(size = 
              axis.title = element_blank(),
              title = element_text(size = 12))
```



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

#Taxonomy - Genus Level

```
# 16S Habitat Overview Plot
# Rarefy data to 1000 reads
set.seed(01221990)

genus_sum_16S <- tapply(taxa_sums(data_16S_rel),
                        tax_table(data_16S_rel)[, "Genus"], sum, na.rm=TRUE)
top10Genus_16S <- names(sort(genus_sum_16S, TRUE))[1:10]

abundant_16S_genus <- subset_taxa(data_16S_rel, Genus %in% top10Genus_16S)
abundant_16S_genus_glom <- tax_glom(abundant_16S_genus, taxrank = "Genus")
abundant_16S_genus_melt <- psmelt(abundant_16S_genus_glom)
```

```
## Warning in psmelt(abundant_16S_genus_glom): The rank names:
## OTU
##  have been renamed to:
## taxa_OTU
## to avoid conflicts with special phyloseq plot attribute names.
```

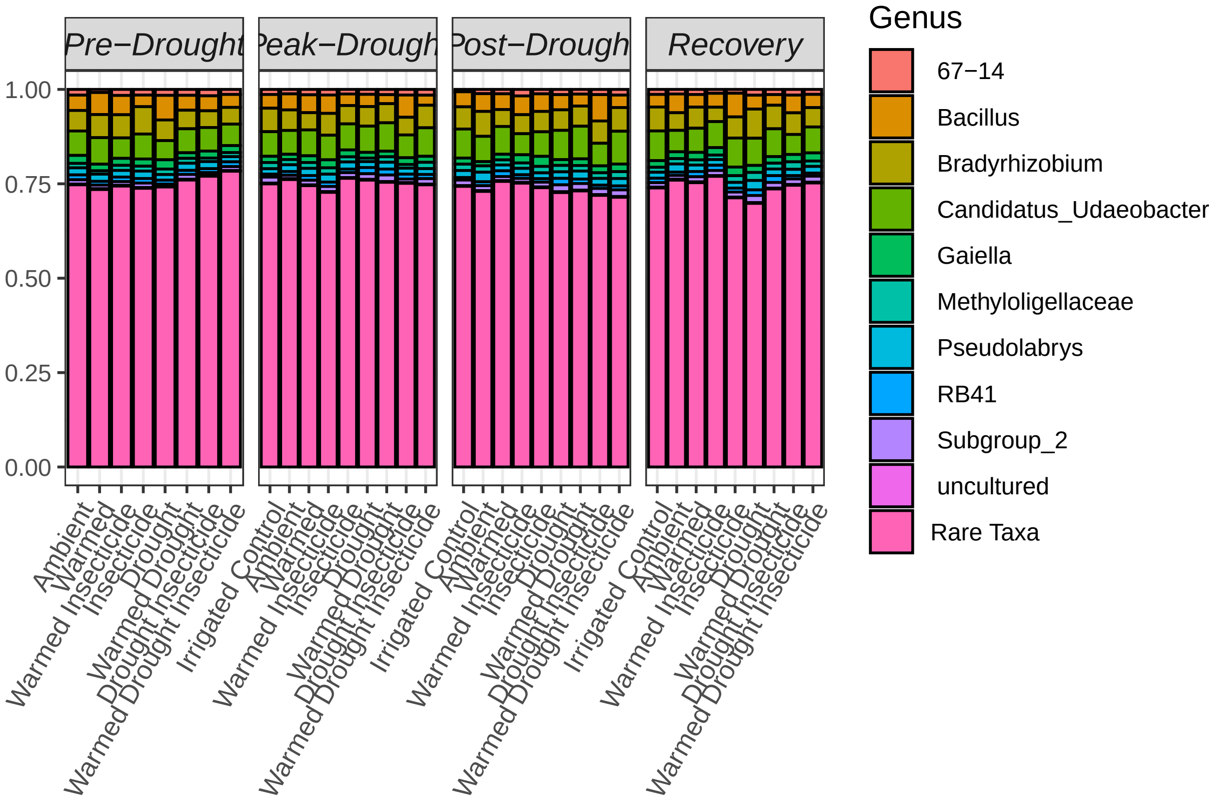
```
sum4 <- abundant_16S_genus_melt %>%
  group_by(Subplot_Descriptions, Drought, Genus) %>%
  summarise(means = mean(Abundance))
```

```
## `summarise()` has grouped output by 'Subplot_Descriptions', 'Drought'. You can
## override using the `.groups` argument.
```

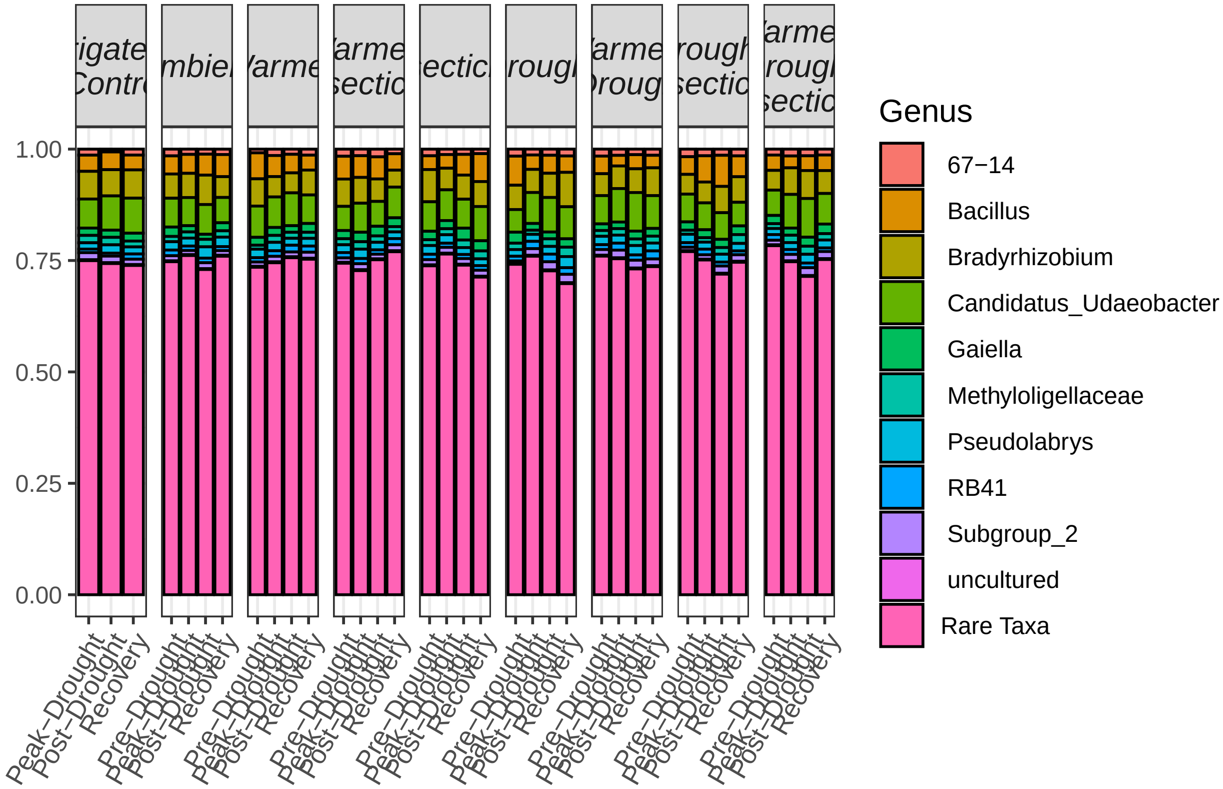
```
#Everything that's left is considered rare
rare <- sum4 %>%
  group_by(Subplot_Descriptions, Drought) %>%
  summarise(means = 1- sum(means)) %>%
  mutate(Genus = "Rare Taxa")
```

