

# First experimental evidence for active farming in ambrosia beetles and strong heredity of garden microbiomes - Fungal Analysis

Janina M.C. Diehl

14.07.2022

## Data Preparation

load required packages

```
library(lme4)
library(permute)
library(lattice)
library(vegan)
library(phyloseq)
library(ggplot2)
library(dplyr)
library(scales)
library(grid)
library(DHARMa)
library(ggeffects)
library(glmmTMB)
library(lmerTest)
library(emmeans)
library(sjPlot)
library(fitdistrplus)
library(GLMMadaptive)
library(microbiome)
library(microbiomeutilities)
library(knitr)
library(ggpubr)
library(doBy)
library(performance)
library(see)
library(patchwork)
library(pairwiseAdonis)
library(cowplot)
library(multcomp)
library(car)
library(forcats)
library(ggrepel)
library(tidyverse)
```

loading the data files for fungi

```
dataLSU <- otu_table(read.table("28S_Rem_zotu_table.txt", sep="\t", header=T, row.names=1, check.names =  
taxLSU <- tax_table(as.matrix(read.table("28S_Rem_zotus97.tax.txt", sep="\t", header=T, row.names=1, fi  
datasampleLSU <- sample_data(read.table("28S_map_removal.txt", sep="\t", header=T, row.names=1)))
```

merge data into phyloseq object

```
(allLSU <- merge_phyloseq(dataLSU, taxLSU, datasampleLSU))
```

```
## phyloseq-class experiment-level object  
## otu_table() OTU Table: [ 212 taxa and 57 samples ]  
## sample_data() Sample Data: [ 57 samples by 16 sample variables ]  
## tax_table() Taxonomy Table: [ 212 taxa by 8 taxonomic ranks ]
```

copy taxonomic classification in tax\_table columns with gaps and add "\_\_spc"

```
dataset.LSU = subset_taxa(allLSU, (Kingdom == "k:Fungi"))  
tax_table(dataset.LSU)[tax_table(dataset.LSU)[,"Phylum"]=="", "Phylum"] <- paste(tax_table(dataset.LSU)[tax_  
tax_table(dataset.LSU)[tax_table(dataset.LSU)[,"Class"]=="", "Class"] <- paste(tax_table(dataset.LSU)[tax_  
tax_table(dataset.LSU)[tax_table(dataset.LSU)[,"Order"]=="", "Order"] <- paste(tax_table(dataset.LSU)[tax_  
tax_table(dataset.LSU)[tax_table(dataset.LSU)[,"Family"]=="", "Family"] <- paste(tax_table(dataset.LSU)[tax_  
tax_table(dataset.LSU)[tax_table(dataset.LSU)[,"Genus"]=="", "Genus"] <- paste(tax_table(dataset.LSU)[tax_  
tax_table(dataset.LSU)[tax_table(dataset.LSU)[,"Species"]=="", "Species"] <- paste(tax_table(dataset.LSU)[
```

start filtering out all ZOTUs that were only assigned to Kingdom (Fungi)

excluding all taxa assigned to Fungi\_spc

```
dataset.LSU.ordi = subset_taxa(dataset.LSU, (Kingdom == "k:Fungi"))  
dataset.LSU.ordi = subset_taxa(dataset.LSU.ordi, Phylum != "k:Fungi_spc")
```

check all the columns for patterns ranging from [a-z] joined by \_\_\_\_ like this [a-z]\_\_\_\_ and substitute it with "" i.e. nothing.

```
tax_table(dataset.LSU.ordi)[, colnames(tax_table(dataset.LSU.ordi))] <- gsub(tax_table(dataset.LSU.ordi  
tax_table(dataset.LSU.ordi)[, colnames(tax_table(dataset.LSU.ordi))] <- gsub(tax_table(dataset.LSU.ordi  
tax_table(dataset.LSU.ordi)[, colnames(tax_table(dataset.LSU.ordi))] <- gsub(tax_table(dataset.LSU.ordi  
tax_table(dataset.LSU.ordi)[, colnames(tax_table(dataset.LSU.ordi))] <- gsub(tax_table(dataset.LSU.ordi  
tax_table(dataset.LSU.ordi) <- tax_table(dataset.LSU.ordi)[,2:8]
```

Format the phyloseq object to add the best taxonomy in phyloseq object (tax\_table and otu\_table).

```
dataset.LSU.ordi <- format_to_besthit(dataset.LSU.ordi)
taxa_names(dataset.LSU.ordi)[1:5]
```

```
## [1] "LSUzotu2:Raffaelea_sulphurea" "LSUzotu7:Raffaelea_sulphurea"
## [3] "LSUzotu1:Chaetomium_globosum" "LSUzotu12:Raffaelea_sulphurea"
## [5] "LSUzotu4:Raffaelea_canadensis"
```

check the table for total reads per sample

```
colSums(otu_table(dataset.LSU.ordi))
```

```
##      B0-19a      B0-19b      B0-22a      B0-22b      B0-24      B0-27      B10-01      B10-06
##      8655      19999      19671      37621      27652      16442      24646      26723
##      B10-09a      B10-09b      B10-18      B10-24      B10-25a      B10-25b      B10-26a      B15-01a
##      26663      39932      6770      10235      24348      14123      26861      12155
##      B15-01b      B15-06      B15-07a      B15-07b      B15-11      B15-20      B15-24      B15-25a
##      24509      2378      19916      9950      19307      14350      21346      8108
##      B15-25b      B16-01a      B16-04      B16-31a      B16-31b      B20-04a      B20-08a      B20-08b
##      27304      20358      7521      15823      23816      34274      27572      19430
##      B20-11      B23-21      B23-23a      B24-02a      B24-26a      B24-26b      B24-28a      B24-29
##      8395      27931      26875      1087      16017      10464      19882      7041
##      B36-14a      B36-14b      B36-30      B36-31a      B36-32      B39-15      B39-16      B39-17a
##      8615      25211      8240      17279      1092      12342      11854      12233
##      B46-03      B7-07      B7-08a      medium1      medium2      negative2      negative1      mock
##      21081      5815      4694      11      98      41      259      23868
##      standard
##      26
```

no losses

check controls from dataset: only ten most abundant ZOTUs are picked for visualisation prae decontam

```
neg.controls<-subset_samples(dataset.LSU.ordi, Treatment=="medium" | Treatment=="negative")
sample_names(neg.controls)
```

```
## [1] "medium1" "medium2" "negative2" "negative1"
```

visualisation of negative controls

```
filtaxa <- names (sort(rowSums(otu_table(neg.controls)),decreasing=T))[1:10]
tax_table(dataset.LSU.ordi)[filtaxa]
```

```
## Taxonomy Table:      [10 taxa by 8 taxonomic ranks]:
##                                Domain  Phylum      Class
## LSUzotu1:Chaetomium_globosum      "Fungi" "Ascomycota" "Sordariomycetes"
```

## LSUzotu2:Raffaelea_sulphurea	"Fungi"	"Ascomycota"	"Sordariomycetes"
## LSUzotu87:Cladosporium_sp._CHTAE11	"Fungi"	"Ascomycota"	"Dothideomycetes"
## LSUzotu26:Chaetomium_spc	"Fungi"	"Ascomycota"	"Sordariomycetes"
## LSUzotu25:Raffaelea_sulphurea	"Fungi"	"Ascomycota"	"Sordariomycetes"
## LSUzotu159:Raffaelea_sulphurea	"Fungi"	"Ascomycota"	"Sordariomycetes"
## LSUzotu4:Raffaelea_canadensis	"Fungi"	"Ascomycota"	"Sordariomycetes"
## LSUzotu72:Raffaelea_sulphurea	"Fungi"	"Ascomycota"	"Sordariomycetes"
## LSUzotu6:Sporothrix_stenoceras	"Fungi"	"Ascomycota"	"Sordariomycetes"
## LSUzotu130:Chaetomiaceae_spc	"Fungi"	"Ascomycota"	"Sordariomycetes"
##	Order	Family	
## LSUzotu1:Chaetomium_globosum	"Sordariales"	"Chaetomiaceae"	
## LSUzotu2:Raffaelea_sulphurea	"Ophiostomatales"	"Ophiostomataceae"	
## LSUzotu87:Cladosporium_sp._CHTAE11	"Capnodiales"	"Cladosporiaceae"	
## LSUzotu26:Chaetomium_spc	"Sordariales"	"Chaetomiaceae"	
## LSUzotu25:Raffaelea_sulphurea	"Ophiostomatales"	"Ophiostomataceae"	
## LSUzotu159:Raffaelea_sulphurea	"Ophiostomatales"	"Ophiostomataceae"	
## LSUzotu4:Raffaelea_canadensis	"Ophiostomatales"	"Ophiostomataceae"	
## LSUzotu72:Raffaelea_sulphurea	"Ophiostomatales"	"Ophiostomataceae"	
## LSUzotu6:Sporothrix_stenoceras	"Ophiostomatales"	"Ophiostomataceae"	
## LSUzotu130:Chaetomiaceae_spc	"Sordariales"	"Chaetomiaceae"	
##	Genus		
## LSUzotu1:Chaetomium_globosum	"Chaetomium"		
## LSUzotu2:Raffaelea_sulphurea	"Raffaelea"		
## LSUzotu87:Cladosporium_sp._CHTAE11	"Cladosporium"		
## LSUzotu26:Chaetomium_spc	"Chaetomium"		
## LSUzotu25:Raffaelea_sulphurea	"Raffaelea"		
## LSUzotu159:Raffaelea_sulphurea	"Raffaelea"		
## LSUzotu4:Raffaelea_canadensis	"Raffaelea"		
## LSUzotu72:Raffaelea_sulphurea	"Raffaelea"		
## LSUzotu6:Sporothrix_stenoceras	"Sporothrix"		
## LSUzotu130:Chaetomiaceae_spc	"Chaetomiaceae_spc"		
##	Species		
## LSUzotu1:Chaetomium_globosum	"Chaetomium_globosum"		
## LSUzotu2:Raffaelea_sulphurea	"Raffaelea_sulphurea"		
## LSUzotu87:Cladosporium_sp._CHTAE11	"Cladosporium_sp._CHTAE11"		
## LSUzotu26:Chaetomium_spc	"Chaetomium_spc"		
## LSUzotu25:Raffaelea_sulphurea	"Raffaelea_sulphurea"		
## LSUzotu159:Raffaelea_sulphurea	"Raffaelea_sulphurea"		
## LSUzotu4:Raffaelea_canadensis	"Raffaelea_canadensis"		
## LSUzotu72:Raffaelea_sulphurea	"Raffaelea_sulphurea"		
## LSUzotu6:Sporothrix_stenoceras	"Sporothrix_stenoceras"		
## LSUzotu130:Chaetomiaceae_spc	"Chaetomiaceae_spc"		
##	best_hit		
## LSUzotu1:Chaetomium_globosum	"LSUzotu1:Chaetomium_globosum"		
## LSUzotu2:Raffaelea_sulphurea	"LSUzotu2:Raffaelea_sulphurea"		
## LSUzotu87:Cladosporium_sp._CHTAE11	"LSUzotu87:Cladosporium_sp._CHTAE11"		
## LSUzotu26:Chaetomium_spc	"LSUzotu26:Chaetomium_spc"		
## LSUzotu25:Raffaelea_sulphurea	"LSUzotu25:Raffaelea_sulphurea"		
## LSUzotu159:Raffaelea_sulphurea	"LSUzotu159:Raffaelea_sulphurea"		
## LSUzotu4:Raffaelea_canadensis	"LSUzotu4:Raffaelea_canadensis"		
## LSUzotu72:Raffaelea_sulphurea	"LSUzotu72:Raffaelea_sulphurea"		
## LSUzotu6:Sporothrix_stenoceras	"LSUzotu6:Sporothrix_stenoceras"		
## LSUzotu130:Chaetomiaceae_spc	"LSUzotu130:Chaetomiaceae_spc"		

```
round(otu_table(dataset.LSU.ordi)[filtaxa], digits = 4)
```

```
## OTU Table:          [10 taxa and 57 samples]
##                      taxa are rows
##
##                      B0-19a B0-19b B0-22a B0-22b B0-24 B0-27
## LSUzotu1:Chaetomium_globosum      258 19888 19631 34239 7239 1340
## LSUzotu2:Raffaelea_sulphurea     7705      9    22      6 15410 10421
## LSUzotu87:Cladosporium_sp._CHTAE11      0      0      0      0      0      0
## LSUzotu26:Chaetomium_spc            0      0      0      0     86     22
## LSUzotu25:Raffaelea_sulphurea        0      0      0      0     97     15
## LSUzotu159:Raffaelea_sulphurea        0      0      0      0      0      0
## LSUzotu4:Raffaelea_canadensis       251      1     11   2284  2764  3024
## LSUzotu72:Raffaelea_sulphurea         1      0      0      0     27     13
## LSUzotu6:Sporothrix_stenoceras        0      0      2      0      0      0
## LSUzotu130:Chaetomiaceae_spc          0      0      0    179     22      3
##
##                      B10-01 B10-06 B10-09a B10-09b B10-18 B10-24
## LSUzotu1:Chaetomium_globosum       170 11309 18613 35695 2094 1159
## LSUzotu2:Raffaelea_sulphurea     16220 13569 5195      7 3573 7565
## LSUzotu87:Cladosporium_sp._CHTAE11      0      0      0      0      0      0
## LSUzotu26:Chaetomium_spc            2    165    184      0     48     25
## LSUzotu25:Raffaelea_sulphurea         4    179    213      0     43     19
## LSUzotu159:Raffaelea_sulphurea       148      0      0      0     10     12
## LSUzotu4:Raffaelea_canadensis         13      0      1      0      2      0
## LSUzotu72:Raffaelea_sulphurea         3     52     54      0      8      7
## LSUzotu6:Sporothrix_stenoceras         0      0      1      0      0      0
## LSUzotu130:Chaetomiaceae_spc          0      0      0      0      0      0
##
##                      B10-25a B10-25b B10-26a B15-01a B15-01b
## LSUzotu1:Chaetomium_globosum       6733      95  4421      33 20163
## LSUzotu2:Raffaelea_sulphurea     13414   7353 18965   9796 2454
## LSUzotu87:Cladosporium_sp._CHTAE11      0      0      0      0      0
## LSUzotu26:Chaetomium_spc           287      0    255      0    125
## LSUzotu25:Raffaelea_sulphurea       305      3    413      0    169
## LSUzotu159:Raffaelea_sulphurea        40    111     20      0      0
## LSUzotu4:Raffaelea_canadensis          2      0      1  1260    257
## LSUzotu72:Raffaelea_sulphurea         72      1     98      0     51
## LSUzotu6:Sporothrix_stenoceras         3      0      0     28      0
## LSUzotu130:Chaetomiaceae_spc          0      0      0      0     20
##
##                      B15-06 B15-07a B15-07b B15-11 B15-20 B15-24
## LSUzotu1:Chaetomium_globosum         4  15782   3284  4560  7674  3569
## LSUzotu2:Raffaelea_sulphurea       1589   2382  4285  8639  3977  8446
## LSUzotu87:Cladosporium_sp._CHTAE11      0      0      0      0      0      0
## LSUzotu26:Chaetomium_spc              0    137     34     58     65     42
## LSUzotu25:Raffaelea_sulphurea         0     97     27     60     73     39
## LSUzotu159:Raffaelea_sulphurea         0      0      0      0      0      0
## LSUzotu4:Raffaelea_canadensis        524    178   1572  4372  1616  7115
## LSUzotu72:Raffaelea_sulphurea         0     40     17     14     11      8
## LSUzotu6:Sporothrix_stenoceras         0      0      0      0      0      0
## LSUzotu130:Chaetomiaceae_spc          0      9      8     27     20     21
##
##                      B15-25a B15-25b B16-01a B16-04 B16-31a
## LSUzotu1:Chaetomium_globosum        466   3571  20350     33 10485
## LSUzotu2:Raffaelea_sulphurea       5434  14262      3  7443    793
## LSUzotu87:Cladosporium_sp._CHTAE11      0      0      0      0      0
## LSUzotu26:Chaetomium_spc             6    189      0      0      3
```

## LSUzotu25:Raffaelea_sulphurea	5	280	0	0	0	
## LSUzotu159:Raffaelea_sulphurea	0	0	0	0	0	
## LSUzotu4:Raffaelea_canadensis	1663	3320	1	1	29	
## LSUzotu72:Raffaelea_sulphurea	1	59	0	0	1	
## LSUzotu6:Sporothrix_stenoceras	1	0	3	0	23	
## LSUzotu130:Chaetomiaceae_spc	0	35	0	0	3	
##	B16-31b	B20-04a	B20-08a	B20-08b	B20-11	
## LSUzotu1:Chaetomium_globosum	10239	34162	21	3445	6388	
## LSUzotu2:Raffaelea_sulphurea	33	30	49	7	3	
## LSUzotu87:Cladosporium_sp._CHTAE11	0	0	0	0	0	
## LSUzotu26:Chaetomium_spc	0	0	0	0	0	
## LSUzotu25:Raffaelea_sulphurea	0	0	5	0	0	
## LSUzotu159:Raffaelea_sulphurea	0	0	0	0	0	
## LSUzotu4:Raffaelea_canadensis	4	1	3	0	2	
## LSUzotu72:Raffaelea_sulphurea	0	0	1	0	0	
## LSUzotu6:Sporothrix_stenoceras	0	0	0	0	0	
## LSUzotu130:Chaetomiaceae_spc	0	0	0	3	0	
##	B23-21	B23-23a	B24-02a	B24-26a	B24-26b	
## LSUzotu1:Chaetomium_globosum	22924	8567	1031	11771	5163	
## LSUzotu2:Raffaelea_sulphurea	4734	14764	43	2542	3448	
## LSUzotu87:Cladosporium_sp._CHTAE11	0	0	0	0	0	
## LSUzotu26:Chaetomium_spc	11	291	1	117	117	
## LSUzotu25:Raffaelea_sulphurea	23	491	0	160	160	
## LSUzotu159:Raffaelea_sulphurea	0	13	0	1	10	
## LSUzotu4:Raffaelea_canadensis	2	1	1	0	0	
## LSUzotu72:Raffaelea_sulphurea	8	147	0	56	25	
## LSUzotu6:Sporothrix_stenoceras	0	0	0	0	0	
## LSUzotu130:Chaetomiaceae_spc	0	0	0	0	0	
##	B24-28a	B24-29	B36-14a	B36-14b	B36-30	
## LSUzotu1:Chaetomium_globosum	6763	4311	3226	6297	6760	
## LSUzotu2:Raffaelea_sulphurea	8373	2555	4297	15265	1367	
## LSUzotu87:Cladosporium_sp._CHTAE11	0	0	0	0	0	
## LSUzotu26:Chaetomium_spc	144	17	35	198	4	
## LSUzotu25:Raffaelea_sulphurea	201	11	48	348	5	
## LSUzotu159:Raffaelea_sulphurea	36	0	4	35	0	
## LSUzotu4:Raffaelea_canadensis	29	1	3	0	3	
## LSUzotu72:Raffaelea_sulphurea	61	0	10	64	2	
## LSUzotu6:Sporothrix_stenoceras	0	0	0	0	0	
## LSUzotu130:Chaetomiaceae_spc	0	0	0	0	0	
##	B36-31a	B36-32	B39-15	B39-16	B39-17a	B46-03
## LSUzotu1:Chaetomium_globosum	16058	367	5922	2153	11890	20657
## LSUzotu2:Raffaelea_sulphurea	499	699	6225	7681	130	15
## LSUzotu87:Cladosporium_sp._CHTAE11	0	0	0	0	0	0
## LSUzotu26:Chaetomium_spc	40	0	9	0	9	0
## LSUzotu25:Raffaelea_sulphurea	37	0	3	0	4	0
## LSUzotu159:Raffaelea_sulphurea	0	0	0	2	0	0
## LSUzotu4:Raffaelea_canadensis	3	0	2	0	5	1
## LSUzotu72:Raffaelea_sulphurea	11	0	0	1	3	0
## LSUzotu6:Sporothrix_stenoceras	0	0	0	0	0	0
## LSUzotu130:Chaetomiaceae_spc	0	0	0	0	0	0
##	B7-07	B7-08a	medium1	medium2	negative2	
## LSUzotu1:Chaetomium_globosum	2230	4660	2	12	11	
## LSUzotu2:Raffaelea_sulphurea	3522	5	7	23	21	
## LSUzotu87:Cladosporium sp. CHTAE11	0	0	0	56	0	