```
## `summarise()` has grouped output by 'Subplot_Descriptions'. You can override
## using the `.groups` argument.
```

```
#concatenate the datasets
sum4 = rbind(sum4, rare)

#order groups
sum4$Genus <- forcats::fct_relevel(sum4$Genus, "Rare Taxa", after = Inf)

# Stacked bar plot for top genus'
# facet_grid by timing of soil sampling "Drought"
#png("T7_warmx_16S_genus_stacked_bar_plot.png", units="in", width=10, height=6, res=300)
ggplot(sum4, aes(x = Subplot_Descriptions, y = means, fill = Genus)) +
        geom_bar(position = "stack", stat = "identity", col = "black") +
        facet_grid(~Drought, scale="free") +
        theme(legend.position = "right", legend.title = element_blank(), axis.line = element_blank()) +
        #scale_fill_manual(values = c("firebrick4","tomato", "darkorange3", "orange1" ,"goldenrod4", "k
        #                              "forestgreen", "olivedrab1", "midnightblue", "powderblue", "mediu
        scale_x_discrete(labels=c("ambient" = "Ambient",
                                  "drought" = "Drought",
                                  "irrigated_control" = "Irrigated Control",
                                  "warmed" = "Warmed",
                                  "warmed_drought" = "Warmed Drought",
                                  "drought_insecticide" = "Drought Insecticide",
                                  "insecticide" = "Insecticide",
                                  'warmed_drought_insecticide' = "Warmed Drought Insecticide",
                                  'warmed_insecticide' = "Warmed Insecticide")) +
        #scale_y_continuous(name = NULL, breaks = NULL) +
        panel_border() +
        theme_bw() +
        ggtitle("") +
        theme(axis.text.x = element_text(angle = 60, hjust = 1)) +
        theme(strip.text = element_text(size = 12, face = "italic"), axis.text.x = element_text(size = 
              axis.title = element_blank(),
              title = element_text(size = 12))
```



```
#dev.off()

# facet_grid by treatment aka "Subplot_Descriptions"
#png("T7_warmx_16S_genus_stacked_bar_plot_by_treatment.png", units="in", width=11, height=6, res=300)
ggplot(sum4, aes(x = Drought, y = means, fill = Genus)) +
        geom_bar(position = "stack", stat = "identity", col = "black") +
        facet_grid(~Subplot_Descriptions, scale="free", labeller = as_labeller(treatment_names)) +
        theme(legend.position = "right", legend.title = element_blank(), axis.line = element_blank()) +
        #scale_fill_manual(values = c("firebrick4","tomato", "darkorange3", "orange1" ,"goldenrod4", "k
        #                             "forestgreen", "olivedrab1", "midnightblue", "powderblue", "mediu
        scale_x_discrete(labels=c("ambient" = "Ambient",
                                  "drought" = "Drought",
                                  "irrigated_control" = "Irrigated Control",
                                  "warmed" = "Warmed",
                                  "warmed_drought" = "Warmed Drought",
                                  "drought_insecticide" = "Drought Insecticide",
                                  "insecticide" = "Insecticide",
                                  'warmed_drought_insecticide' = "Warmed Drought Insecticide",
                                  'warmed_insecticide' = "Warmed Insecticide")) +
        #scale_y_continuous(name = NULL, breaks = NULL) +
        panel_border() +
        theme_bw() +
        ggtitle("") +
        theme(axis.text.x = element_text(angle = 60, hjust = 1)) +
        theme(strip.text = element_text(size = 12, face = "italic"), axis.text.x = element_text(size = 
              axis.title = element_blank(),
              title = element_text(size = 12))
```



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

Create phyloseq object that doesn't include insecticide treatments

```
data_16S_filtered_noinsect <- subset_samples(data_16S_filtered, Subplot_Descriptions%in%c("ambient", "i
summarize_phyloseq(data_16S_filtered_noinsect)
```

```
## Compositional = NO2

## 1] Min. number of reads = 30832] Max. number of reads = 1470113] Total number of reads = 39375314] A
##         (i.e. exactly one read detected across all samples)0.003758903903621710] Number of sample va

## [[1]]
## [1] "1] Min. number of reads = 3083"
##
## [[2]]
## [1] "2] Max. number of reads = 147011"
##
## [[3]]
## [1] "3] Total number of reads = 3937531"
##
## [[4]]
## [1] "4] Average number of reads = 34845.407079646"
##
## [[5]]
## [1] "5] Median number of reads = 27520"
##
```

```
## [[6]]
## [1] "7] Sparsity = 0.981825366979626"
##
## [[7]]
## [1] "6] Any OTU sum to 1 or less? YES"
##
## [[8]]
## [1] "8] Number of singletons = 19518"
##
## [[9]]
## [1] "9] Percent of OTUs that are singletons \n        (i.e. exactly one read detected across all samp
##
## [[10]]
## [1] "10] Number of sample variables are: 17"
##
## [[11]]
##  [1] "RSTF_SampleID"            "Barcode"
##  [3] "LinkerPrimerSequence"     "Reverse_Primer"
##  [5] "MiSeqRun"                 "Treatment"
##  [7] "Replicate"                "Rep"
##  [9] "Footprint_Treatment_full" "Footprint"
## [11] "Footprint_Location"       "Subplot"
## [13] "Subplot_Location"         "Subplot_Descriptions"
## [15] "Unique_ID"                "Drought"
## [17] "Datetime_UTC"
```

```
#  Rarefy the samples without replacement. Rarefaction is used to simulate even number of reads per samp
ps.rarefied = rarefy_even_depth(data_16S_filtered_noinsect, rngseed=1, sample.size = 0.9*min(sample_sums
```

```
## ‘set.seed(1)‘ was used to initialize repeatable random subsampling.
```

```
## Please record this for your records so others can reproduce.
```

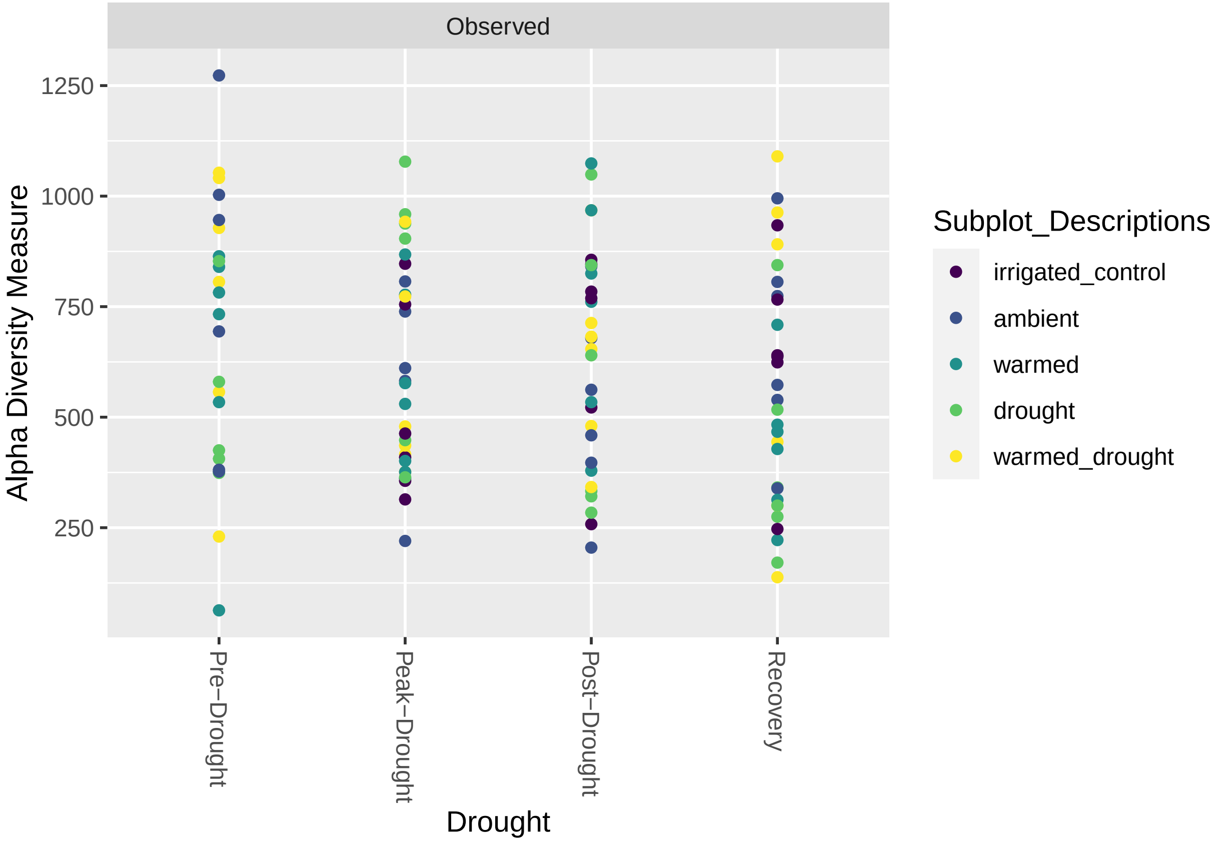
```
## Try ‘set.seed(1); .Random.seed‘ for the full vector
```

```
## ...
```

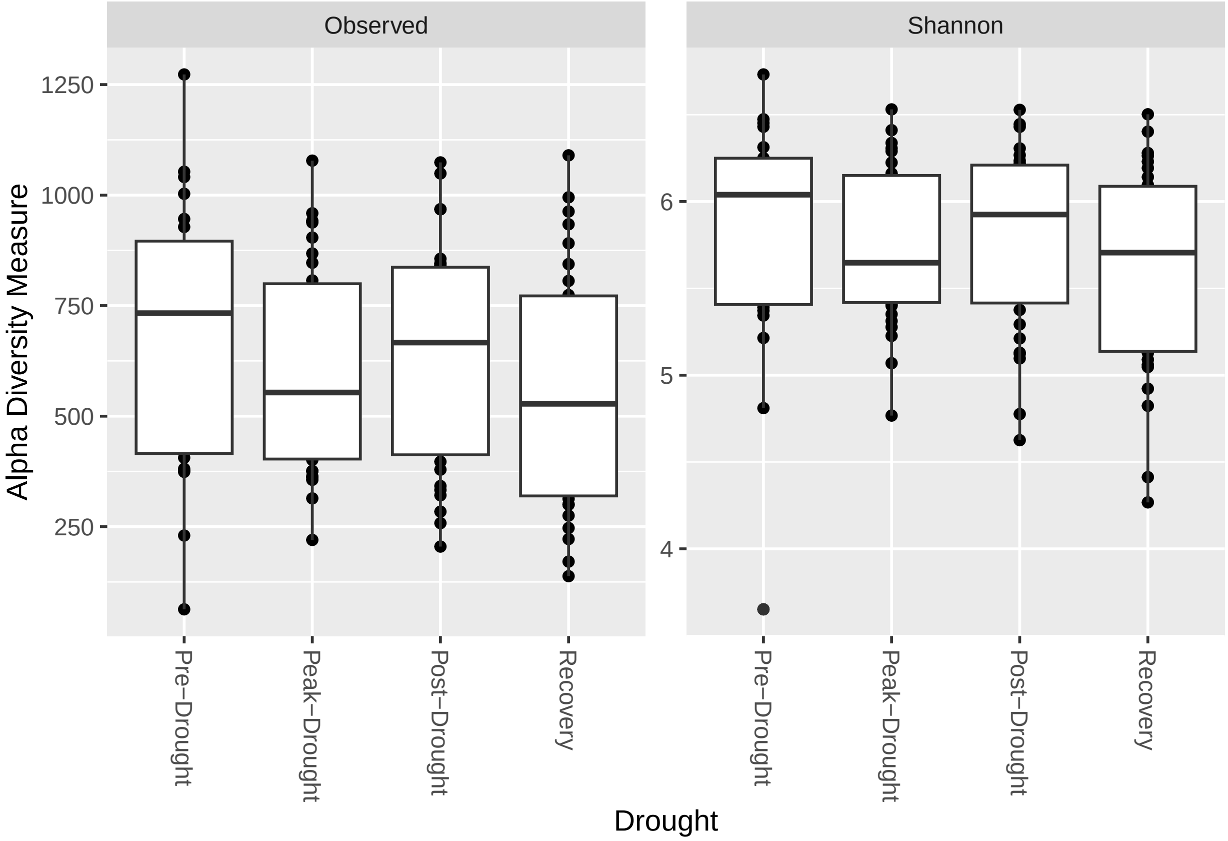
```
## 36760OTUs were removed because they are no longer
## present in any sample after random subsampling
```