## LSUzotu26:Chaetomium_spc	5	0	0	1	2
## LSUzotu25:Raffaelea_sulphurea	5	0	0	2	0
## LSUzotu159:Raffaelea_sulphurea	0	0	0	3	0
## LSUzotu4:Raffaelea_canadensis	0	14	0	0	2
## LSUzotu72:Raffaelea_sulphurea	0	0	0	0	0
## LSUzotu6:Sporothrix_stenoceras	0	0	0	0	1
## LSUzotu130:Chaetomiaceae_spc	0	2	0	0	2
##	negative1	mock	standard		
## LSUzotu1:Chaetomium_globosum	219	4546	6		
## LSUzotu2:Raffaelea_sulphurea	24	747	16		
## LSUzotu87:Cladosporium_sp._CHTAE11	6	0	0		
## LSUzotu26:Chaetomium_spc	0	4	0		
## LSUzotu25:Raffaelea_sulphurea	1	2	0		
## LSUzotu159:Raffaelea_sulphurea	0	0	0		
## LSUzotu4:Raffaelea_canadensis	0	1822	0		
## LSUzotu72:Raffaelea_sulphurea	2	3	0		
## LSUzotu6:Sporothrix_stenoceras	1	15551	0		
## LSUzotu130:Chaetomiaceae_spc	0	22	0		

negative2 seems to have cross contaminations from samples → consider cross-contamination within samples of this PCR plate? → switched primers, mock and neg2 and corrected this now in tax\_table Weirdly Sporothrix is not abundant in samples (highest contaminant here), C.globosum (Zotu 1) highly abundant in all samples, R.canadensis (Zotu4) only half of the samples higher abundant → Line B15 + B0, R. sulphurea (Zotu 2) most samples higher abundant with some lower ones, rest show now abundances in samples

**check for ten most abundant ZOTUs in pos.controls**

```
pos.controls<-subset_samples(dataset.LSU.ordi, Treatment=="standard" | Treatment=="mock")
sample_names(pos.controls)
```

```
## [1] "mock"      "standard"
```

```
filtaxa <- names (sort(rowSums(otu_table(pos.controls)) ,decreasing=T))[1:10]
tax_table(dataset.LSU.ordi)[filtaxa]
```

```
## Taxonomy Table:      [10 taxa by 8 taxonomic ranks]:
##
##      Domain  Phylum  Class
## LSUzotu6:Sporothrix_stenoceras "Fungi" "Ascomycota" "Sordariomycetes"
## LSUzotu1:Chaetomium_globosum   "Fungi" "Ascomycota" "Sordariomycetes"
## LSUzotu4:Raffaelea_canadensis  "Fungi" "Ascomycota" "Sordariomycetes"
## LSUzotu2:Raffaelea_sulphurea   "Fungi" "Ascomycota" "Sordariomycetes"
## LSUzotu66:Sordariomycetes_spc  "Fungi" "Ascomycota" "Sordariomycetes"
## LSUzotu54:Sporothrix_stenoceras "Fungi" "Ascomycota" "Sordariomycetes"
## LSUzotu138:Sporothrix_stenoceras "Fungi" "Ascomycota" "Sordariomycetes"
## LSUzotu158:Sporothrix_stenoceras "Fungi" "Ascomycota" "Sordariomycetes"
## LSUzotu128:Sordariomycetes_spc  "Fungi" "Ascomycota" "Sordariomycetes"
## LSUzotu198:Sordariomycetes_spc  "Fungi" "Ascomycota" "Sordariomycetes"
##
##      Order  Family
## LSUzotu6:Sporothrix_stenoceras "Ophiostomatales" "Ophiostomataceae"
## LSUzotu1:Chaetomium_globosum   "Sordariales"     "Chaetomiaceae"
## LSUzotu4:Raffaelea_canadensis  "Ophiostomatales" "Ophiostomataceae"
## LSUzotu2:Raffaelea_sulphurea   "Ophiostomatales" "Ophiostomataceae"
## LSUzotu66:Sordariomycetes_spc  "Sordariomycetes_spc" "Sordariomycetes_spc"
```

```

## LSUzotu54:Sporothrix_stenoceras "Ophiostomatales" "Ophiostomataceae"
## LSUzotu138:Sporothrix_stenoceras "Ophiostomatales" "Ophiostomataceae"
## LSUzotu158:Sporothrix_stenoceras "Ophiostomatales" "Ophiostomataceae"
## LSUzotu128:Sordariomycetes_spc "Sordariomycetes_spc" "Sordariomycetes_spc"
## LSUzotu198:Sordariomycetes_spc "Sordariomycetes_spc" "Sordariomycetes_spc"
##
## Genus Species
## LSUzotu6:Sporothrix_stenoceras "Sporothrix" "Sporothrix_stenoceras"
## LSUzotu1:Chaetomium_globosum "Chaetomium" "Chaetomium_globosum"
## LSUzotu4:Raffaelea_canadensis "Raffaelea" "Raffaelea_canadensis"
## LSUzotu2:Raffaelea_sulphurea "Raffaelea" "Raffaelea_sulphurea"
## LSUzotu66:Sordariomycetes_spc "Sordariomycetes_spc" "Sordariomycetes_spc"
## LSUzotu54:Sporothrix_stenoceras "Sporothrix" "Sporothrix_stenoceras"
## LSUzotu138:Sporothrix_stenoceras "Sporothrix" "Sporothrix_stenoceras"
## LSUzotu158:Sporothrix_stenoceras "Sporothrix" "Sporothrix_stenoceras"
## LSUzotu128:Sordariomycetes_spc "Sordariomycetes_spc" "Sordariomycetes_spc"
## LSUzotu198:Sordariomycetes_spc "Sordariomycetes_spc" "Sordariomycetes_spc"
##
## best_hit
## LSUzotu6:Sporothrix_stenoceras "LSUzotu6:Sporothrix_stenoceras"
## LSUzotu1:Chaetomium_globosum "LSUzotu1:Chaetomium_globosum"
## LSUzotu4:Raffaelea_canadensis "LSUzotu4:Raffaelea_canadensis"
## LSUzotu2:Raffaelea_sulphurea "LSUzotu2:Raffaelea_sulphurea"
## LSUzotu66:Sordariomycetes_spc "LSUzotu66:Sordariomycetes_spc"
## LSUzotu54:Sporothrix_stenoceras "LSUzotu54:Sporothrix_stenoceras"
## LSUzotu138:Sporothrix_stenoceras "LSUzotu138:Sporothrix_stenoceras"
## LSUzotu158:Sporothrix_stenoceras "LSUzotu158:Sporothrix_stenoceras"
## LSUzotu128:Sordariomycetes_spc "LSUzotu128:Sordariomycetes_spc"
## LSUzotu198:Sordariomycetes_spc "LSUzotu198:Sordariomycetes_spc"

```

```
round(otu_table(dataset.LSU.ordi)[filtaxa], digits = 4)
```

```

## OTU Table:          [10 taxa and 57 samples]
##                      taxa are rows
##
## B0-19a B0-19b B0-22a B0-22b B0-24 B0-27 B10-01
## LSUzotu6:Sporothrix_stenoceras      0      0      2      0      0      0      0
## LSUzotu1:Chaetomium_globosum    258  19888  19631  34239  7239  1340   170
## LSUzotu4:Raffaelea_canadensis     251      1     11   2284  2764  3024    13
## LSUzotu2:Raffaelea_sulphurea    7705      9     22      6 15410 10421 16220
## LSUzotu66:Sordariomycetes_spc      0      0      0      0      0      0      0
## LSUzotu54:Sporothrix_stenoceras     0      0      0      0      0      0      0
## LSUzotu138:Sporothrix_stenoceras    0      0      0      0      0      0      0
## LSUzotu158:Sporothrix_stenoceras    0      0      0      0      0      0      0
## LSUzotu128:Sordariomycetes_spc     0      0      0      0      0      0      0
## LSUzotu198:Sordariomycetes_spc     0      0      0      0      0      0      0
##
## B10-06 B10-09a B10-09b B10-18 B10-24 B10-25a
## LSUzotu6:Sporothrix_stenoceras      0      1      0      0      0      3
## LSUzotu1:Chaetomium_globosum    11309  18613  35695  2094  1159   6733
## LSUzotu4:Raffaelea_canadensis      0      1      0      2      0      2
## LSUzotu2:Raffaelea_sulphurea    13569   5195      7  3573  7565  13414
## LSUzotu66:Sordariomycetes_spc      0      0      0      0      0      0
## LSUzotu54:Sporothrix_stenoceras     0      0      0      0      0      0
## LSUzotu138:Sporothrix_stenoceras    0      0      0      0      0      0
## LSUzotu158:Sporothrix_stenoceras    0      0      0      0      0      0
## LSUzotu128:Sordariomycetes_spc     0      0      0      0      0      0
## LSUzotu198:Sordariomycetes_spc     0      0      0      0      0      0

```



##	B10-25b	B10-26a	B15-01a	B15-01b	B15-06	B15-07a
## LSUzotu6:Sporothrix_stenoceras	0	0	28	0	0	0
## LSUzotu1:Chaetomium_globosum	95	4421	33	20163	4	15782
## LSUzotu4:Raffaelea_canadensis	0	1	1260	257	524	178
## LSUzotu2:Raffaelea_sulphurea	7353	18965	9796	2454	1589	2382
## LSUzotu66:Sordariomycetes_spc	0	0	1	0	0	0
## LSUzotu54:Sporothrix_stenoceras	0	0	1	0	0	0
## LSUzotu138:Sporothrix_stenoceras	0	0	0	0	0	0
## LSUzotu158:Sporothrix_stenoceras	0	0	0	0	0	0
## LSUzotu128:Sordariomycetes_spc	0	0	0	0	0	0
## LSUzotu198:Sordariomycetes_spc	0	0	0	0	0	0
##	B15-07b	B15-11	B15-20	B15-24	B15-25a	B15-25b
## LSUzotu6:Sporothrix_stenoceras	0	0	0	0	1	0
## LSUzotu1:Chaetomium_globosum	3284	4560	7674	3569	466	3571
## LSUzotu4:Raffaelea_canadensis	1572	4372	1616	7115	1663	3320
## LSUzotu2:Raffaelea_sulphurea	4285	8639	3977	8446	5434	14262
## LSUzotu66:Sordariomycetes_spc	0	0	0	0	0	0
## LSUzotu54:Sporothrix_stenoceras	0	0	0	0	0	0
## LSUzotu138:Sporothrix_stenoceras	0	0	0	0	0	0
## LSUzotu158:Sporothrix_stenoceras	0	0	0	0	0	0
## LSUzotu128:Sordariomycetes_spc	0	0	0	0	0	0
## LSUzotu198:Sordariomycetes_spc	0	0	0	0	0	0
##	B16-01a	B16-04	B16-31a	B16-31b	B20-04a	B20-08a
## LSUzotu6:Sporothrix_stenoceras	3	0	23	0	0	0
## LSUzotu1:Chaetomium_globosum	20350	33	10485	10239	34162	21
## LSUzotu4:Raffaelea_canadensis	1	1	29	4	1	3
## LSUzotu2:Raffaelea_sulphurea	3	7443	793	33	30	49
## LSUzotu66:Sordariomycetes_spc	0	0	0	0	0	0
## LSUzotu54:Sporothrix_stenoceras	0	0	0	0	0	0
## LSUzotu138:Sporothrix_stenoceras	0	0	2	0	0	0
## LSUzotu158:Sporothrix_stenoceras	0	0	0	0	0	0
## LSUzotu128:Sordariomycetes_spc	0	0	0	0	0	0
## LSUzotu198:Sordariomycetes_spc	0	0	0	0	0	0
##	B20-08b	B20-11	B23-21	B23-23a	B24-02a	B24-26a
## LSUzotu6:Sporothrix_stenoceras	0	0	0	0	0	0
## LSUzotu1:Chaetomium_globosum	3445	6388	22924	8567	1031	11771
## LSUzotu4:Raffaelea_canadensis	0	2	2	1	1	0
## LSUzotu2:Raffaelea_sulphurea	7	3	4734	14764	43	2542
## LSUzotu66:Sordariomycetes_spc	0	0	0	0	0	0
## LSUzotu54:Sporothrix_stenoceras	0	0	0	0	0	0
## LSUzotu138:Sporothrix_stenoceras	0	0	0	0	0	0
## LSUzotu158:Sporothrix_stenoceras	0	0	0	0	0	0
## LSUzotu128:Sordariomycetes_spc	0	0	0	0	0	0
## LSUzotu198:Sordariomycetes_spc	0	0	0	0	0	0
##	B24-26b	B24-28a	B24-29	B36-14a	B36-14b	B36-30
## LSUzotu6:Sporothrix_stenoceras	0	0	0	0	0	0
## LSUzotu1:Chaetomium_globosum	5163	6763	4311	3226	6297	6760
## LSUzotu4:Raffaelea_canadensis	0	29	1	3	0	3
## LSUzotu2:Raffaelea_sulphurea	3448	8373	2555	4297	15265	1367
## LSUzotu66:Sordariomycetes_spc	0	0	0	0	0	0
## LSUzotu54:Sporothrix_stenoceras	0	0	0	0	0	0
## LSUzotu138:Sporothrix_stenoceras	0	0	0	0	0	0
## LSUzotu158:Sporothrix_stenoceras	0	0	0	0	0	0
## LSUzotu128:Sordariomycetes_spc	0	0	0	0	0	0

## LSUzotu198:Sordariomycetes_spc	0	0	0	0	0	0
##	B36-31a	B36-32	B39-15	B39-16	B39-17a	B46-03
## LSUzotu6:Sporothrix_stenoceras	0	0	0	0	0	0
## LSUzotu1:Chaetomium_globosum	16058	367	5922	2153	11890	20657
## LSUzotu4:Raffaelea_canadensis	3	0	2	0	5	1
## LSUzotu2:Raffaelea_sulphurea	499	699	6225	7681	130	15
## LSUzotu66:Sordariomycetes_spc	0	0	0	0	0	0
## LSUzotu54:Sporothrix_stenoceras	0	0	0	0	0	0
## LSUzotu138:Sporothrix_stenoceras	0	0	0	0	0	0
## LSUzotu158:Sporothrix_stenoceras	0	0	0	0	0	0
## LSUzotu128:Sordariomycetes_spc	0	0	0	0	0	0
## LSUzotu198:Sordariomycetes_spc	0	0	0	0	0	0
##	B7-07	B7-08a	medium1	medium2	negative2	
## LSUzotu6:Sporothrix_stenoceras	0	0	0	0	1	
## LSUzotu1:Chaetomium_globosum	2230	4660	2	12	11	
## LSUzotu4:Raffaelea_canadensis	0	14	0	0	2	
## LSUzotu2:Raffaelea_sulphurea	3522	5	7	23	21	
## LSUzotu66:Sordariomycetes_spc	0	0	0	0	0	
## LSUzotu54:Sporothrix_stenoceras	0	0	0	0	0	
## LSUzotu138:Sporothrix_stenoceras	0	0	0	0	0	
## LSUzotu158:Sporothrix_stenoceras	0	0	0	0	0	
## LSUzotu128:Sordariomycetes_spc	0	0	0	0	0	
## LSUzotu198:Sordariomycetes_spc	0	0	0	0	0	
##	negative1	mock	standard			
## LSUzotu6:Sporothrix_stenoceras	1	15551	0			
## LSUzotu1:Chaetomium_globosum	219	4546	6			
## LSUzotu4:Raffaelea_canadensis	0	1822	0			
## LSUzotu2:Raffaelea_sulphurea	24	747	16			
## LSUzotu66:Sordariomycetes_spc	0	412	0			
## LSUzotu54:Sporothrix_stenoceras	0	163	0			
## LSUzotu138:Sporothrix_stenoceras	0	136	0			
## LSUzotu158:Sporothrix_stenoceras	0	112	0			
## LSUzotu128:Sordariomycetes_spc	0	56	0			
## LSUzotu198:Sordariomycetes_spc	0	38	0			

=> fungi from Zymo communitiy standard not sequenced, mock-community (R. sulphurea, R. canadensis, C. globosum, O. stenoceras, yeast (C. pichia)) represented in mock sample

## plot controls

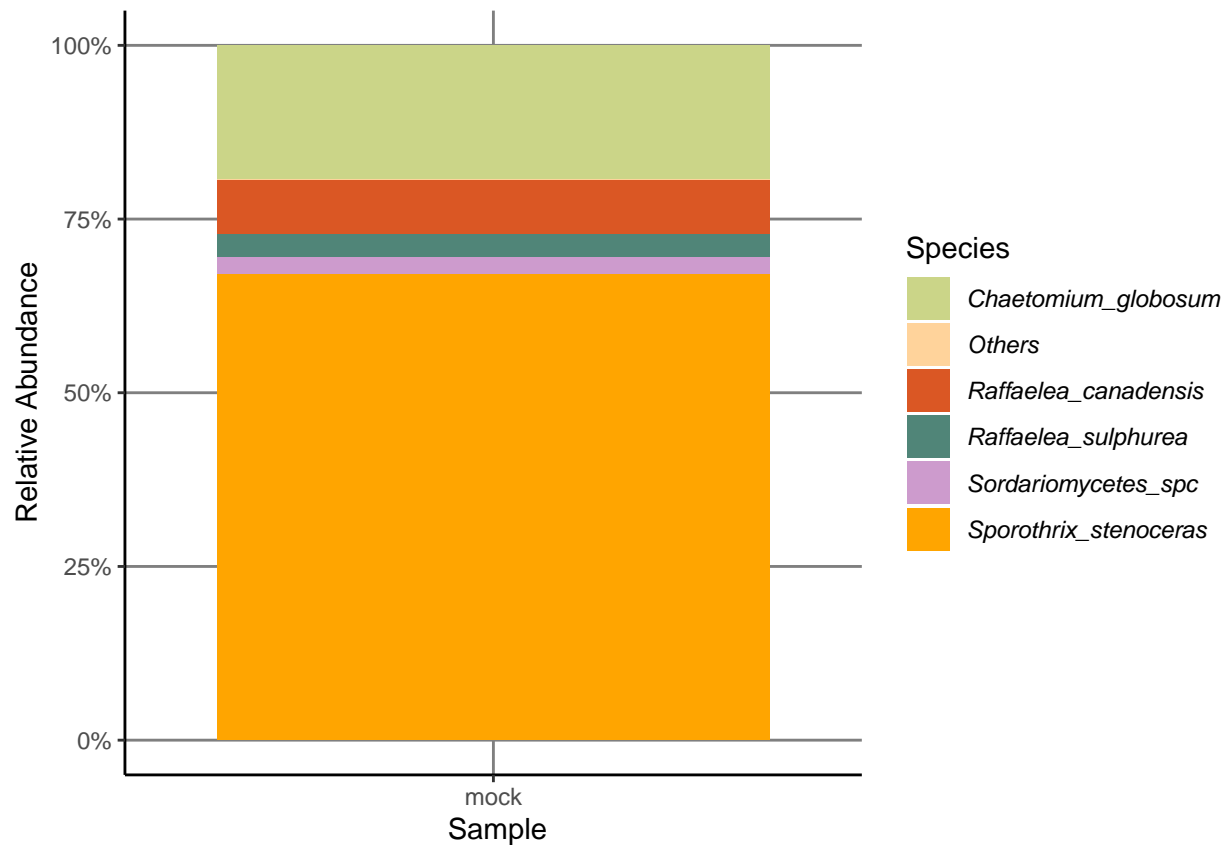
prepare data for plotting

```
mock <- subset_samples(dataset.LSU.ordi, Sample == "mock")
std <- subset_samples(dataset.LSU.ordi, Sample == "standard")

negs <- subset_samples(dataset.LSU.ordi, Treatment == "negative" | Treatment == "medium")

Fungi_Species.mock <- mock %>%
  tax_glom(taxrank = "Species") %>%
  transform_sample_counts(function(x) {x/sum(x)} ) %>%
  psmelt() %>%
  arrange(Species)
```

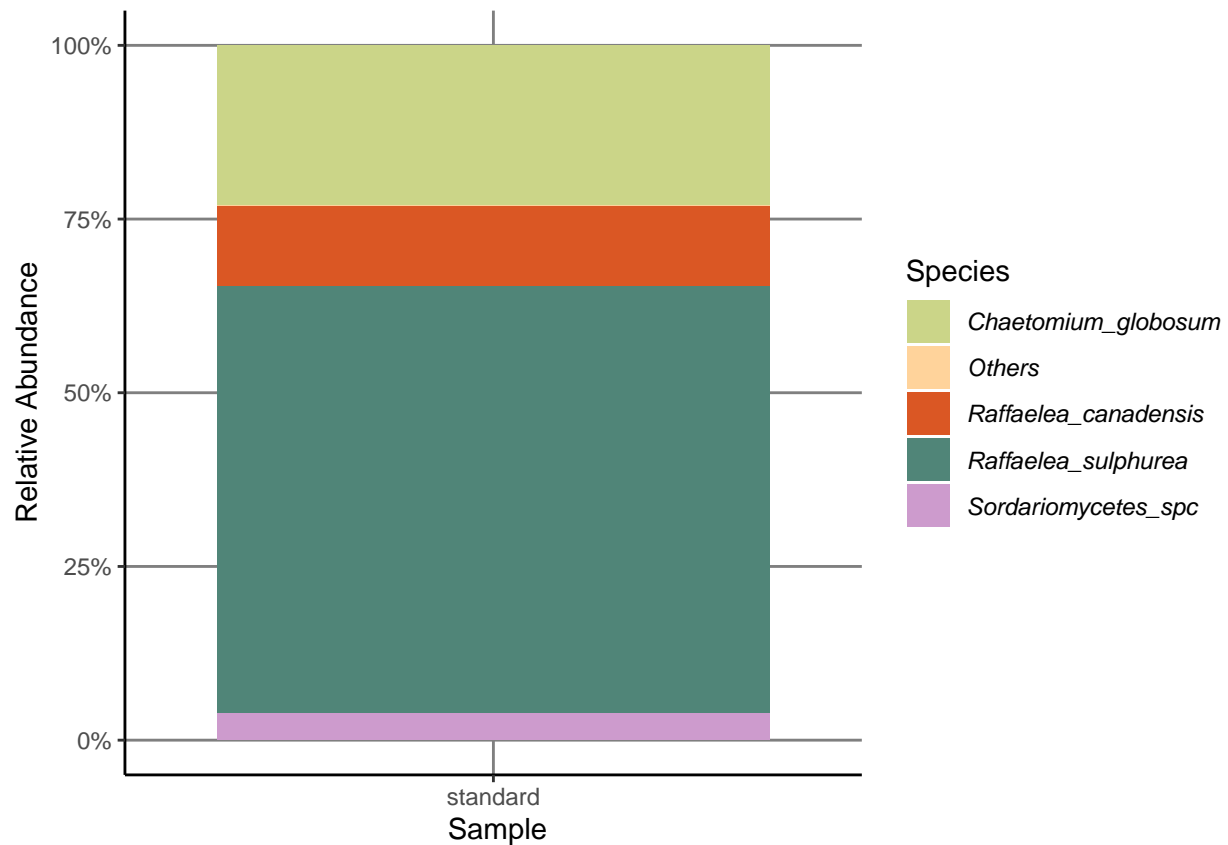




```
Fungi_Species.std_plot <-ggplot(Fungi_Species.std, aes(x = Sample, y = Abundance, fill = Species)) +
  geom_bar(stat = "identity", color = NA, position="fill") +
  scale_fill_manual(values = Family_colors, name = "Species")
```

```
g2<-Fungi_Species.std_plot +
  theme(plot.title = element_text(size = 20, face = "bold")) +
  theme(text = element_text(size=20, face = "bold"))+
  theme(axis.text.x=element_text(size = rel(0.5)))+
  theme(legend.title = element_text(size = 20), legend.text = element_text(size = 14))+
  theme_classic()+
  #gets rid of background
  theme(panel.grid.major = element_line(colour = "grey50"))+
  labs(x=" Sample", y="Relative Abundance")+
  scale_y_continuous(labels=percent_format())+
  theme(legend.text = element_text(face = "italic"))
```

g2

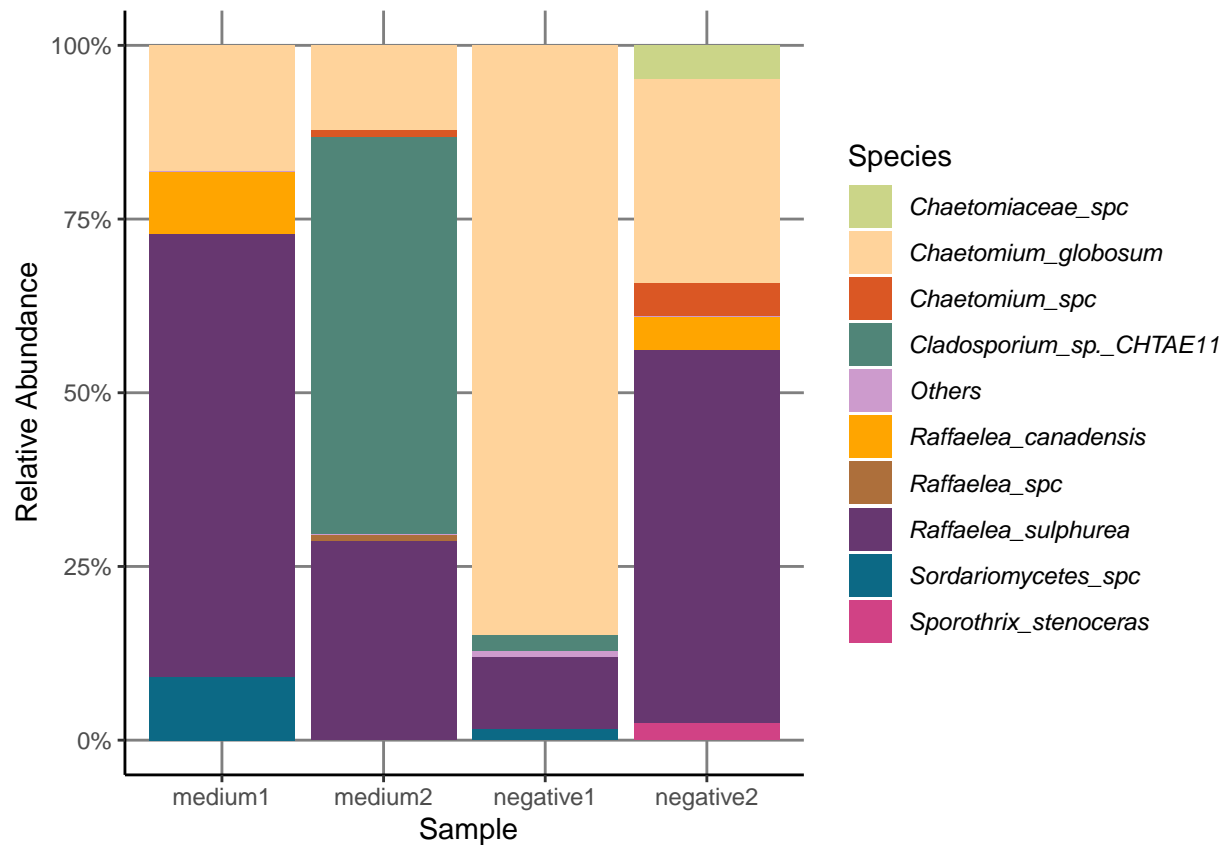


negative controls prae decontam

```
Fungi_Species.neg_plot <-ggplot(Fungi_Species.neg, aes(x = Sample, y = Abundance, fill = Species)) +
  geom_bar(stat = "identity", color = NA, position="fill") +
  scale_fill_manual(values = Family_colors, name = "Species")

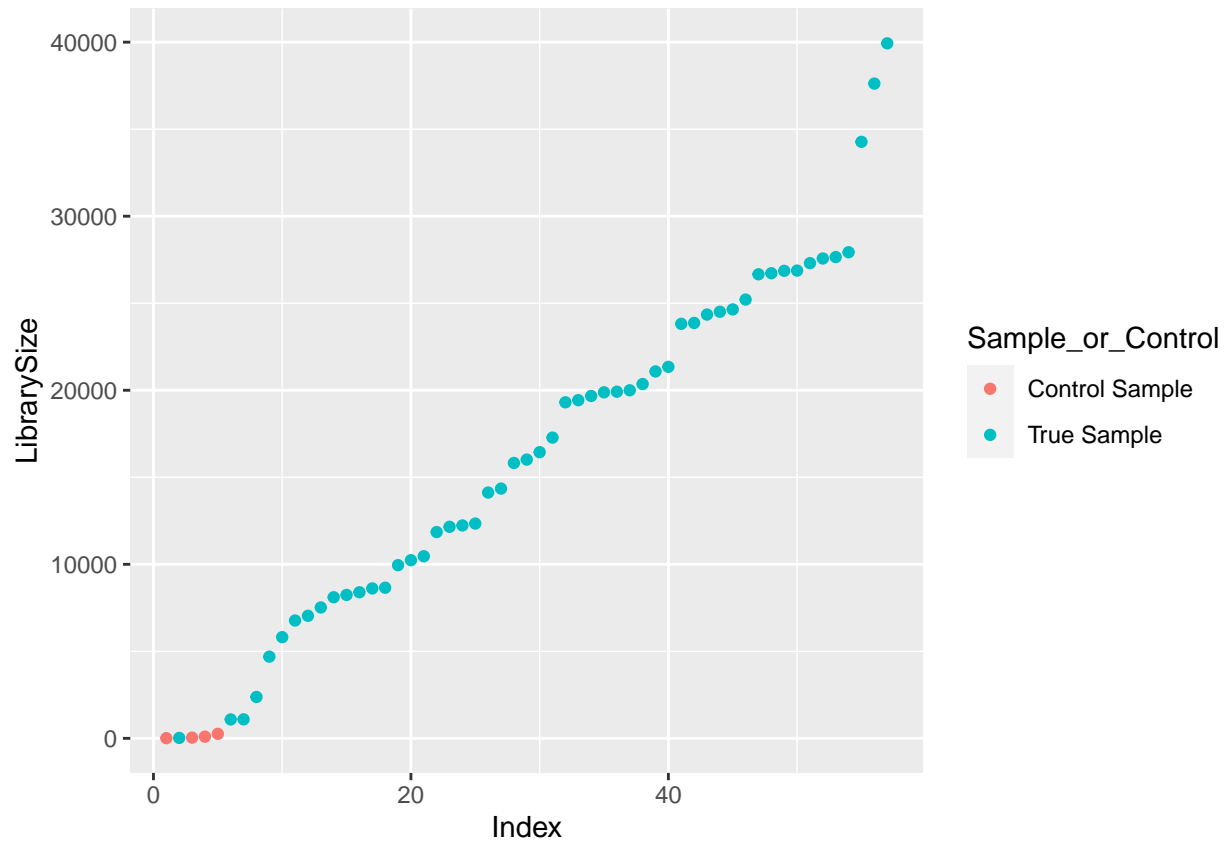
g3<-Fungi_Species.neg_plot +
  theme(plot.title = element_text(size = 20, face = "bold")) +
  theme(text = element_text(size=20, face = "bold"))+
  theme(axis.text.x=element_text(size = rel(0.5)))+
  theme(legend.title = element_text(size = 20), legend.text = element_text(size = 14))+
  theme_classic()+
  #gets rid of background
  theme(panel.grid.major = element_line(colour = "grey50"))+
  labs(x=" Sample", y="Relative Abundance")+
  scale_y_continuous(labels=percent_format())+
  theme(legend.text = element_text(face = "italic"))

g3
```



## run decontam

```
library(decontam)
df <- as.data.frame(sample_data(dataset.LSU.ordi)) # Put sample_data into a ggplot-friendly data.frame
df$LibrarySize <- sample_sums(dataset.LSU.ordi)
df <- df[order(df$LibrarySize),]
df$Index <- seq(nrow(df))
ggplot(data=df, aes(x=Index, y=LibrarySize, color=Sample_or_Control)) + geom_point()
```



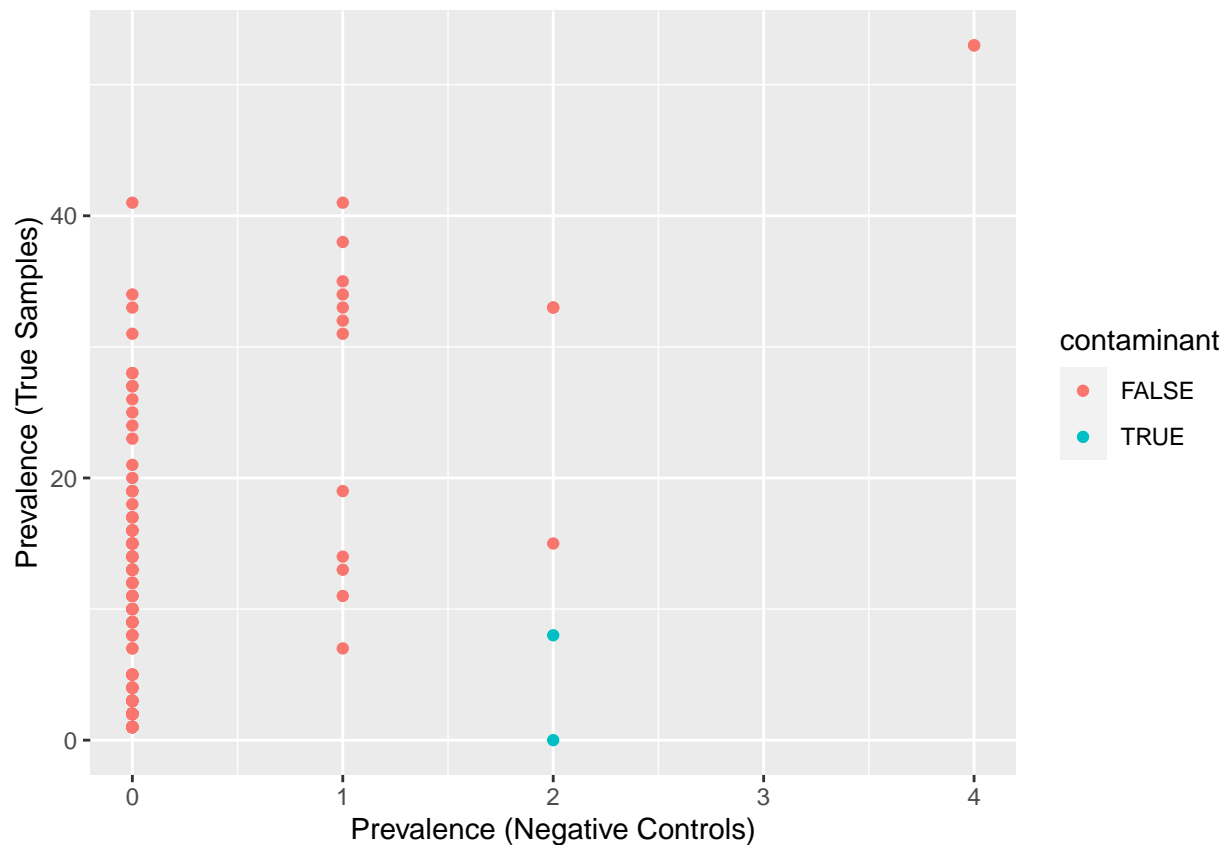
```
sample_data(dataset.LSU.ordi)$is.neg <- sample_data(dataset.LSU.ordi)$Sample_or_Control == "Control Sample"
contamdf.prev <- isContaminant(dataset.LSU.ordi, method="prevalence", neg="is.neg")
table(contamdf.prev$contaminant)
```

```
##
## FALSE TRUE
##    209    2
```

```
ps.pa <- transform_sample_counts(dataset.LSU.ordi, function(abund) 1*(abund>0))
ps.pa.neg <- prune_samples(sample_data(ps.pa)$Sample_or_Control == "Control Sample", ps.pa)
ps.pa.pos <- prune_samples(sample_data(ps.pa)$Sample_or_Control == "True Sample", ps.pa)
```