```
## ...
```

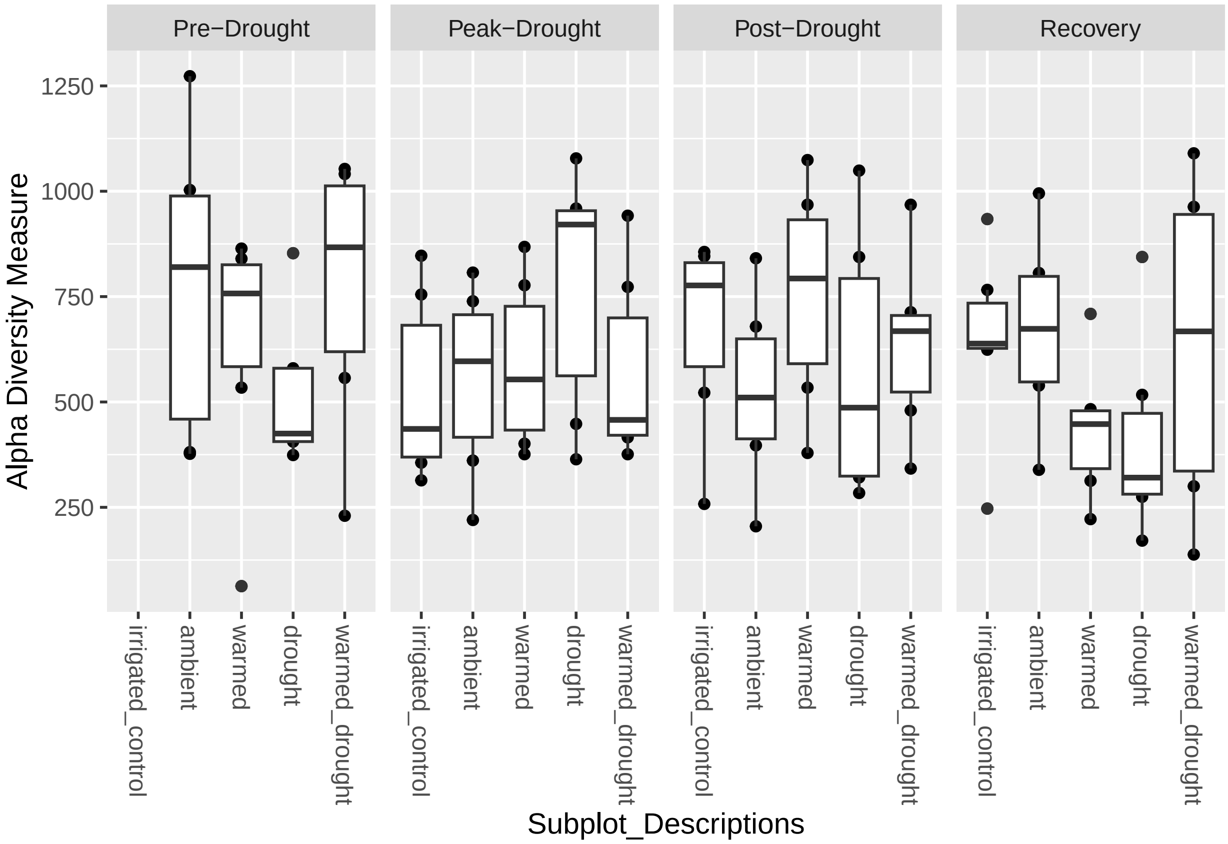
```
# Plotting Alpha Diversity
treatment_names <- c("ambient" = "Ambient",
                     "drought" = "Drought",
                     "irrigated_control" = "Irrigated Control",
                     "warmed" = "Warmed",
                     "warmed_drought" = "Warmed + Drought")

plot_richness(ps.rarefied, x="Drought", color="Subplot_Descriptions", measures=c("Observed"))
```



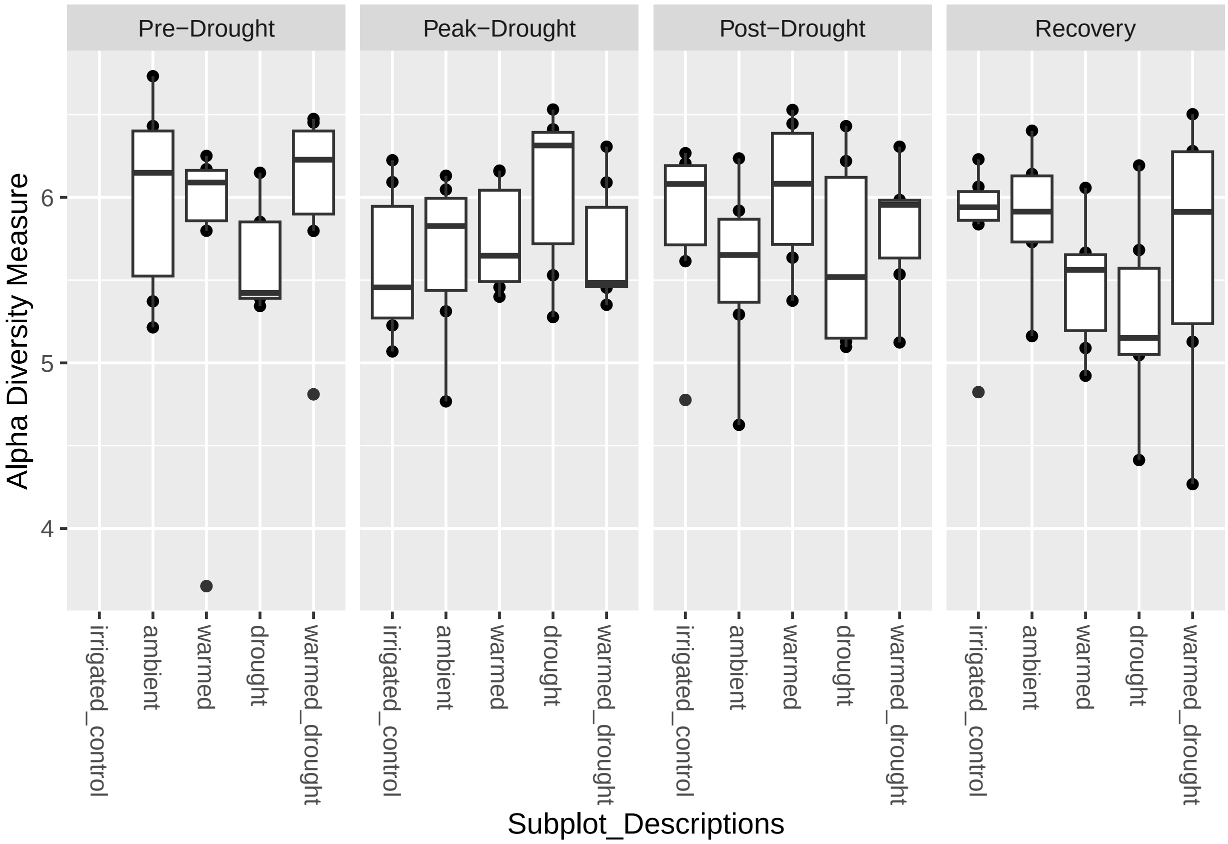
```
plot_richness(ps.rarefied, x="Drought", measures=c("Observed", "Shannon")) + geom_boxplot()
```



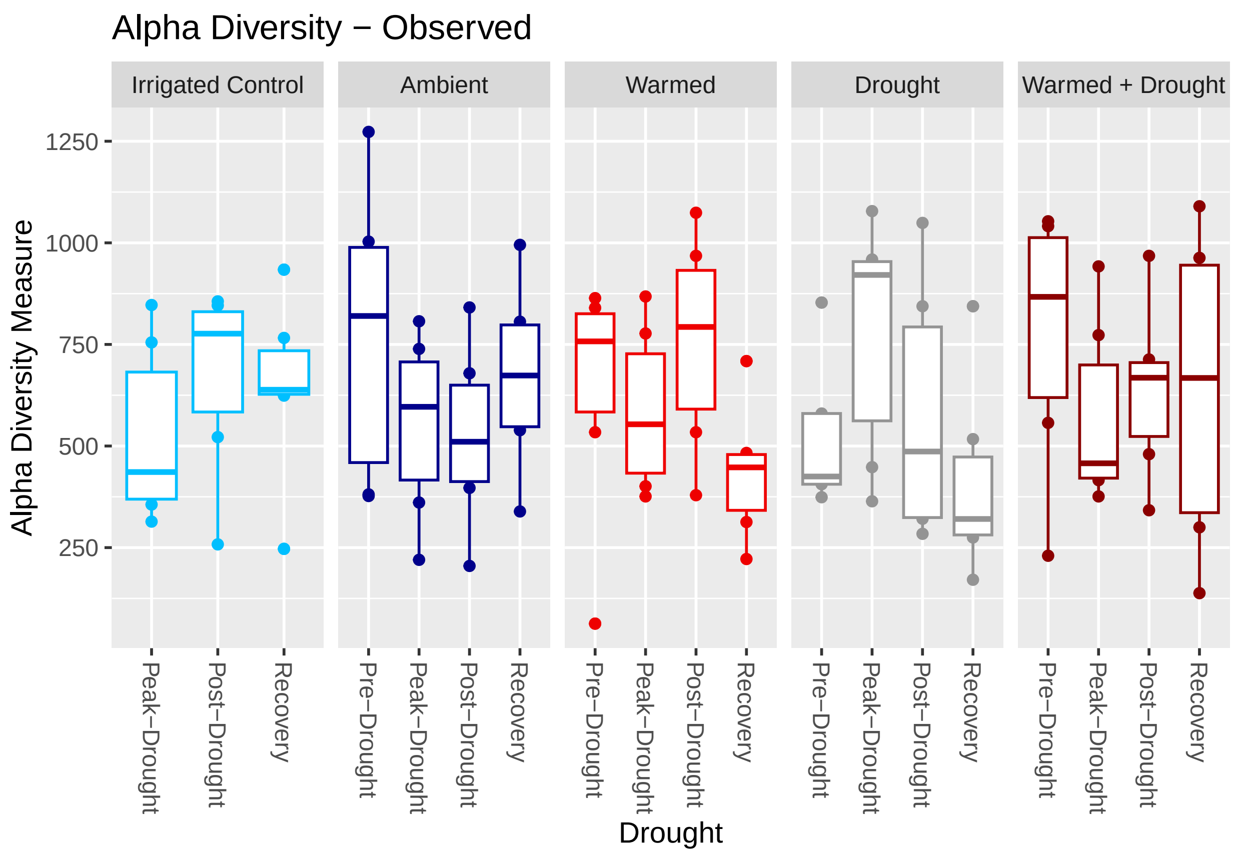
```
plot_richness(ps.rarefied, x="Subplot_Descriptions", measures=c("Observed")) + geom_boxplot() + facet_g
```



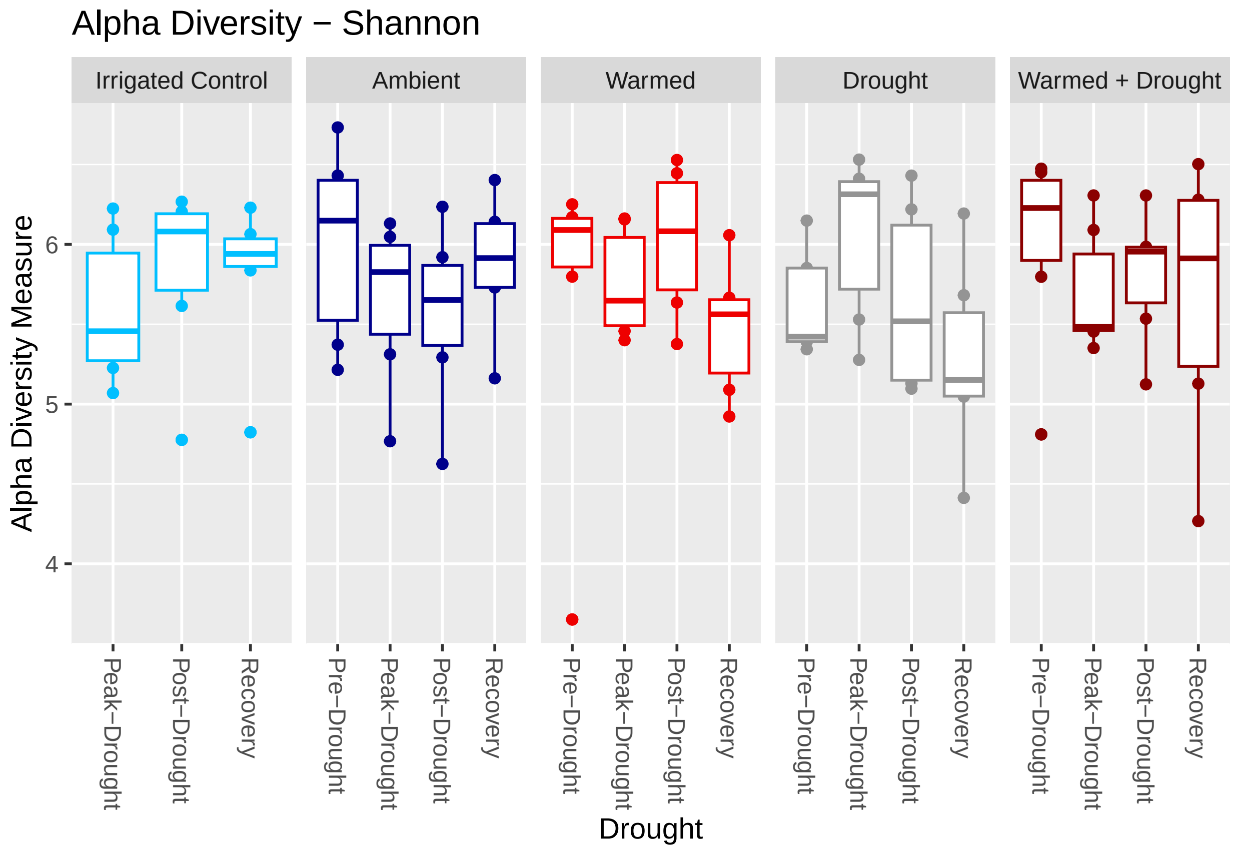
```
plot_richness(ps.rarefied, x="Subplot_Descriptions", measures=c("Shannon")) + geom_boxplot() + facet_gri
```