Make data.frame of prevalence in positive and negative samples

```
df.pa <- data.frame(pa.pos=taxa_sums(ps.pa.pos), pa.neg=taxa_sums(ps.pa.neg), contaminant=contamdf.prev$contaminant)
ggplot(data=df.pa, aes(x=pa.neg, y=pa.pos, color=contaminant)) + geom_point() + xlab("Prevalence (Negative)") + ylab("Prevalence (Positive)")
```



```
ps.noncontam_dataset.LSU.ordi <- prune_taxa(!contamdf.prev$contaminant, dataset.LSU.ordi)
ps.noncontam_dataset.LSU.ordi
```

```
## phyloseq-class experiment-level object
## otu_table() OTU Table: [ 209 taxa and 57 samples ]
## sample_data() Sample Data: [ 57 samples by 17 sample variables ]
## tax_table() Taxonomy Table: [ 209 taxa by 8 taxonomic ranks ]
```

```
smin <- min(sample_sums(ps.noncontam_dataset.LSU.ordi))
smean <- mean(sample_sums(ps.noncontam_dataset.LSU.ordi))
smax <- max(sample_sums(ps.noncontam_dataset.LSU.ordi))
cat("The minimum sample read count is:",smin) #11
```

```
## The minimum sample read count is: 11
```

```
cat("The average sample read count is:",smean) #15670.75
```

```
## The average sample read count is: 15670.75
```

```
cat("The maximum sample read count is:",smax) #39932
```

```
## The maximum sample read count is: 39932
```



```
contaminants <- subset(contamdf.prev, contaminant == "TRUE")
```

create a list of all excluded contaminants

check controls post decontam

prepare data for plotting

```
mock2 <- subset_samples(ps.noncontam_dataset.LSU.ordi, Sample == "mock")
std2 <- subset_samples(ps.noncontam_dataset.LSU.ordi, Sample == "standard")

negs2 <- subset_samples(ps.noncontam_dataset.LSU.ordi, Treatment == "negative" | Treatment == "medium")

Fungi_Species.mock2 <- mock2 %>%
  tax_glom(taxrank = "Species") %>%
  transform_sample_counts(function(x) {x/sum(x)} ) %>%
  psmelt() %>%
  arrange(Species)

Fungi_Species.mock2$Species<-as.character(Fungi_Species.mock2$Species)
Fungi_Species.mock2$Species[Fungi_Species.mock2$Abundance<0.01]<- "Others"
Fungi_Species.mock2$Class<-as.character(Fungi_Species.mock2$Class)
Fungi_Species.mock2$Class[Fungi_Species.mock2$Abundance<0.01]<- "Others"

Fungi_Species.std2 <- std2 %>%
  tax_glom(taxrank = "Species") %>%
  transform_sample_counts(function(x) {x/sum(x)} ) %>%
  psmelt() %>%
  arrange(Species)

Fungi_Species.std2$Species<-as.character(Fungi_Species.std2$Species)
Fungi_Species.std2$Species[Fungi_Species.std2$Abundance<0.01]<- "Others"
Fungi_Species.std2$Class<-as.character(Fungi_Species.std2$Class)
Fungi_Species.std2$Class[Fungi_Species.std2$Abundance<0.01]<- "Others"

Fungi_Species.neg2 <- negs2 %>%
  tax_glom(taxrank = "Species") %>%
  transform_sample_counts(function(x) {x/sum(x)} ) %>%
  psmelt() %>%
  arrange(Species)

Fungi_Species.neg2$Species<-as.character(Fungi_Species.neg2$Species)
Fungi_Species.neg2$Species[Fungi_Species.neg2$Abundance<0.01]<- "Others"
Fungi_Species.neg2$Class<-as.character(Fungi_Species.neg2$Class)
Fungi_Species.neg2$Class[Fungi_Species.neg2$Abundance<0.01]<- "Others"
```

positive controls post decontam

```
Fungi_Species.mock2_plot <-ggplot(Fungi_Species.mock2, aes(x = Sample, y = Abundance, fill = Species)) +
  geom_bar(stat = "identity", color = NA, position="fill") +
```

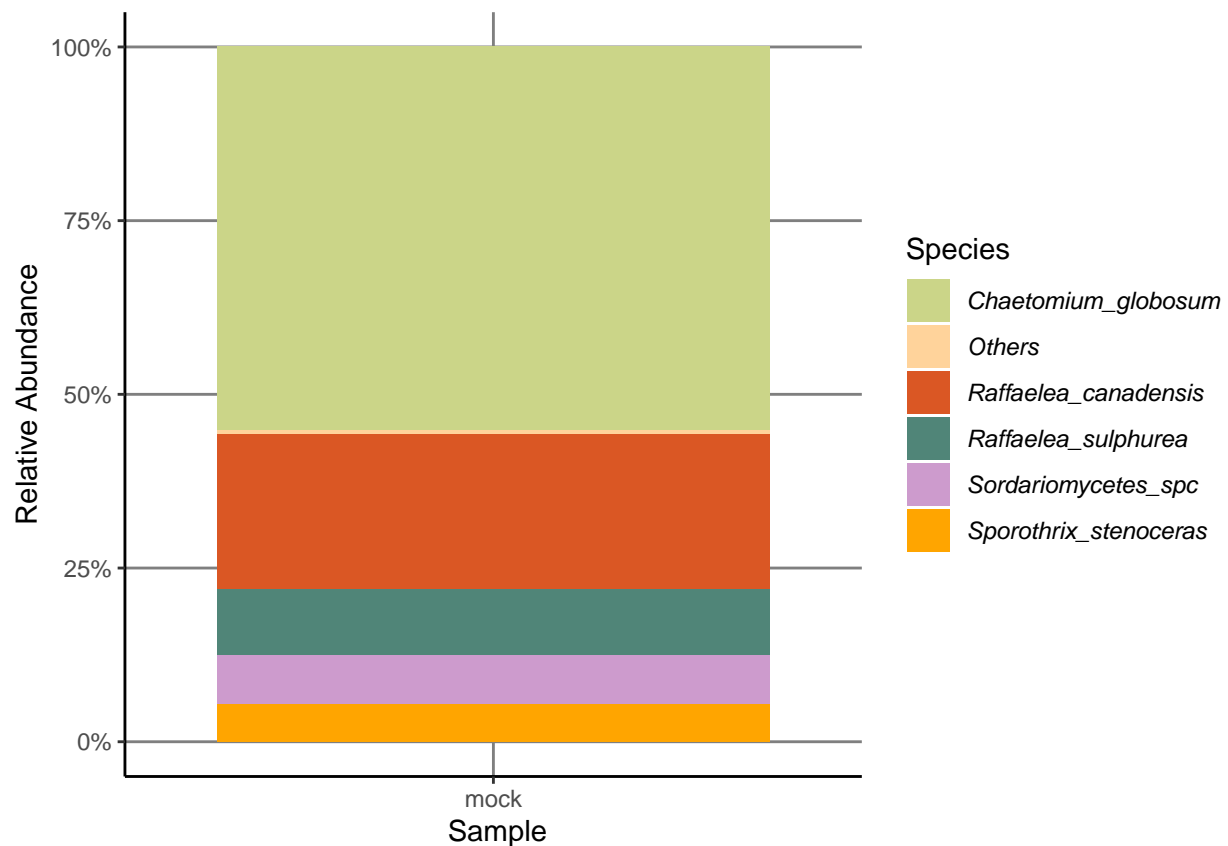
```

scale_fill_manual(values = Family_colors, name = "Species")

g4<-Fungi_Species.mock2_plot +
  theme(plot.title = element_text(size = 20, face = "bold")) +
  theme(text = element_text(size=20, face = "bold"))+
  theme(axis.text.x=element_text(size = rel(0.5)))+
  theme(legend.title = element_text(size = 20), legend.text = element_text(size = 14))+
  theme_classic()+
  theme(panel.grid.major = element_line(colour = "grey50"))+
  labs(x=" Sample", y="Relative Abundance")+
  scale_y_continuous(labels=percent_format())+
  theme(legend.text = element_text(face = "italic"))

```

g4



standard controls post decontam

```

Fungi_Species.std2_plot <-ggplot(Fungi_Species.std2, aes(x = Sample, y = Abundance, fill = Species)) +
  geom_bar(stat = "identity", color = NA, position="fill") +
  scale_fill_manual(values = Family_colors, name = "Species")

g5<-Fungi_Species.std2_plot +
  theme(plot.title = element_text(size = 20, face = "bold")) +
  theme(text = element_text(size=20, face = "bold"))+
  theme(axis.text.x=element_text(size = rel(0.5)))+
  theme(legend.title = element_text(size = 20), legend.text = element_text(size = 14))+

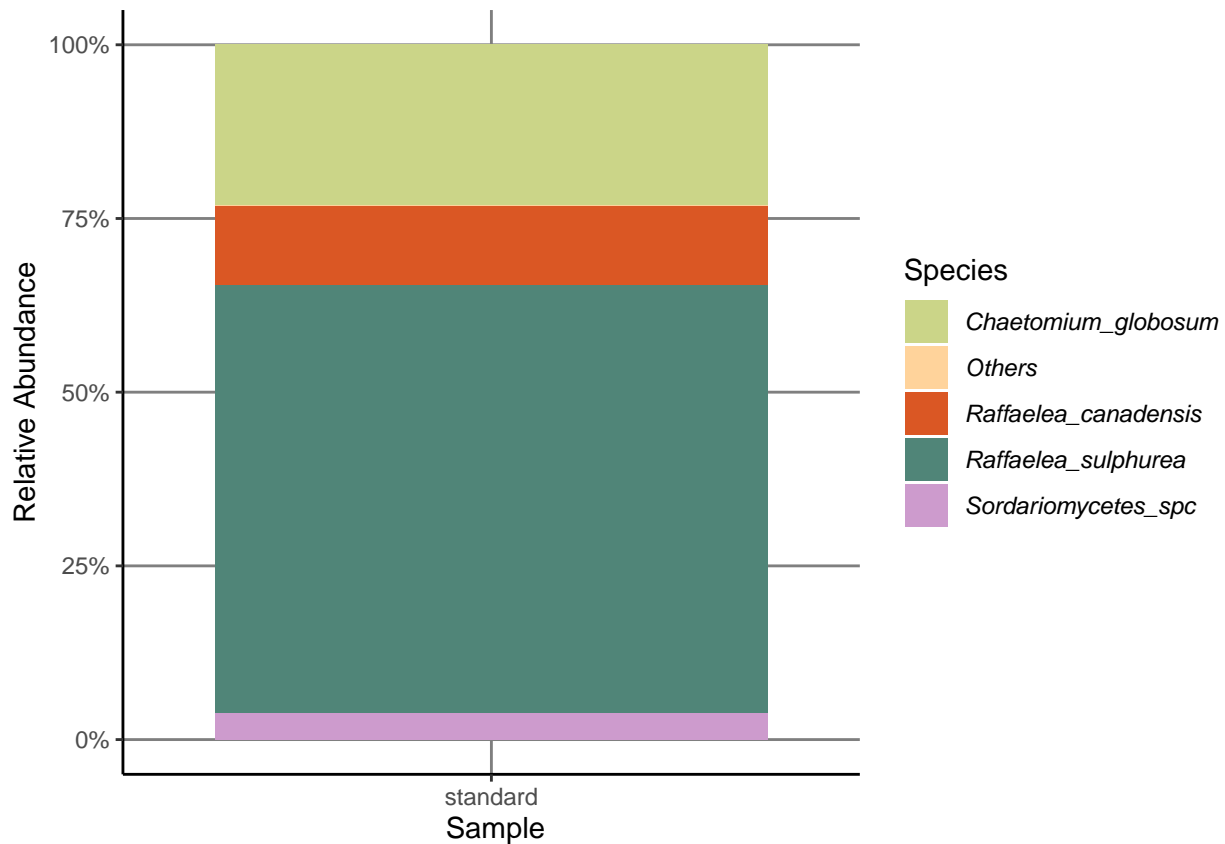
```

```

theme_classic()+                #gets rid of background
theme(panel.grid.major = element_line(colour = "grey50"))+
labs(x=" Sample", y="Relative Abundance")+
scale_y_continuous(labels=percent_format())+
theme(legend.text = element_text(face = "italic"))

```

g5



negative controls post decontam

```

Fungi_Species.neg2_plot <-ggplot(Fungi_Species.neg2, aes(x = Sample, y = Abundance, fill = Species)) +
  geom_bar(stat = "identity", color = NA, position="fill") +
  scale_fill_manual(values = Family_colors, name = "Species")

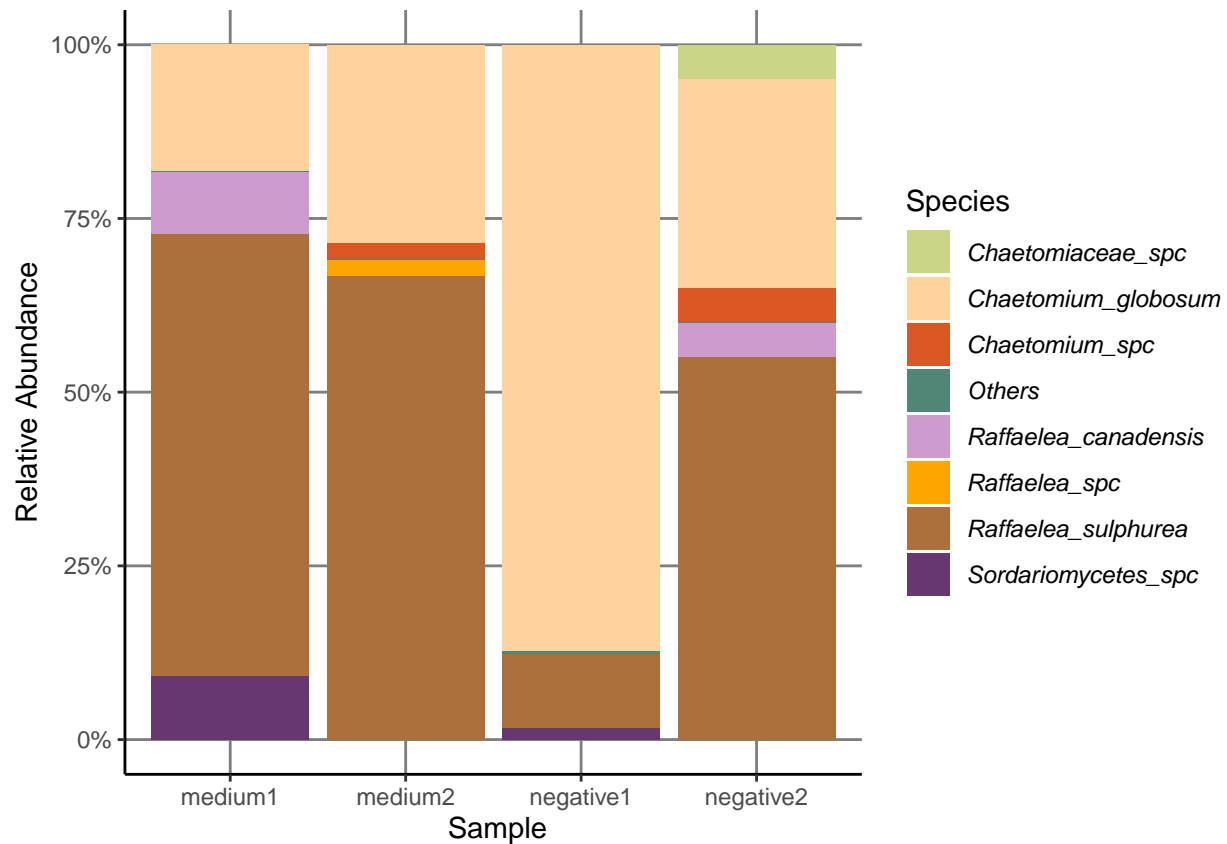
```

```

g6<-Fungi_Species.neg2_plot +
  theme(plot.title = element_text(size = 20, face = "bold")) +
  theme(text = element_text(size=20, face = "bold"))+
  theme(axis.text.x=element_text(size = rel(0.5)))+
  theme(legend.title = element_text(size = 20), legend.text = element_text(size = 14))+
  theme_classic()+                #gets rid of background
  theme(panel.grid.major = element_line(colour = "grey50"))+
  labs(x=" Sample", y="Relative Abundance")+
  scale_y_continuous(labels=percent_format())+
  theme(legend.text = element_text(face = "italic"))

```

g6



exclude negative and positive samples for a look of the ten most abundant ZOTUS in our samples

```
fun.without.controls <- subset_samples(ps.noncontam_dataset.LSU.ordi, Treatment!="medium")
fun.without.controls2 <- subset_samples(fun.without.controls, Treatment!="negative")
fun.without.controls3 <- subset_samples(fun.without.controls2, Treatment!="mock")
fwc <- subset_samples(fun.without.controls3, Treatment!="standard")
```

get rid of no read taxa and unimportant metadata in dataset

```
fwc <- prune_taxa(taxa_sums(fwc) > 0, fwc)

sample_data(fwc) <- sample_data(fwc)[,c("Sample", "Nest", "Lineage", "Treatment", "age_sampling_d.", "Group")]
```