```
#png("T7_warmx_16S_observed_alpha_diversity_no_insects.png", units="in", width=10, height=6, res=300)
plot_richness(ps.rarefied, x="Drought", measures=c("Observed"), color = "Subplot_Descriptions") +
        geom_boxplot() +
        labs(title = "Alpha Diversity - Observed") +
        facet_grid(.~Subplot_Descriptions, scale="free", labeller = as_labeller(treatment_names)) +
        scale_colour_manual(values = c("ambient" = "darkblue", "drought" = "gray58",
        "irrigated_control" = "deepskyblue", "warmed" = "red2", "warmed_drought" = "darkred")) +
        theme(legend.position = "none")
```



```
#dev.off()

#png("T7_warmx_16S_shannon_alpha_diversity_no_insects.png", units="in", width=10, height=6, res=300)
plot_richness(ps.rarefied, x="Drought", measures=c("Shannon"), color = "Subplot_Descriptions") +
        geom_boxplot() +
        labs(title = "Alpha Diversity - Shannon") +
        facet_grid(.~Subplot_Descriptions, scale="free", labeller = as_labeller(treatment_names)) +
        scale_colour_manual(values = c("ambient" = "darkblue", "drought" = "gray58",
        "irrigated_control" = "deepskyblue", "warmed" = "red2", "warmed_drought" = "darkred")) +
        theme(legend.position = "none")
```



```
#dev.off()

richness <- estimate_richness(ps.rarefied)
head(richness)
```

```
##     Observed    Chao1 se.chao1      ACE    se.ACE  Shannon   Simpson InvSimpson
## X18      374 421.0426 15.07474 410.8929  8.870340 5.422226 0.9933222  149.75044
## X20      557 672.3871 24.93554 658.7261 11.358652 5.797220 0.9947132  189.15186
## X27      377 428.7708 16.11213 418.3584  9.013134 5.371822 0.9921342  127.13250
## X31      406 463.6780 16.59482 454.8852  9.007781 5.390647 0.9916406  119.62622
## X37      381 436.8621 16.27162 429.0176  8.869382 5.214496 0.9892282   92.83479
## X38       63  63.0000  0.00000  63.0000  3.013198 3.651438 0.9614018   25.90795
##        Fisher
## X18 116.44862
## X20 209.81353
## X27 117.78389
## X31 131.01499
## X37 119.57395
## X38  11.47039
```

```
kruskal.test(richness$Shannon ~ sample_data(ps.rarefied)$Subplot_Descriptions)
```

```
##
##  Kruskal-Wallis rank sum test
##
```

```
## data:  richness$Shannon by sample_data(ps.rarefied)$Subplot_Descriptions
## Kruskal-Wallis chi-squared = 1.1143, df = 4, p-value = 0.892
```

```
pairwise.wilcox.test(richness$Shannon, sample_data(ps.rarefied)$Subplot_Descriptions, p.adj = "bonf")
```

```
##
##  Pairwise comparisons using Wilcoxon rank sum exact test
##
## data:  richness$Shannon and sample_data(ps.rarefied)$Subplot_Descriptions
##
##                irrigated_control ambient warmed drought
## ambient        1                 -       -      -
## warmed         1                 1       -      -
## drought        1                 1       1      -
## warmed_drought 1                 1       1      1
##
## P value adjustment method: bonferroni
```

```
kruskal.test(richness$Shannon ~ sample_data(ps.rarefied)$Drought)
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  richness$Shannon by sample_data(ps.rarefied)$Drought
## Kruskal-Wallis chi-squared = 2.8993, df = 3, p-value = 0.4074
```

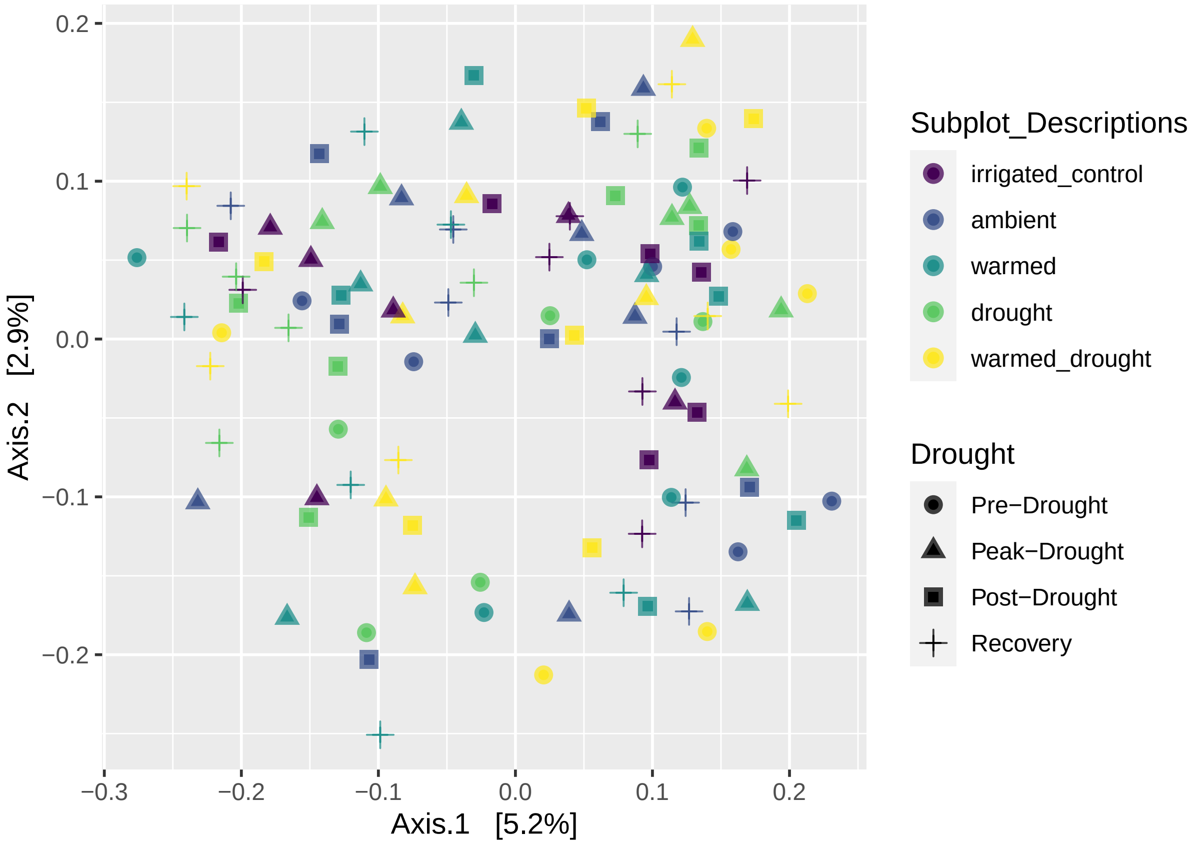
```
pairwise.wilcox.test(richness$Shannon, sample_data(ps.rarefied)$Drought, p.adj = "bonf")
```

```
##
##  Pairwise comparisons using Wilcoxon rank sum exact test
##
## data:  richness$Shannon and sample_data(ps.rarefied)$Drought
##
##              Pre-Drought Peak-Drought Post-Drought
## Peak-Drought 1.00        -            -
## Post-Drought 1.00        1.00         -
## Recovery     0.63        1.00         1.00
##
## P value adjustment method: bonferroni
```

Beta Diversity

```
# PCoA plot using the unweighted UniFrac as distance
wunifrac_dist = phyloseq::distance(ps.rarefied, method="unifrac", weighted=F)
ordination = ordinate(ps.rarefied, method="PCoA", distance=wunifrac_dist)
plot_ordination(ps.rarefied, ordination, color="Subplot_Descriptions", shape = "Drought") +
        geom_point(size=3, alpha=0.75) +
        theme(aspect.ratio=1)
```

```
## Warning: Using shapes for an ordinal variable is not advised
```



```r
# Test whether the treatments ("Subplot_Descriptions") differ significantly from each other using the p
adonis2(wunifrac_dist ~ sample_data(ps.rarefied)$Subplot_Descriptions + sample_data(ps.rarefied)$Drough
```

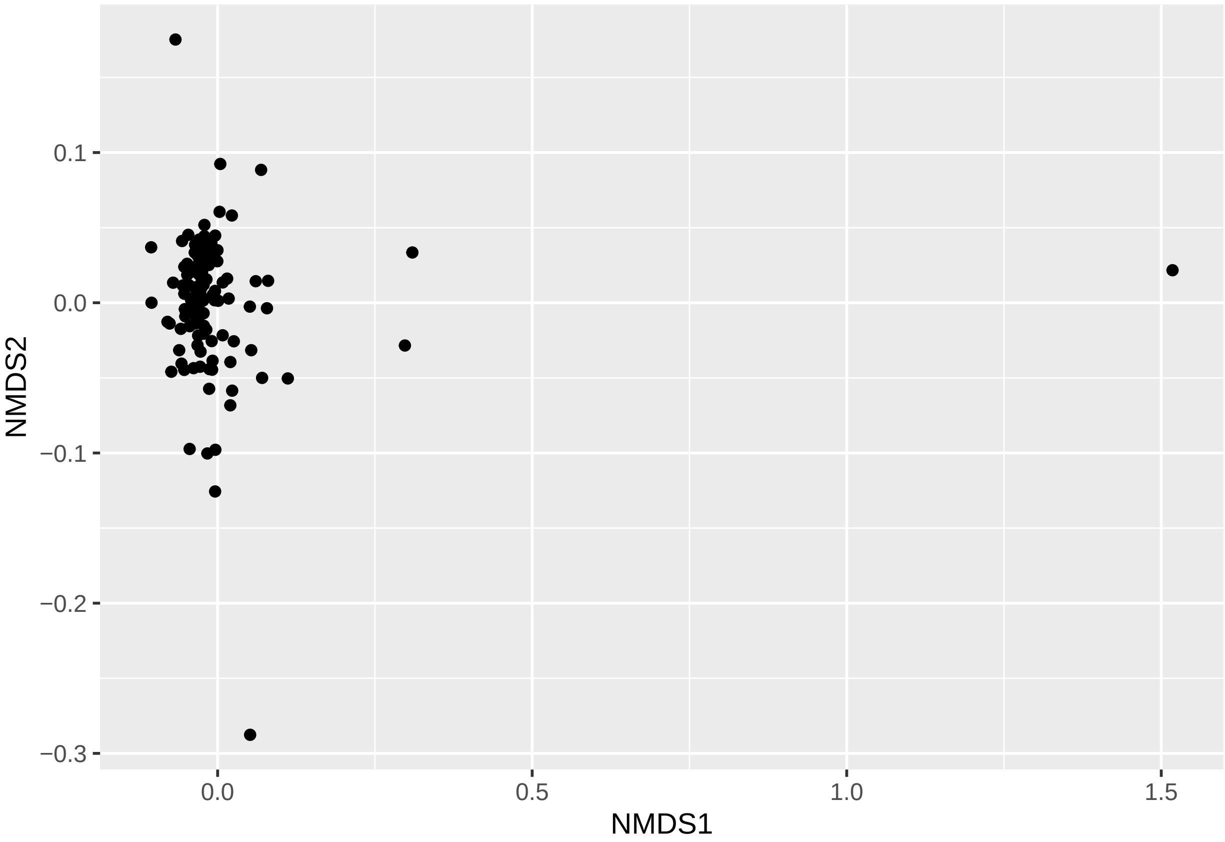
```
## Permutation test for adonis under reduced model
## Terms added sequentially (first to last)
## Permutation: free
## Number of permutations: 999
##
## adonis2(formula = wunifrac_dist ~ sample_data(ps.rarefied)$Subplot_Descriptions + sample_data(ps.rar
##                                               Df SumOfSqs      R2      F
## sample_data(ps.rarefied)$Subplot_Descriptions   4    1.436 0.03618 1.0138
## sample_data(ps.rarefied)$Drought                3    1.076 0.02711 1.0131
## Residual                                      105   37.178 0.93671
## Total                                         112   39.690 1.00000
##                                               Pr(>F)
## sample_data(ps.rarefied)$Subplot_Descriptions  0.316
## sample_data(ps.rarefied)$Drought               0.327
## Residual
## Total
```

Ordination using Phyloseq package