now have a look at the 15 most abundant ZOTUS in our samples

```
most.abundant<-subset_samples(fwc, Treatment=="control" | Treatment=="removal" | Treatment=="2nd-founda
```

```
filtaxa <- names (sort(rowSums(otu_table(most.abundant)) ,decreasing=T))[1:15]
tax_table(fwc)[filtaxa]
```

```
## Taxonomy Table:      [15 taxa by 8 taxonomic ranks]:
##
##      Domain  Phylum      Class
## LSUzotu1:Chaetomium_globosum "Fungi" "Ascomycota" "Sordariomycetes"
## LSUzotu2:Raffaelea_sulphurea "Fungi" "Ascomycota" "Sordariomycetes"
## LSUzotu3:Acremonium_biseptum "Fungi" "Ascomycota" "Sordariomycetes"
## LSUzotu4:Raffaelea_canadensis "Fungi" "Ascomycota" "Sordariomycetes"
## LSUzotu5:Raffaelea_canadensis "Fungi" "Ascomycota" "Sordariomycetes"
## LSUzotu8:Penicillium_commune "Fungi" "Ascomycota" "Eurotiomycetes"
## LSUzotu19:Chaetomium_globosum "Fungi" "Ascomycota" "Sordariomycetes"
## LSUzotu23:Raffaelea_sulphurea "Fungi" "Ascomycota" "Sordariomycetes"
## LSUzotu25:Raffaelea_sulphurea "Fungi" "Ascomycota" "Sordariomycetes"
## LSUzotu7:Raffaelea_sulphurea "Fungi" "Ascomycota" "Sordariomycetes"
## LSUzotu26:Chaetomium_spc "Fungi" "Ascomycota" "Sordariomycetes"
## LSUzotu9:Raffaelea_spc "Fungi" "Ascomycota" "Sordariomycetes"
## LSUzotu36:Raffaelea_sulphurea "Fungi" "Ascomycota" "Sordariomycetes"
## LSUzotu11:Raffaelea_canadensis "Fungi" "Ascomycota" "Sordariomycetes"
## LSUzotu37:Raffaelea_canadensis "Fungi" "Ascomycota" "Sordariomycetes"
##
##      Order      Family
## LSUzotu1:Chaetomium_globosum "Sordariales" "Chaetomiaceae"
## LSUzotu2:Raffaelea_sulphurea "Ophiostomatales" "Ophiostomataceae"
## LSUzotu3:Acremonium_biseptum "Hypocreales" "Bionectriaceae"
## LSUzotu4:Raffaelea_canadensis "Ophiostomatales" "Ophiostomataceae"
## LSUzotu5:Raffaelea_canadensis "Ophiostomatales" "Ophiostomataceae"
## LSUzotu8:Penicillium_commune "Eurotiales" "Trichocomaceae"
## LSUzotu19:Chaetomium_globosum "Sordariales" "Chaetomiaceae"
## LSUzotu23:Raffaelea_sulphurea "Ophiostomatales" "Ophiostomataceae"
## LSUzotu25:Raffaelea_sulphurea "Ophiostomatales" "Ophiostomataceae"
## LSUzotu7:Raffaelea_sulphurea "Ophiostomatales" "Ophiostomataceae"
## LSUzotu26:Chaetomium_spc "Sordariales" "Chaetomiaceae"
## LSUzotu9:Raffaelea_spc "Ophiostomatales" "Ophiostomataceae"
## LSUzotu36:Raffaelea_sulphurea "Ophiostomatales" "Ophiostomataceae"
## LSUzotu11:Raffaelea_canadensis "Ophiostomatales" "Ophiostomataceae"
## LSUzotu37:Raffaelea_canadensis "Ophiostomatales" "Ophiostomataceae"
##
##      Genus      Species
## LSUzotu1:Chaetomium_globosum "Chaetomium" "Chaetomium_globosum"
## LSUzotu2:Raffaelea_sulphurea "Raffaelea" "Raffaelea_sulphurea"
## LSUzotu3:Acremonium_biseptum "Acremonium" "Acremonium_biseptum"
## LSUzotu4:Raffaelea_canadensis "Raffaelea" "Raffaelea_canadensis"
## LSUzotu5:Raffaelea_canadensis "Raffaelea" "Raffaelea_canadensis"
## LSUzotu8:Penicillium_commune "Penicillium" "Penicillium_commune"
## LSUzotu19:Chaetomium_globosum "Chaetomium" "Chaetomium_globosum"
## LSUzotu23:Raffaelea_sulphurea "Raffaelea" "Raffaelea_sulphurea"
## LSUzotu25:Raffaelea_sulphurea "Raffaelea" "Raffaelea_sulphurea"
## LSUzotu7:Raffaelea_sulphurea "Raffaelea" "Raffaelea_sulphurea"
## LSUzotu26:Chaetomium_spc "Chaetomium" "Chaetomium_spc"
## LSUzotu9:Raffaelea_spc "Raffaelea" "Raffaelea_spc"
## LSUzotu36:Raffaelea_sulphurea "Raffaelea" "Raffaelea_sulphurea"
## LSUzotu11:Raffaelea_canadensis "Raffaelea" "Raffaelea_canadensis"
## LSUzotu37:Raffaelea_canadensis "Raffaelea" "Raffaelea_canadensis"
##
## best_hit
```

```
## LSUzotu1:Chaetomium_globosum "LSUzotu1:Chaetomium_globosum"
## LSUzotu2:Raffaelea_sulphurea "LSUzotu2:Raffaelea_sulphurea"
## LSUzotu3:Acremonium_biseptum "LSUzotu3:Acremonium_biseptum"
## LSUzotu4:Raffaelea_canadensis "LSUzotu4:Raffaelea_canadensis"
## LSUzotu5:Raffaelea_canadensis "LSUzotu5:Raffaelea_canadensis"
## LSUzotu8:Penicillium_commune "LSUzotu8:Penicillium_commune"
## LSUzotu19:Chaetomium_globosum "LSUzotu19:Chaetomium_globosum"
## LSUzotu23:Raffaelea_sulphurea "LSUzotu23:Raffaelea_sulphurea"
## LSUzotu25:Raffaelea_sulphurea "LSUzotu25:Raffaelea_sulphurea"
## LSUzotu7:Raffaelea_sulphurea "LSUzotu7:Raffaelea_sulphurea"
## LSUzotu26:Chaetomium_spc "LSUzotu26:Chaetomium_spc"
## LSUzotu9:Raffaelea_spc "LSUzotu9:Raffaelea_spc"
## LSUzotu36:Raffaelea_sulphurea "LSUzotu36:Raffaelea_sulphurea"
## LSUzotu11:Raffaelea_canadensis "LSUzotu11:Raffaelea_canadensis"
## LSUzotu37:Raffaelea_canadensis "LSUzotu37:Raffaelea_canadensis"
```

```
round(otu_table(fwc)[filtaxa], digits = 4)
```

```
## OTU Table:          [15 taxa and 51 samples]
##                      taxa are rows
##
##                      B0-19a B0-19b B0-22a B0-22b B0-24 B0-27 B10-01
## LSUzotu1:Chaetomium_globosum      258  19888  19631  34239  7239  1340   170
## LSUzotu2:Raffaelea_sulphurea     7705    9      22    6 15410 10421 16220
## LSUzotu3:Acremonium_biseptum        0    0    0    0    0    0    0
## LSUzotu4:Raffaelea_canadensis     251    1    11   2284  2764  3024   13
## LSUzotu5:Raffaelea_canadensis        0   45    4    0    1    1  3221
## LSUzotu8:Penicillium_commune        0    0    0    0    0    0    0
## LSUzotu19:Chaetomium_globosum        0    0    0    2   99   20    3
## LSUzotu23:Raffaelea_sulphurea        0    0    0    0  172   30    6
## LSUzotu25:Raffaelea_sulphurea        0    0    0    0   97   15    4
## LSUzotu7:Raffaelea_sulphurea     386    0    0    0  548  380  109
## LSUzotu26:Chaetomium_spc            0    0    0    0   86   22    2
## LSUzotu9:Raffaelea_spc              2    0    0    0  170  192    2
## LSUzotu36:Raffaelea_sulphurea        0    0    0    0   63   19    2
## LSUzotu11:Raffaelea_canadensis        0    0    0    0    0    0  535
## LSUzotu37:Raffaelea_canadensis        3    0    0    0  123  187    0
##
##                      B10-06 B10-09a B10-09b B10-18 B10-24 B10-25a
## LSUzotu1:Chaetomium_globosum     11309  18613  35695  2094  1159  6733
## LSUzotu2:Raffaelea_sulphurea     13569  5195    7   3573  7565 13414
## LSUzotu3:Acremonium_biseptum        0    0    0    0    0    0
## LSUzotu4:Raffaelea_canadensis        0    1    0    2    0    2
## LSUzotu5:Raffaelea_canadensis     161    7   2030  540   885  861
## LSUzotu8:Penicillium_commune        0    0    0    0    0    0
## LSUzotu19:Chaetomium_globosum     192   491    0   48   24  452
## LSUzotu23:Raffaelea_sulphurea     312   673    0   39   35  255
## LSUzotu25:Raffaelea_sulphurea     179   213    0   43   19  305
## LSUzotu7:Raffaelea_sulphurea     125   115    0   40   64  116
## LSUzotu26:Chaetomium_spc          165   184    0   48   25  287
## LSUzotu9:Raffaelea_spc              4    0    0    1    1   14
## LSUzotu36:Raffaelea_sulphurea     139   150    0   41   20  267
## LSUzotu11:Raffaelea_canadensis      11    0    0   38  102  132
## LSUzotu37:Raffaelea_canadensis        0    0    1    0    0    0
##
##                      B10-25b B10-26a B15-01a B15-01b B15-06 B15-07a
## LSUzotu1:Chaetomium_globosum       95  4421    33  20163    4  15782
```

## LSUzotu2:Raffaelea_sulphurea	7353	18965	9796	2454	1589	2382
## LSUzotu3:Acremonium_biseptum	0	0	102	0	96	29
## LSUzotu4:Raffaelea_canadensis	0	1	1260	257	524	178
## LSUzotu5:Raffaelea_canadensis	3875	413	0	0	0	0
## LSUzotu8:Penicillium_commune	0	0	0	0	0	0
## LSUzotu19:Chaetomium_globosum	1	354	0	355	0	278
## LSUzotu23:Raffaelea_sulphurea	6	247	0	202	0	240
## LSUzotu25:Raffaelea_sulphurea	3	413	0	169	0	97
## LSUzotu7:Raffaelea_sulphurea	58	93	63	47	20	47
## LSUzotu26:Chaetomium_spc	0	255	0	125	0	137
## LSUzotu9:Raffaelea_spc	0	5	187	34	14	7
## LSUzotu36:Raffaelea_sulphurea	2	269	0	114	0	80
## LSUzotu11:Raffaelea_canadensis	754	59	0	0	0	0
## LSUzotu37:Raffaelea_canadensis	0	0	197	6	52	6
##	B15-07b	B15-11	B15-20	B15-24	B15-25a	B15-25b
## LSUzotu1:Chaetomium_globosum	3284	4560	7674	3569	466	3571
## LSUzotu2:Raffaelea_sulphurea	4285	8639	3977	8446	5434	14262
## LSUzotu3:Acremonium_biseptum	0	0	0	0	0	0
## LSUzotu4:Raffaelea_canadensis	1572	4372	1616	7115	1663	3320
## LSUzotu5:Raffaelea_canadensis	0	0	0	0	0	0
## LSUzotu8:Penicillium_commune	0	0	1	0	0	0
## LSUzotu19:Chaetomium_globosum	50	79	71	62	12	292
## LSUzotu23:Raffaelea_sulphurea	23	95	137	67	1	146
## LSUzotu25:Raffaelea_sulphurea	27	60	73	39	5	280
## LSUzotu7:Raffaelea_sulphurea	49	91	79	115	50	73
## LSUzotu26:Chaetomium_spc	34	58	65	42	6	189
## LSUzotu9:Raffaelea_spc	105	223	58	345	76	1081
## LSUzotu36:Raffaelea_sulphurea	25	39	43	58	4	177
## LSUzotu11:Raffaelea_canadensis	0	0	0	0	0	0
## LSUzotu37:Raffaelea_canadensis	123	144	24	257	124	412
##	B16-01a	B16-04	B16-31a	B16-31b	B20-04a	B20-08a
## LSUzotu1:Chaetomium_globosum	20350	33	10485	10239	34162	21
## LSUzotu2:Raffaelea_sulphurea	3	7443	793	33	30	49
## LSUzotu3:Acremonium_biseptum	0	0	0	0	0	26272
## LSUzotu4:Raffaelea_canadensis	1	1	29	4	1	3
## LSUzotu5:Raffaelea_canadensis	1	10	113	376	52	573
## LSUzotu8:Penicillium_commune	0	0	1520	6660	0	1
## LSUzotu19:Chaetomium_globosum	0	0	2	0	0	0
## LSUzotu23:Raffaelea_sulphurea	0	0	0	0	0	0
## LSUzotu25:Raffaelea_sulphurea	0	0	0	0	0	5
## LSUzotu7:Raffaelea_sulphurea	0	34	7	0	0	1
## LSUzotu26:Chaetomium_spc	0	0	3	0	0	0
## LSUzotu9:Raffaelea_spc	0	0	2	0	0	0
## LSUzotu36:Raffaelea_sulphurea	0	0	1	0	0	3
## LSUzotu11:Raffaelea_canadensis	0	0	1	2	3	0
## LSUzotu37:Raffaelea_canadensis	0	0	2	0	0	1
##	B20-08b	B20-11	B23-21	B23-23a	B24-02a	B24-26a
## LSUzotu1:Chaetomium_globosum	3445	6388	22924	8567	1031	11771
## LSUzotu2:Raffaelea_sulphurea	7	3	4734	14764	43	2542
## LSUzotu3:Acremonium_biseptum	5057	0	0	1	1	0
## LSUzotu4:Raffaelea_canadensis	0	2	2	1	1	0
## LSUzotu5:Raffaelea_canadensis	5452	1957	19	297	0	280
## LSUzotu8:Penicillium_commune	0	0	0	0	0	0
## LSUzotu19:Chaetomium_globosum	0	0	22	530	0	218

## LSUzotu23:Raffaelea_sulphurea	0	0	48	253	1	149
## LSUzotu25:Raffaelea_sulphurea	0	0	23	491	0	160
## LSUzotu7:Raffaelea_sulphurea	0	1	46	67	0	47
## LSUzotu26:Chaetomium_spc	0	0	11	291	1	117
## LSUzotu9:Raffaelea_spc	0	0	0	2	1	0
## LSUzotu36:Raffaelea_sulphurea	0	0	16	308	0	67
## LSUzotu11:Raffaelea_canadensis	1	0	0	54	0	21
## LSUzotu37:Raffaelea_canadensis	0	0	0	0	0	0
##	B24-26b	B24-28a	B24-29	B36-14a	B36-14b	B36-30
## LSUzotu1:Chaetomium_globosum	5163	6763	4311	3226	6297	6760
## LSUzotu2:Raffaelea_sulphurea	3448	8373	2555	4297	15265	1367
## LSUzotu3:Acremonium_biseptum	0	8	0	0	0	0
## LSUzotu4:Raffaelea_canadensis	0	29	1	3	0	3
## LSUzotu5:Raffaelea_canadensis	323	1842	62	407	743	23
## LSUzotu8:Penicillium_commune	0	0	0	0	0	0
## LSUzotu19:Chaetomium_globosum	240	300	10	42	379	7
## LSUzotu23:Raffaelea_sulphurea	140	163	22	19	140	6
## LSUzotu25:Raffaelea_sulphurea	160	201	11	48	348	5
## LSUzotu7:Raffaelea_sulphurea	33	87	21	50	109	24
## LSUzotu26:Chaetomium_spc	117	144	17	35	198	4
## LSUzotu9:Raffaelea_spc	5	17	0	0	8	0
## LSUzotu36:Raffaelea_sulphurea	92	126	4	21	205	1
## LSUzotu11:Raffaelea_canadensis	56	232	0	53	145	0
## LSUzotu37:Raffaelea_canadensis	0	0	0	0	0	0
##	B36-31a	B36-32	B39-15	B39-16	B39-17a	B46-03
## LSUzotu1:Chaetomium_globosum	16058	367	5922	2153	11890	20657
## LSUzotu2:Raffaelea_sulphurea	499	699	6225	7681	130	15
## LSUzotu3:Acremonium_biseptum	0	0	0	0	1	0
## LSUzotu4:Raffaelea_canadensis	3	0	2	0	5	1
## LSUzotu5:Raffaelea_canadensis	19	19	98	1914	32	385
## LSUzotu8:Penicillium_commune	0	0	0	0	1	0
## LSUzotu19:Chaetomium_globosum	95	2	3	0	25	0
## LSUzotu23:Raffaelea_sulphurea	50	0	1	0	6	0
## LSUzotu25:Raffaelea_sulphurea	37	0	3	0	4	0
## LSUzotu7:Raffaelea_sulphurea	15	4	73	83	1	0
## LSUzotu26:Chaetomium_spc	40	0	9	0	9	0
## LSUzotu9:Raffaelea_spc	0	0	0	0	0	0
## LSUzotu36:Raffaelea_sulphurea	32	0	0	1	9	0
## LSUzotu11:Raffaelea_canadensis	0	1	0	10	0	0
## LSUzotu37:Raffaelea_canadensis	0	0	0	0	0	0
##	B7-07	B7-08a				
## LSUzotu1:Chaetomium_globosum	2230	4660				
## LSUzotu2:Raffaelea_sulphurea	3522	5				
## LSUzotu3:Acremonium_biseptum	0	2				
## LSUzotu4:Raffaelea_canadensis	0	14				
## LSUzotu5:Raffaelea_canadensis	0	0				
## LSUzotu8:Penicillium_commune	0	0				
## LSUzotu19:Chaetomium_globosum	11	0				
## LSUzotu23:Raffaelea_sulphurea	3	2				
## LSUzotu25:Raffaelea_sulphurea	5	0				
## LSUzotu7:Raffaelea_sulphurea	23	0				
## LSUzotu26:Chaetomium_spc	5	0				
## LSUzotu9:Raffaelea_spc	0	1				
## LSUzotu36:Raffaelea_sulphurea	5	0				



```
## LSUzotu11:Raffaelea_canadensis      0      0
## LSUzotu37:Raffaelea_canadensis      0      0
```

```
Fungi_Species.all. <- ps.noncontam_dataset.LSU.ordi %>%
  tax_glom(taxrank = "Species") %>%
  transform_sample_counts(function(x) {x/sum(x)} ) %>%
  psmelt() %>%
  filter(Abundance > 0.05) %>%
  arrange(Species)

Fungi_Species.all.$Treatment <- factor(Fungi_Species.all.$Treatment, levels = c("control", "2nd-foundat.

Fungi_Species_plot <-ggplot(Fungi_Species.all., aes(x = Sample, y = Abundance, fill = Species)) +
  geom_bar(stat = "identity", position="fill") +
  scale_fill_manual(values = Family_colors, name = "Species")

Fungi_Species_plot +
  facet_wrap("Treatment", scales = "free_x", drop = TRUE)+
  theme(plot.title = element_text(size = 16, face = "bold")) +
  theme(text = element_text(size=16, face = "bold"))+
  theme(legend.title = element_text(size = 20), legend.text = element_text(size = 14))+
  theme_classic()+      theme(panel.grid.major = element_line(colour = "grey50"))+
  labs(x="Sample", y="Relative Abundance")+
  scale_y_continuous(labels=percent_format())+
  theme(axis.text.x=element_text(angle = 90, hjust = 1, size = rel(1)))+
  theme(legend.text = element_text(face = "italic"))
```



```
##
## [[6]]
## [1] "7] Sparsity = 0.822267520869734"
##
## [[7]]
## [1] "6] Any OTU sum to 1 or less? YES"
##
## [[8]]
## [1] "8] Number of singletons = 4"
##
## [[9]]
## [1] "9] Percent of OTUs that are singletons \n          (i.e. exactly one read detected across all sampl
##
## [[10]]
## [1] "10] Number of sample variables are: 6"
##
## [[11]]
## [1] "Sample"          "Nest"          "Linage"          "Treatment"
## [5] "age_sampling_d." "Group"
```

```
replicates <- subset.data.frame(sample_data(fwc))
table(replicates$Treatment)
```

```
##
## 2nd-foundation      control      removal
##           11           20           20
```

```
sample_data(fwc)$ColSums <- colSums(otu_table(fwc))
```

## Analysis

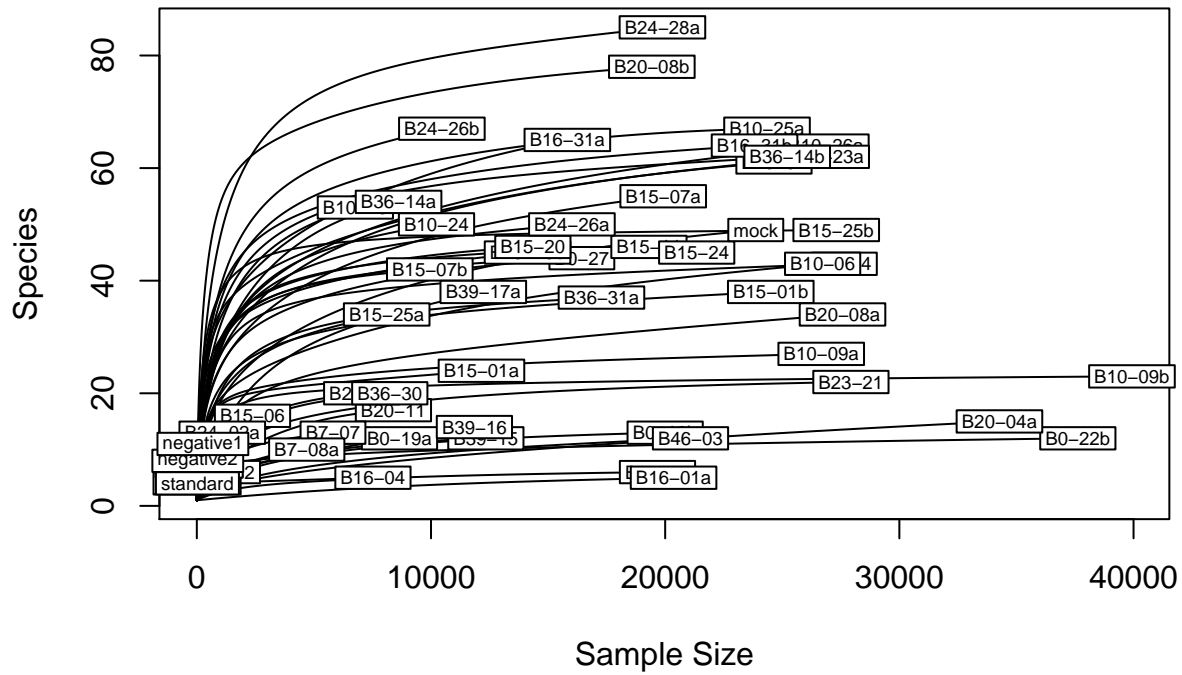
### Rarefaction curves

1. all samples with controls
2. all samples with controls after decontam
3. all samples without controls

Rarefaction is used to simulate even number of reads per sample.

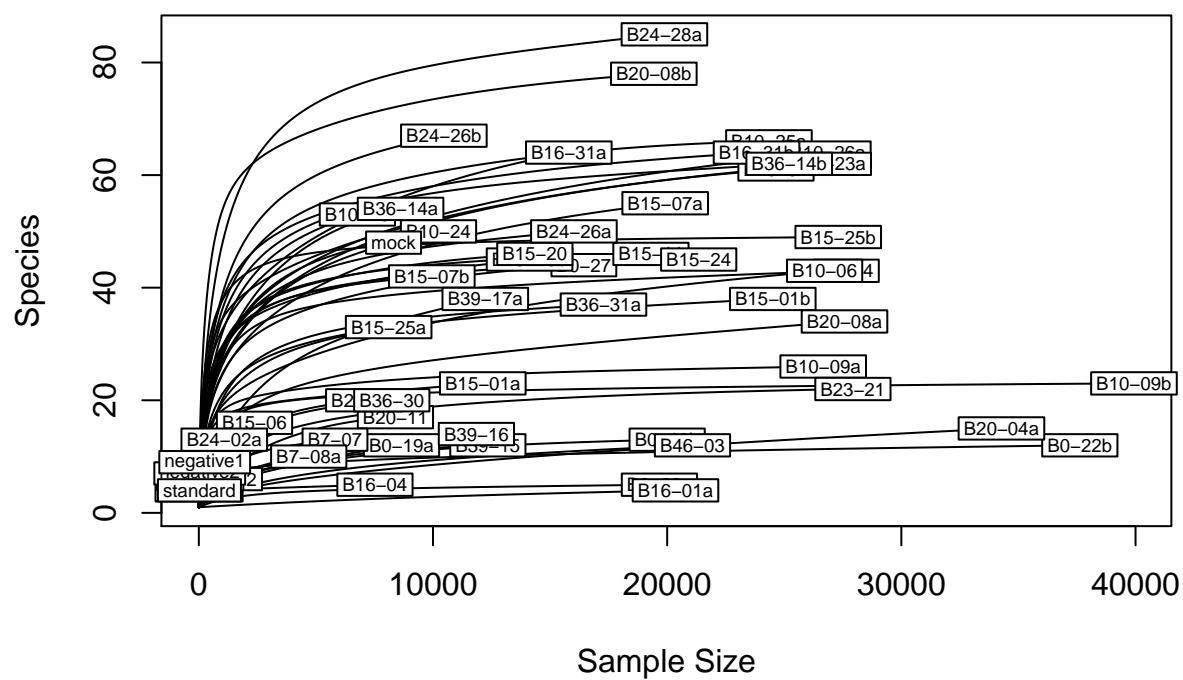
```
rarecurve(t(otu_table(dataset.LSU.ordi)), cex=0.6, step = 20, main = "Rarefaction curve of all samples")
```

### Rarefaction curve of all samples (incl. controls)



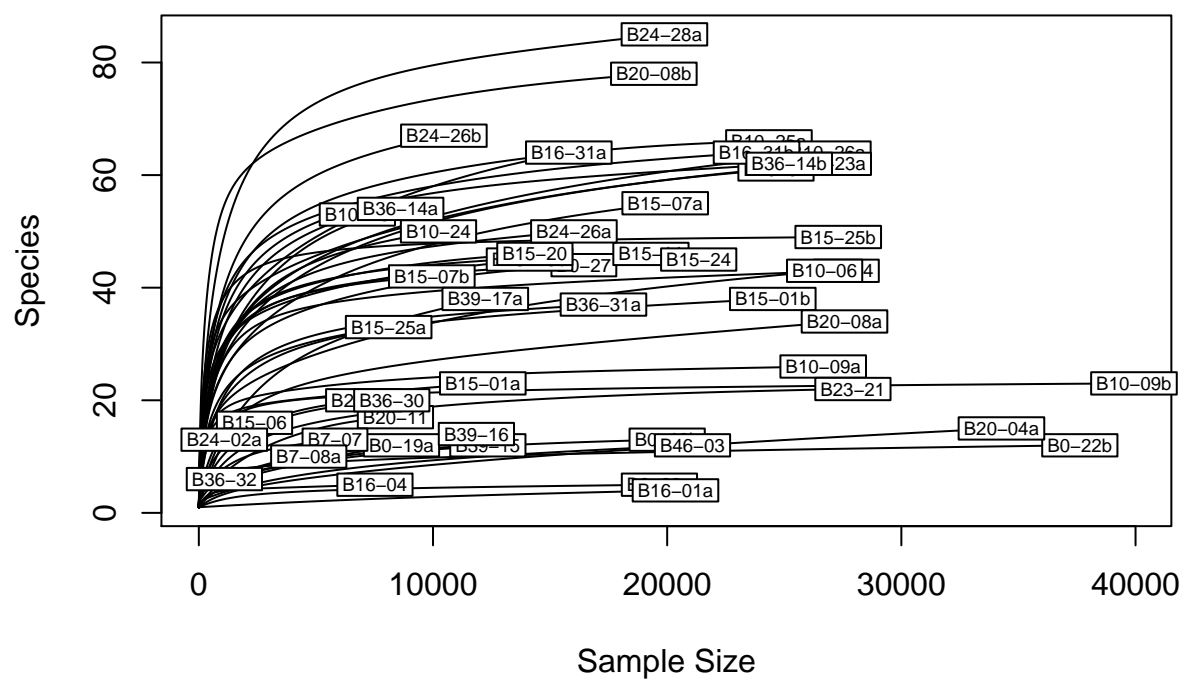
```
rarecurve(t(otu_table(ps.noncontam_dataset.LSU.ordi)), cex=0.6, step = 20, main = "Rarefaction curve of
```

## Rarefaction curve of all samples (incl. controls) after decontaminatic



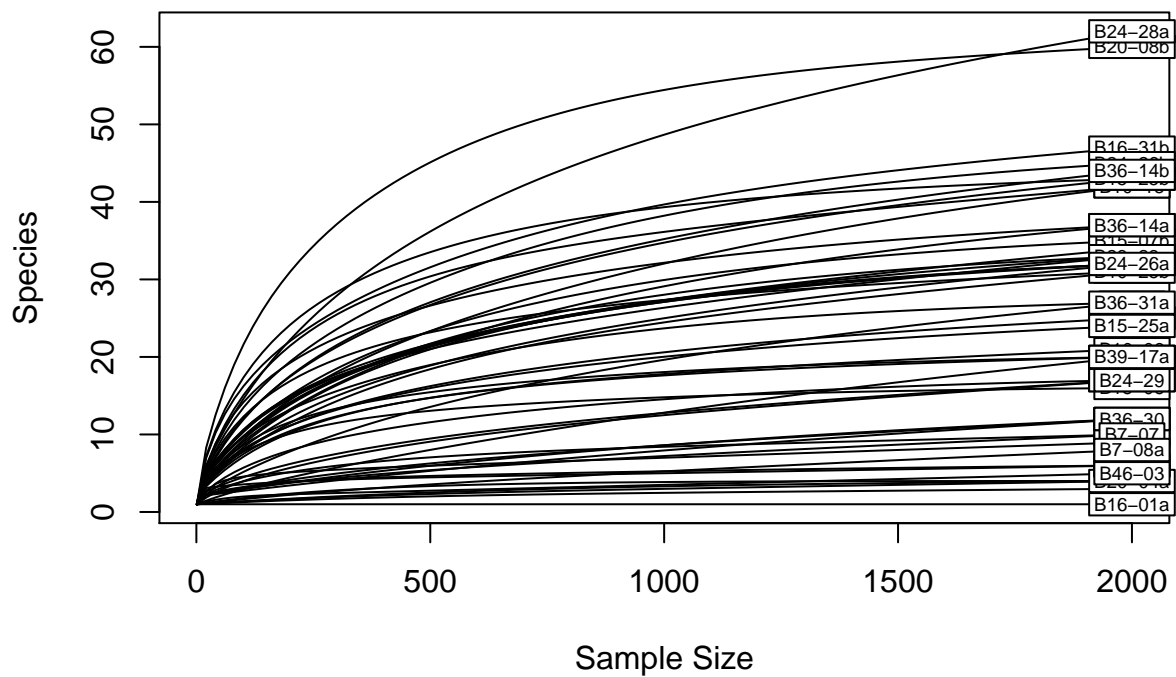
```
rarecurve(t(otu_table(fwc)), cex=0.6, step = 20, main = "Rarefaction curve of samples (excl. controls)");
```

## Rarefaction curve of samples (excl. controls)



rarefy to 2000 reads

```
ps2 <-fwc
set.seed(1)
ps2 <- rarefy_even_depth(ps2,sample.size=2000, replace=FALSE, rngseed = 1)
rarecurve(t(otu_table(ps2)), cex=0.6, step = 20)
```



```
replicates3 <- subset.data.frame(sample_data(ps2))
table(replicates3$Treatment)
```

```
##
## 2nd-foundation      control      removal
##           11           19           19
```

subset data to two comparison groups

```
#control vs. removal
CR <- subset_samples(fwc, Treatment!="2nd-foundation")
CRrf <- subset_samples(ps2, Treatment!="2nd-foundation")

#removal vs. 2nd attempt
R2nd <- subset_samples(fwc, Treatment!="control")
R2ndrf <- subset_samples(ps2, Treatment!="control")
```

## Alpha diversity

### Diversity plots

This returns a table with selected diversity indicators (Shannon diversity & Observed richness).

```
tabCR <- microbiome::alpha(CRrf, index = c("diversity_shannon","observed"))
tabR2nd <- microbiome::alpha(R2ndrf, index = c("diversity_shannon","observed"))
```

**Prepare data for vizualisation** Now, get the metadata (sample\_data) from the phyloseq object

```
ps1.meta.CR <- meta(CRrf)
ps1.meta.R2nd <- meta(R2ndrf)
```

Add the diversity table to metadata

```
ps1.meta.CR$Shannon <- tabCR$diversity_shannon
ps1.meta.CR$Observed <- tabCR$observed

ps1.meta.R2nd$Shannon <- tabR2nd$diversity_shannon
ps1.meta.R2nd$Observed <- tabR2nd$observed
```

**plot** control vs. removal (rarefied, 2000 reads)

```
ps1.meta.CR$Treatment <- as.factor(ps1.meta.CR$Treatment)
ps1.meta.CR$Treatment <- factor(ps1.meta.CR$Treatment, levels = c("control", "removal"))
```

*#Shannon diversity index*

```
(a <- summaryBy(Shannon ~ Treatment, ps1.meta.CR, FUN = c(mean, sd)))
```

```
## Treatment Shannon.mean Shannon.sd
## 1 control 0.9571066 0.4540992
## 2 removal 0.7803962 0.5793697
```

```
p1 <- ggplot(ps1.meta.CR, aes(x = Treatment, y = Shannon))
shan_fun <- p1 + geom_jitter(position = position_jitter(w=0.2, h=0.1))+
  geom_pointrange(data = a, mapping = aes(x = Treatment, y = Shannon.mean, ymin = Shannon.mean-Shannon.sd, ymax = Shannon.mean+Shannon.sd),
  theme_bw()+
  ylim(NA, 1.95)+
  labs(x = "Treatment group", y = "Shannon Diversity Index")+
  theme(axis.title = element_text(size = 18, face = "bold"), axis.text = element_text(size = 16))
```

*#Observed richness*

```
(b <- summaryBy(Observed ~ Treatment, ps1.meta.CR, FUN = c(mean, sd)))
```

```
## Treatment Observed.mean Observed.sd
## 1 control 20.68421 12.79346
## 2 removal 24.36842 15.64612
```

```
p2 <- ggplot(ps1.meta.CR, aes(x = Treatment, y = Observed))
obs_fun <- p2 + geom_jitter(position = position_jitter(w=0.2, h=0.1))+
  geom_pointrange(data = b, mapping = aes(x = Treatment, y = Observed.mean, ymin = Observed.mean-Observed.sd, ymax = Observed.mean+Observed.sd),
  theme_bw()+
```



```

ylim(NA, 70)
labs(x = "Treatment group", y = "Observed richness")+
theme(axis.title = element_text(size = 18, face = "bold"), axis.text = element_text(size = 16))

## NULL

removal vs. 2nd attempt (rarefied, 2000 reads)

ps1.meta.R2nd$Treatment <- as.factor(ps1.meta.R2nd$Treatment)
ps1.meta.R2nd$Treatment <- factor(ps1.meta.R2nd$Treatment, levels = c("removal", "2nd-foundation"), lab

#Shannon diversity index
(c <- summaryBy(Shannon ~ Treatment, ps1.meta.R2nd, FUN = c(mean, sd)))

##      Treatment Shannon.mean Shannon.sd
## 1      removal    0.7803962  0.5793697
## 2 2nd-attempt    1.2684611  0.7431823

p3 <- ggplot(ps1.meta.R2nd, aes(x = Treatment, y = Shannon))
shan_fun2 <- p3 + geom_jitter(position = position_jitter(w=0.2, h=0.1))+
  geom_pointrange(data = c, mapping = aes(x = Treatment, y = Shannon.mean, ymin = Shannon.mean-Shannon
  theme_bw()+
  ylim(NA, 2.7))+
  labs(x = "Treatment group", y = "Shannon Diversity Index")+
  theme(axis.title = element_text(size = 18, face = "bold"), axis.text = element_text(size = 16))

#Observed richness
(d <- summaryBy(Observed ~ Treatment, ps1.meta.R2nd, FUN = c(mean, sd)))

##      Treatment Observed.mean Observed.sd
## 1      removal    24.36842    15.64612
## 2 2nd-attempt    32.81818    17.28478

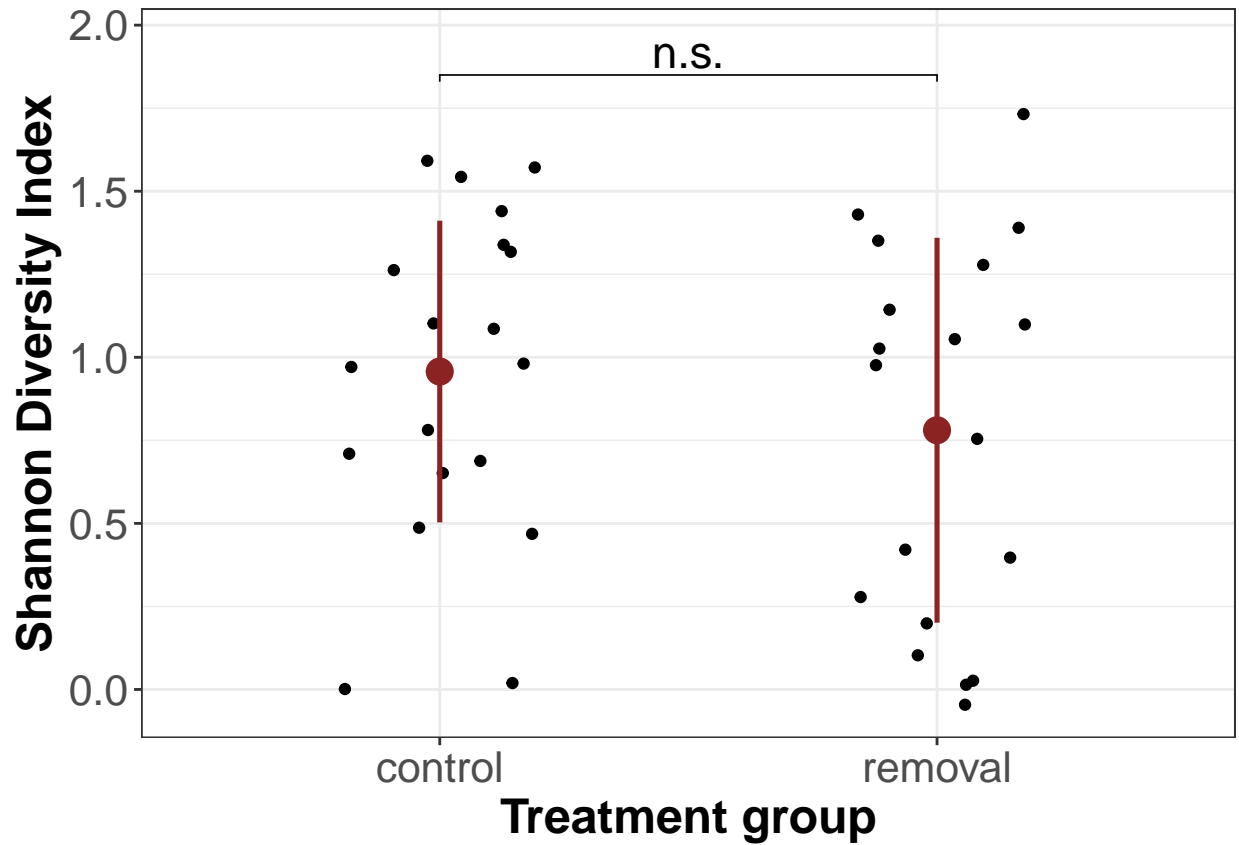
p4 <- ggplot(ps1.meta.R2nd, aes(x = Treatment, y = Observed))
obs_fun2 <- p4 + geom_jitter(position = position_jitter(w=0.2, h=0.1))+
  geom_pointrange(data = d, mapping = aes(x = Treatment, y = Observed.mean, ymin = Observed.mean-Obse
  theme_bw()+
  ylim(NA, 70))+
  labs(x = "Treatment group", y = "Observed richness")+
  theme(axis.title = element_text(size = 18, face = "bold"), axis.text = element_text(size = 16))

add statistical data output to plot

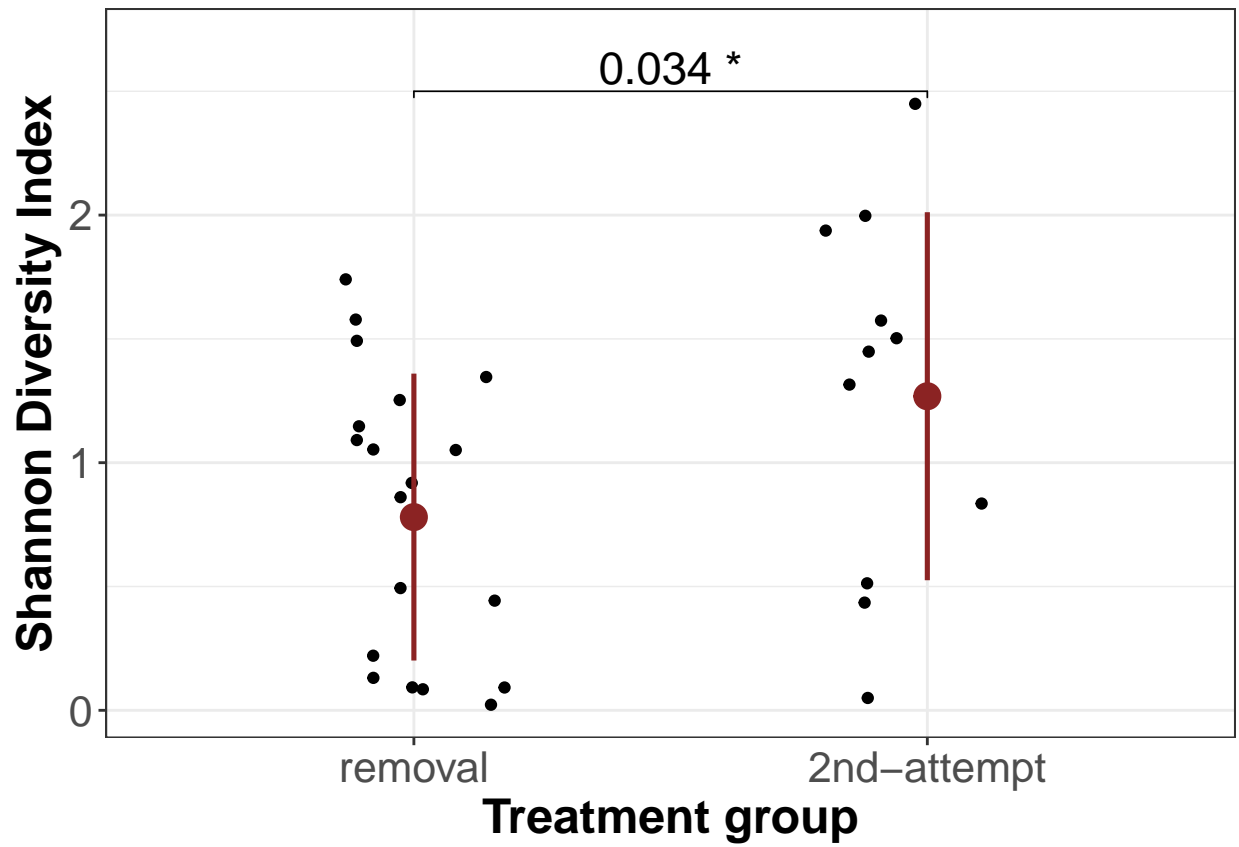
shan_fun <- shan_fun + font("axis.title", size = 18)+
  font("xylab", size = 18)+
  geom_bracket(
    xmin = "control", xmax = "removal", y.position = 1.85, label = "n.s.", label.size = 6, tip.l

shan_fun

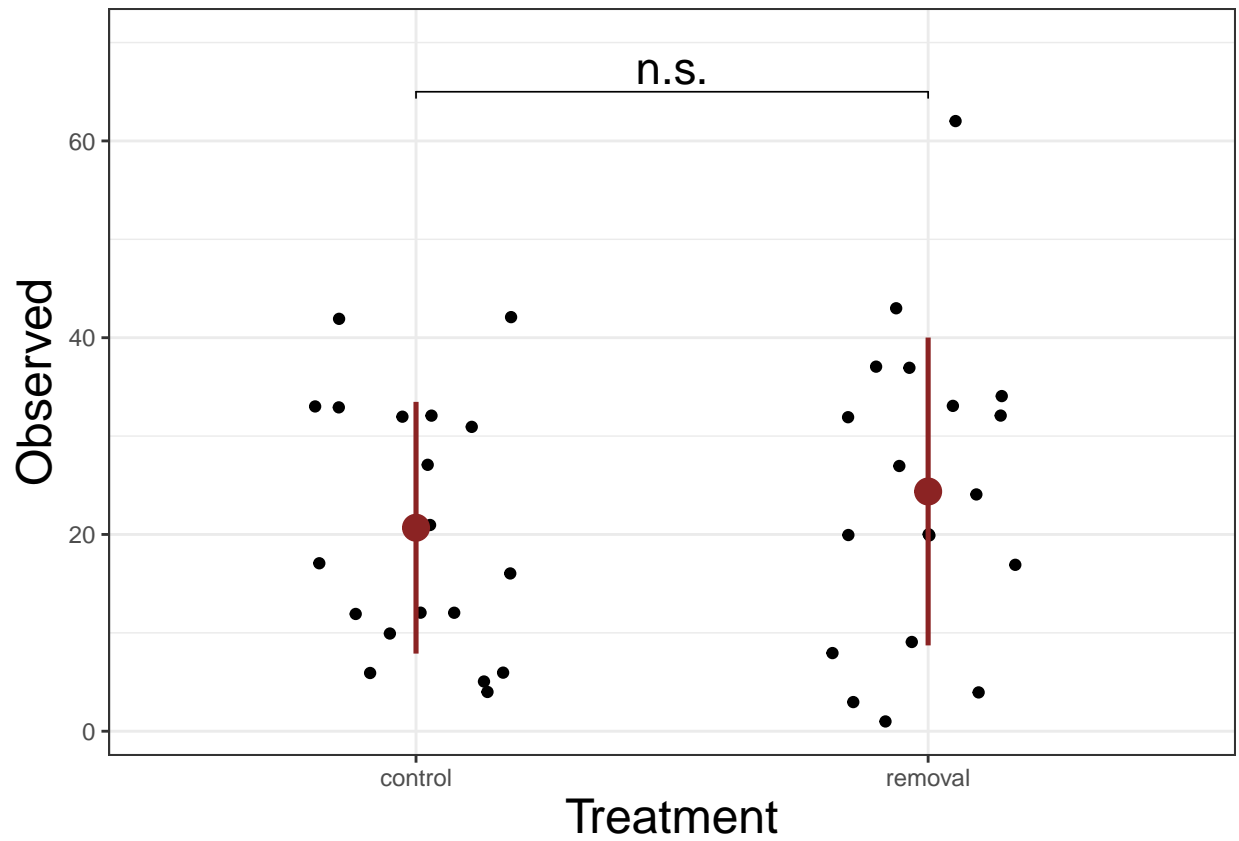
```