```r
ord_16S <- prune_taxa(names(sort(taxa_sums(ps.rarefied), TRUE)[1:50]), ps.rarefied)

# ordination
jaccard_pcoa_16S <- ordinate(
        physeq = ord_16S,
        method = "NMDS",
        distance = "jaccard"
        )
```

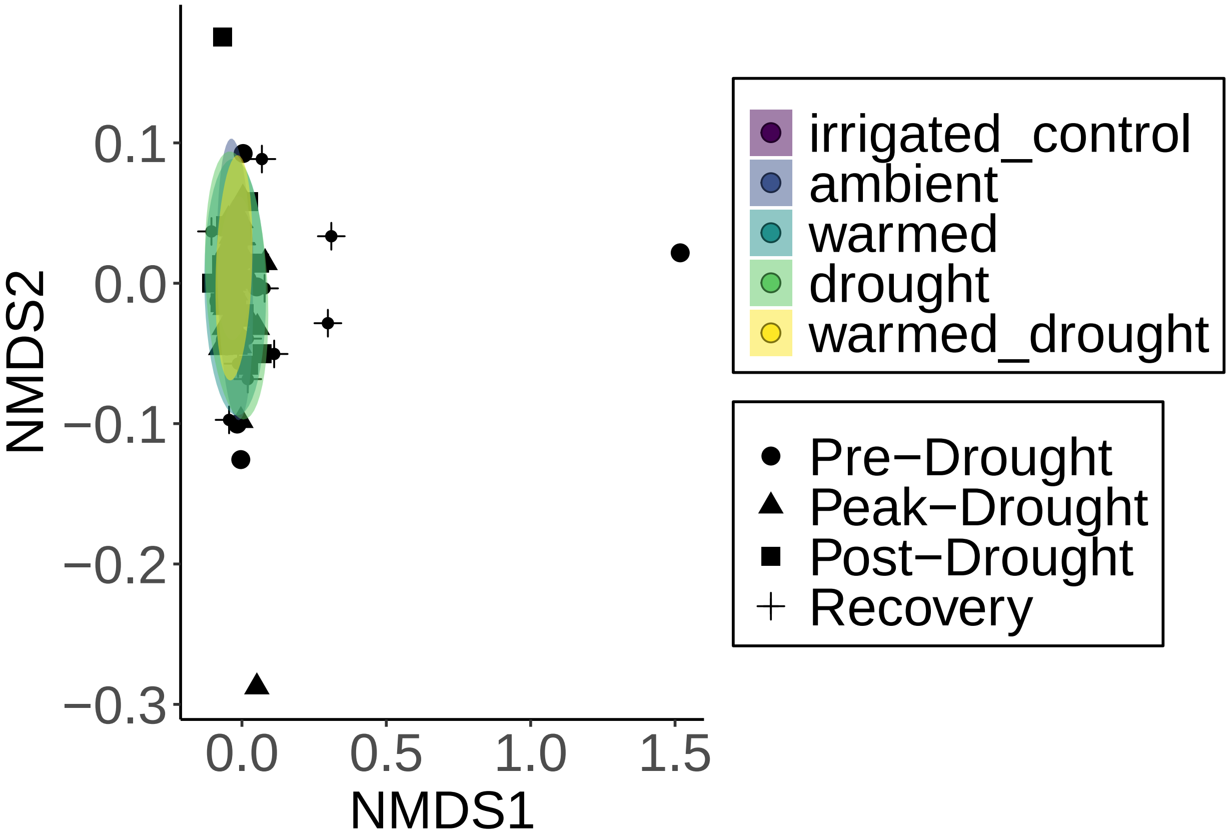
```
## Square root transformation
## Wisconsin double standardization
## Run 0 stress 0.1369458
## Run 1 stress 0.1369458
## ... Procrustes: rmse 0.0002418106  max resid 0.000971151
## ... Similar to previous best
## Run 2 stress 0.1390981
## Run 3 stress 0.1410165
## Run 4 stress 0.1377956
## Run 5 stress 0.1424688
## Run 6 stress 0.1375706
## Run 7 stress 0.1372971
## ... Procrustes: rmse 0.008945154  max resid 0.07385745
## Run 8 stress 0.1393521
## Run 9 stress 0.1394098
## Run 10 stress 0.142491
## Run 11 stress 0.1369479
## ... Procrustes: rmse 0.0007447091  max resid 0.004461067
## ... Similar to previous best
## Run 12 stress 0.1381146
## Run 13 stress 0.1420626
## Run 14 stress 0.1410847
## Run 15 stress 0.1374233
## ... Procrustes: rmse 0.006427872  max resid 0.03987364
## Run 16 stress 0.1395362
## Run 17 stress 0.1432958
## Run 18 stress 0.1370693
## ... Procrustes: rmse 0.003604675  max resid 0.02919269
## Run 19 stress 0.1369476
## ... Procrustes: rmse 0.0008088631  max resid 0.003719665
## ... Similar to previous best
## Run 20 stress 0.1379677
## *** Best solution repeated 3 times
```

```r
# plot ordination
plot_ordination(ord_16S, jaccard_pcoa_16S, "samples")
```



```
#png("T7_warmx_16S_NMDS_ordination.png", units="in", width=8, height=6, res=300)
plot_ordination(
  physeq = ord_16S,                                                         #phyloseq object
  ordination = jaccard_pcoa_16S) +                                                #ordination
  geom_point(aes(fill = Subplot_Descriptions, shape = Drought), size = 3) + #sets fill color to samplet
        stat_ellipse(aes(fill = factor(Subplot_Descriptions)), geom = "polygon", alpha = .5) +
  #scale_shape_manual(values = c(21, 22, 25)) +
  #scale_fill_manual(values = sample_colors) +
  theme_classic() +                                                      #changes theme, removes grey b
  theme(
    legend.text = element_text(size = 20),                               #changes legend size
    legend.title = element_blank(),                                      #removes legend title
    legend.background = element_rect(fill = "white", color = "black"))+
  theme(axis.text.y.left = element_text(size = 20),
        axis.text.x = element_text(size = 20),
        axis.title.x = element_text(size = 20),
        axis.title.y = element_text(size = 20))+
  guides(fill = guide_legend(override.aes = list(shape = 21)))
```

```
## Warning: Using shapes for an ordinal variable is not advised
```



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

OTU differential abundance testing with DESeq2