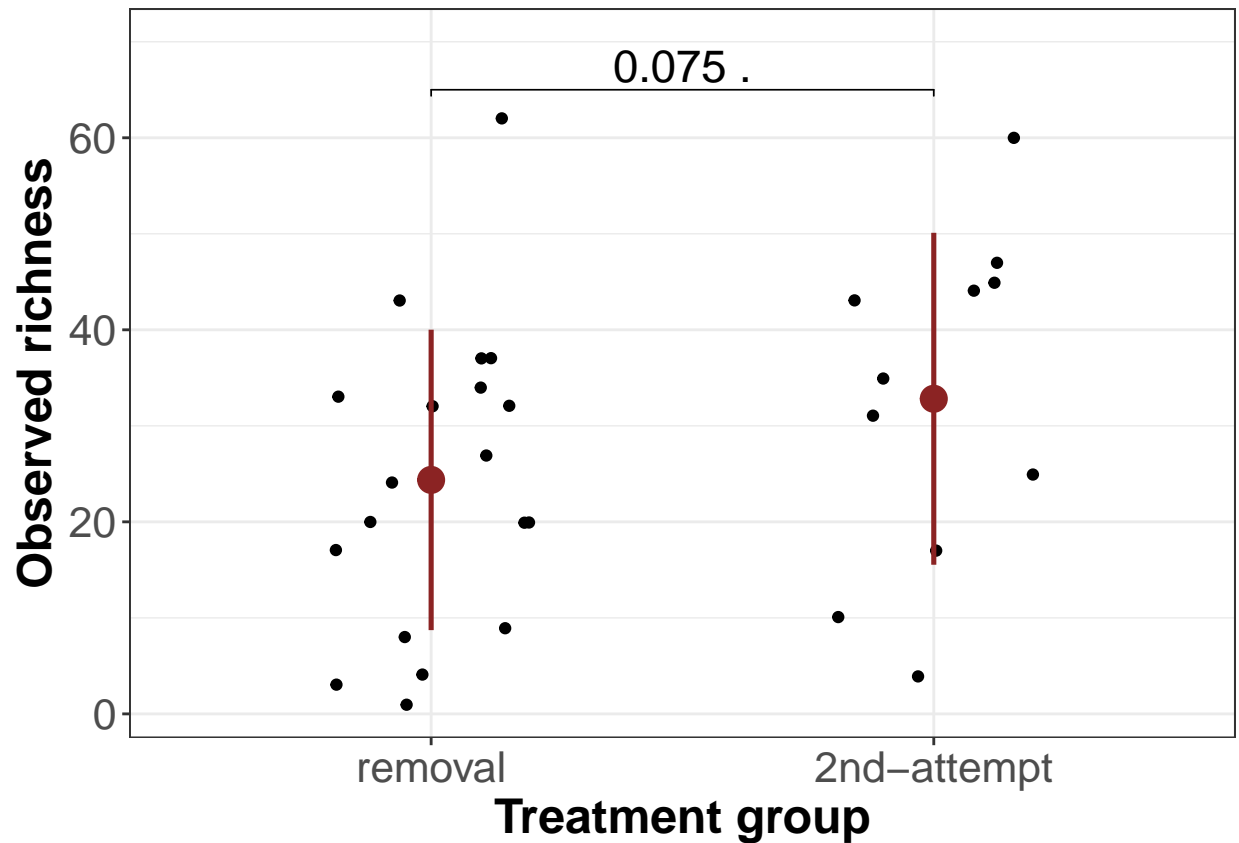
```
shan_fun2 <- shan_fun2 + font("axis.title", size = 18)+
  font("xylab", size = 18)+
  geom_bracket(
    xmin = "removal", xmax = "2nd-attempt", y.position = 2.5, label = "0.034 *", label.size = 6
  )
shan_fun2
```



```
obs_fun <- obs_fun + font("axis.title", size = 18)+
  font("xylab", size = 18)+
  geom_bracket(
    xmin = "control", xmax = "removal", y.position = 65, label = "n.s.", label.size = 6,tip.len=
obs_fun
```



```
obs_fun2 <- obs_fun2 + font("axis.title", size = 18)+
  font("xylab", size = 18)+
  geom_bracket(
    xmin = "removal", xmax = "2nd-attempt", y.position = 65, label = "0.075 .", label.size = 6,
  )
obs_fun2
```



## Statistics

Testing differences in alpha diversity

create dataframe for analysis

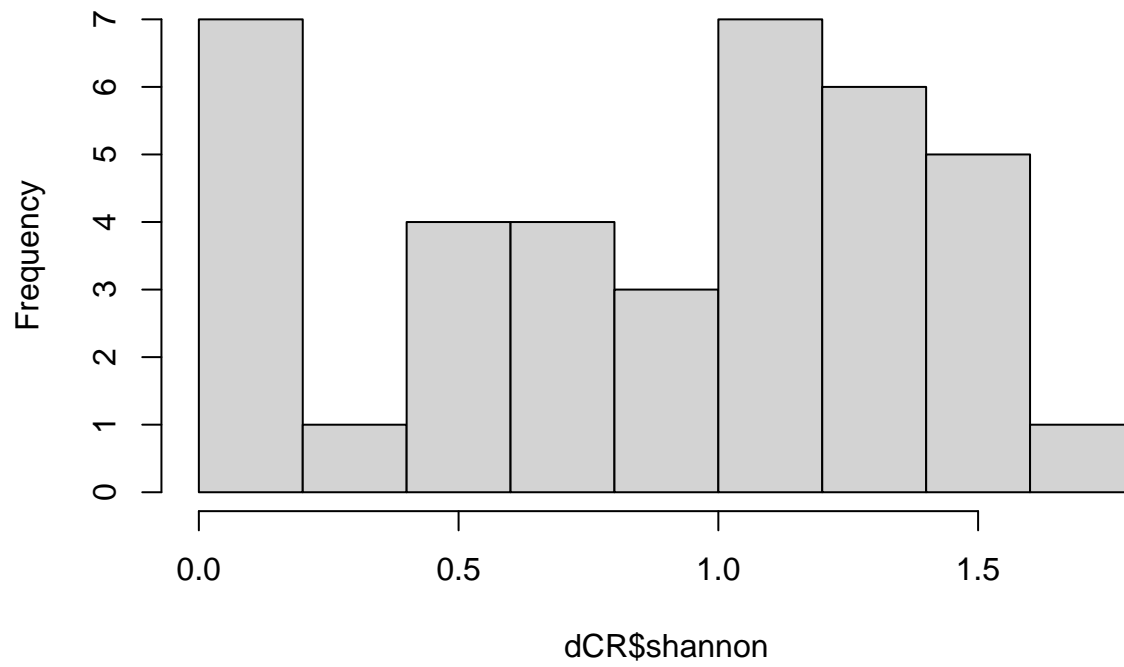
```
#control vs. removal
dCR <- meta(CRrf)
shannon <- diversity(CRrf, "shannon")
dCR$shannon <- shannon$shannon
observed <- alpha(CRrf, index = "observed", zeroes = TRUE)
dCR$observed <- observed$observed

#removal vs. 2nd attempt
dR2nd <- meta(R2ndrf)
shannon <- diversity(R2ndrf, "shannon")
dR2nd$shannon <- shannon$shannon
observed <- alpha(R2ndrf, index = "observed", zeroes = TRUE)
dR2nd$observed <- observed$observed
```

Shannon diversity test distribution

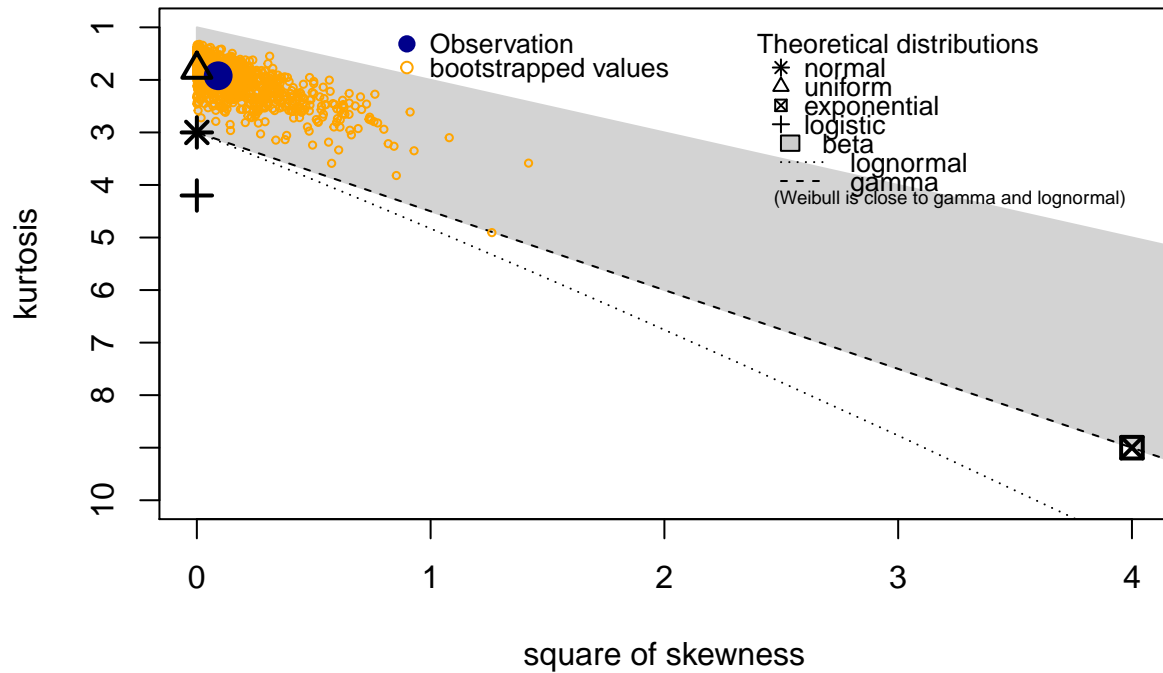
```
hist(dCR$shannon)
```

**Histogram of dCR\$shannon**



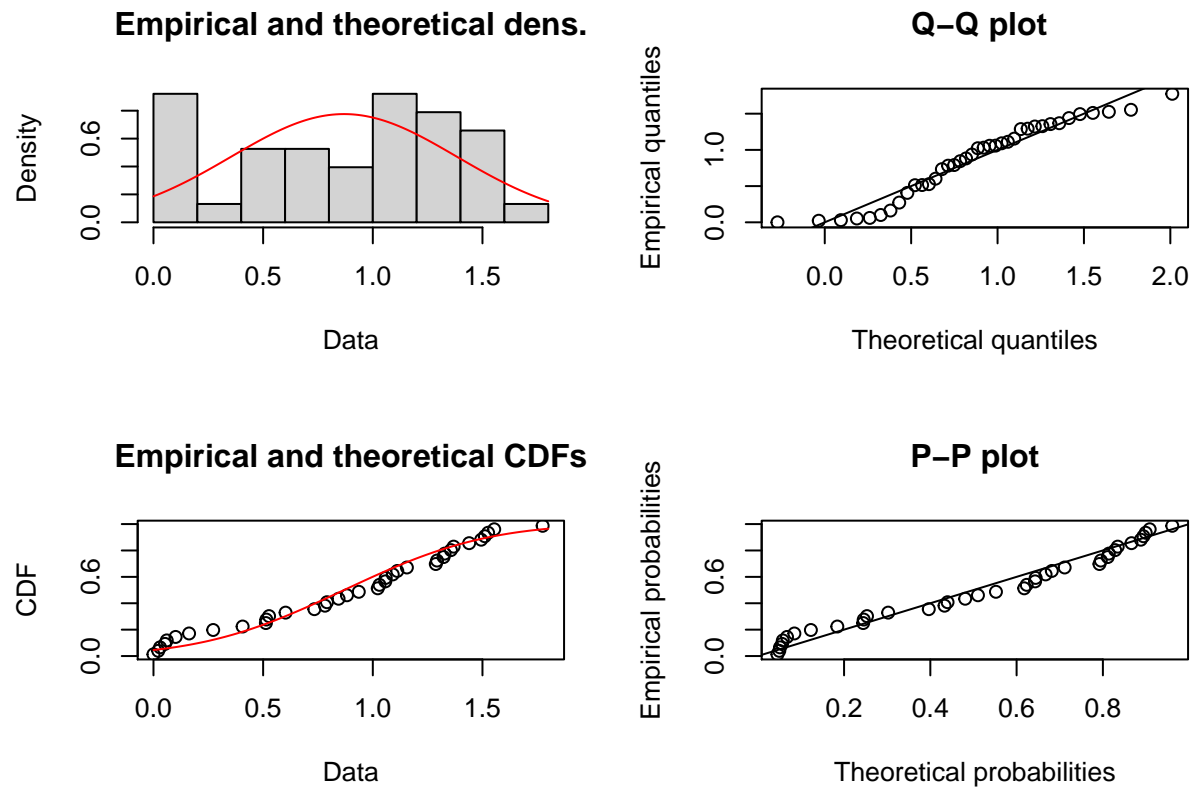
```
descdist(dCR$shannon, boot = 1000)
```

## Cullen and Frey graph



```
## summary statistics
## -----
## min: 0    max: 1.774394
## median: 0.9793887
## mean: 0.8687514
## estimated sd: 0.5211839
## estimated skewness: -0.3021322
## estimated kurtosis: 1.921262
```

```
fit.norm <- fitdist(dCR$shannon, distr = "norm", method = "mme") # looks good
plot(fit.norm)
```



→ gamma

test with subset 1

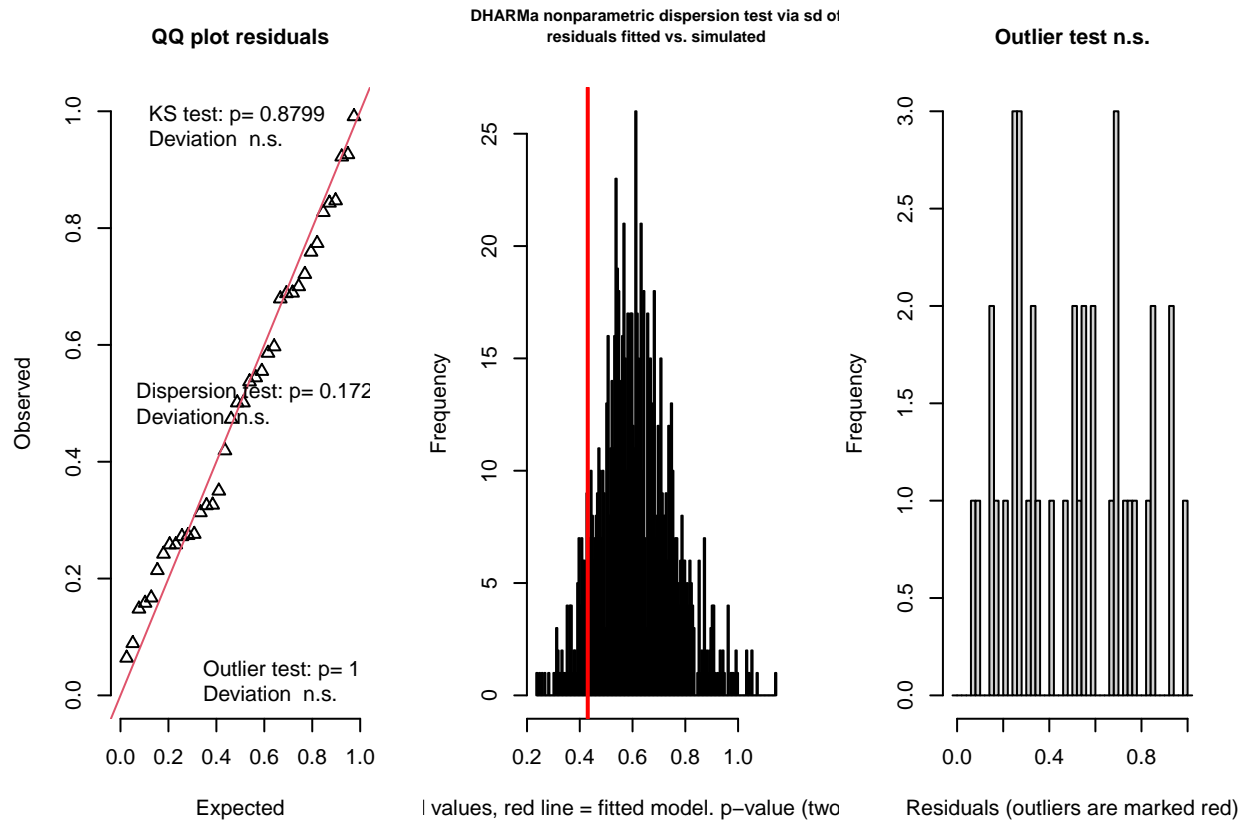
```
dCR$Treatment <- as.factor(dCR$Treatment)
dCR <- within(dCR, Treatment <- relevel(Treatment, ref = "control"))

shan1 <- glm(shannon ~ Treatment + Linage, data = dCR)
Anova(shan1, type = "II")
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: shannon
##          LR Chisq Df Pr(>Chisq)
## Treatment  1.2779  1  0.25830
## Linage     22.5454 10  0.01255 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
res_shan1 <- simulateResiduals(shan1, n = 1000)
testResiduals(res_shan1)
```





```
## $uniformity
##
## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.095368, p-value = 0.8799
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.70198, p-value = 0.172
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 38, p-value = 1
```

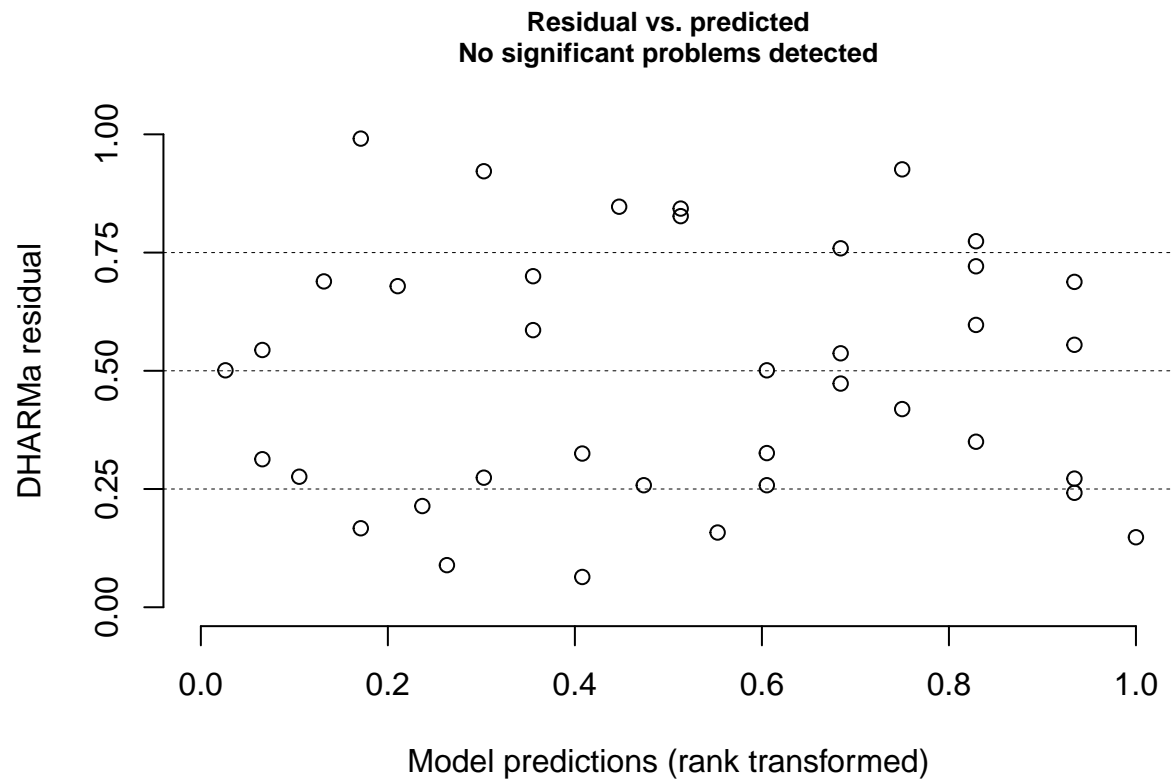
```

## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.00000000 0.09251276
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
## 0

## $uniformity
##
## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.095368, p-value = 0.8799
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.70198, p-value = 0.172
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 38, p-value = 1
## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.00000000 0.09251276
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
## 0

```

```
plotResiduals(res_shan1)
```

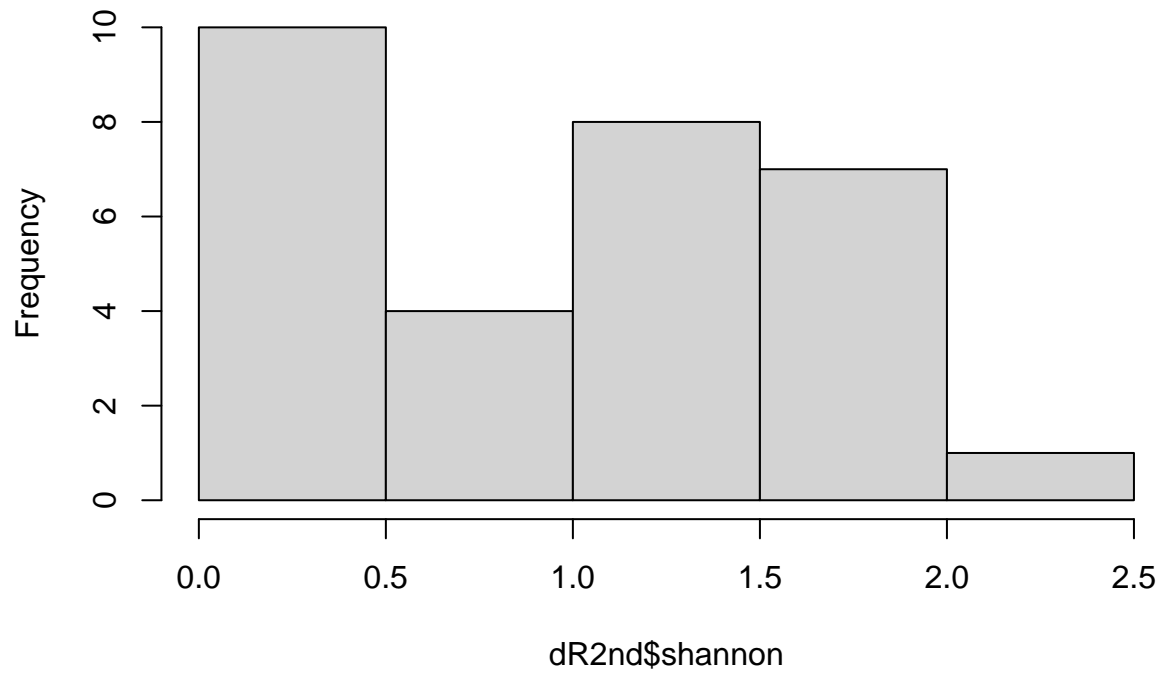


-> strong Lineage effect in B24, Lineage with high abundant *Erwinia*, *Ochrobactrum* and *Microbacterium* inside

test with subset 2

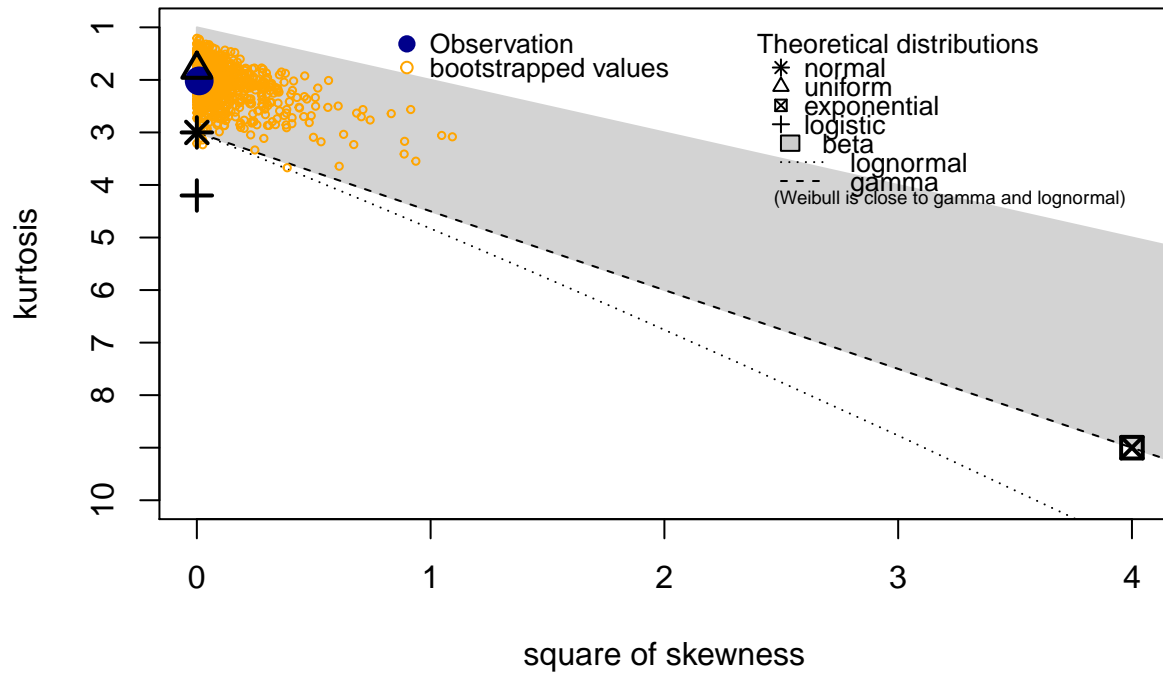
```
hist(dR2nd$shannon)
```

**Histogram of dR2nd\$shannon**



```
descdist(dR2nd$shannon, boot = 1000)
```

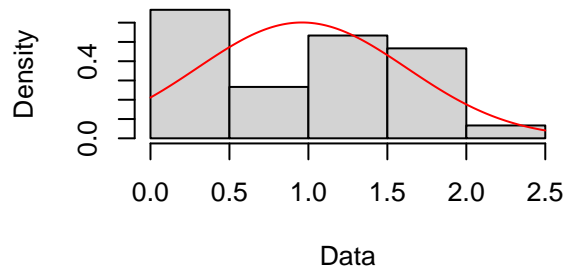
## Cullen and Frey graph



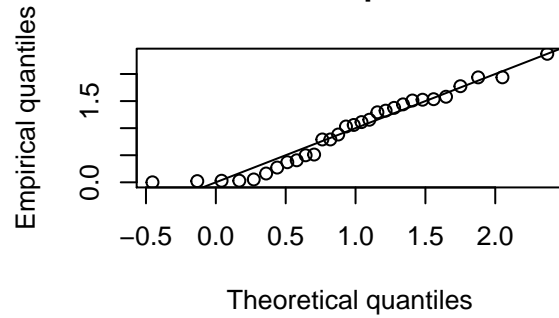
```
## summary statistics
## -----
## min: 0    max: 2.372804
## median: 1.044649
## mean: 0.9593534
## estimated sd: 0.675297
## estimated skewness: 0.1004296
## estimated kurtosis: 2.019042
```

```
fit.norm <- fitdist(dR2nd$shannon, distr = "norm", method = "mme") # looks good
plot(fit.norm)
```

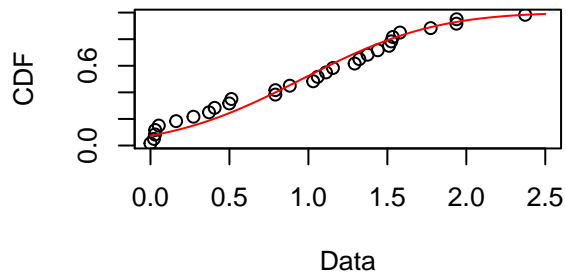
**Empirical and theoretical dens.**



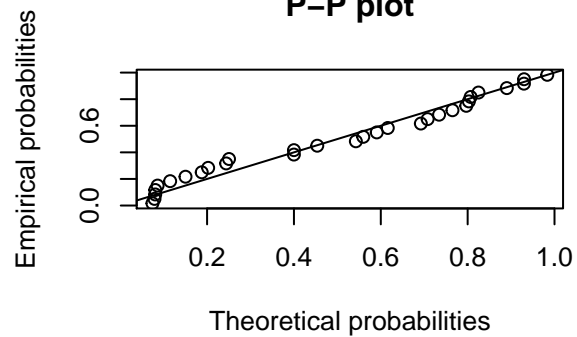
**Q-Q plot**



**Empirical and theoretical CDFs**



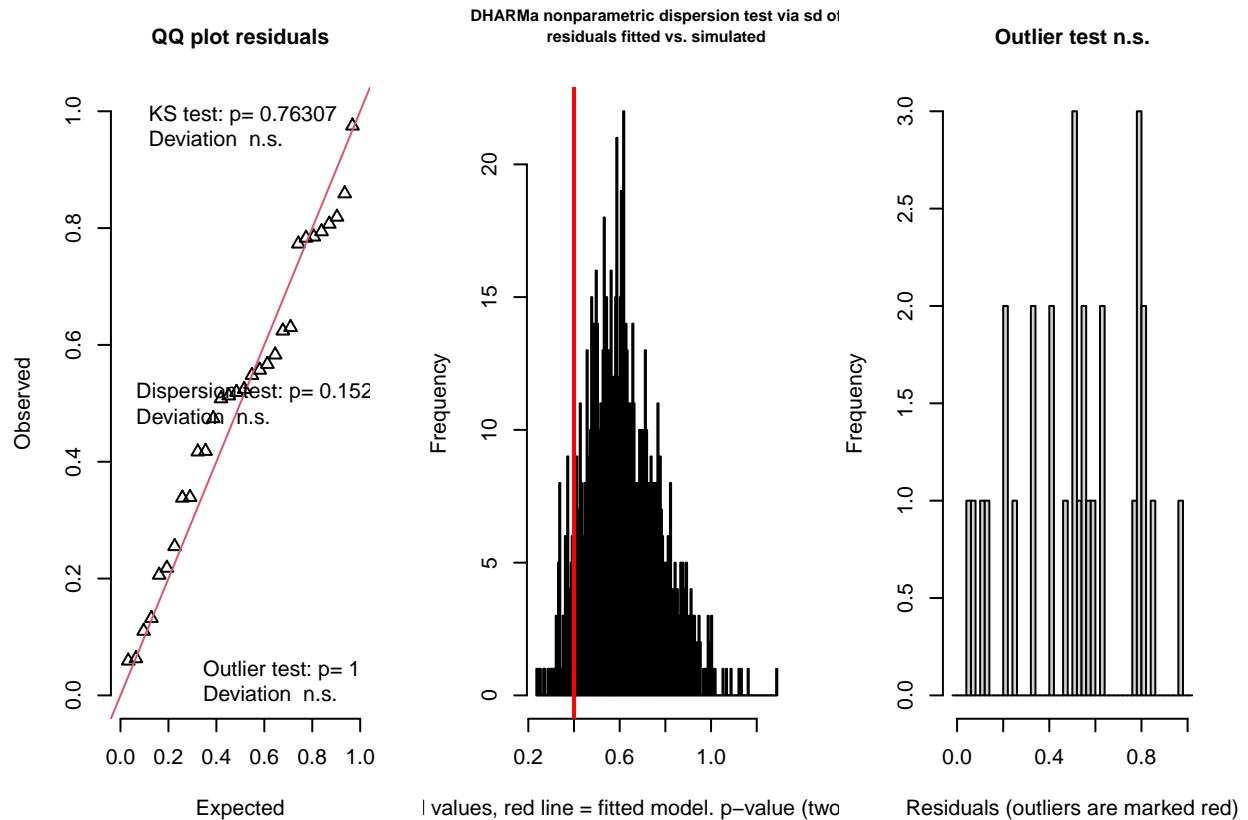
**P-P plot**



```
shan <- glm(shannon ~ Treatment + Linage, data = dR2nd)
Anova(shan, type = "II")
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: shannon
##      LR Chisq Df Pr(>Chisq)
## Treatment  4.5059 1  0.03378 *
## Linage    14.1368 9  0.11754
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
res_shan <- simulateResiduals(shan, n = 1000)
testResiduals(res_shan)
```



```
## $uniformity
##
## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.117, p-value = 0.7631
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.65369, p-value = 0.152
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 30, p-value = 1
```

```

## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.0000000 0.1157033
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
## 0

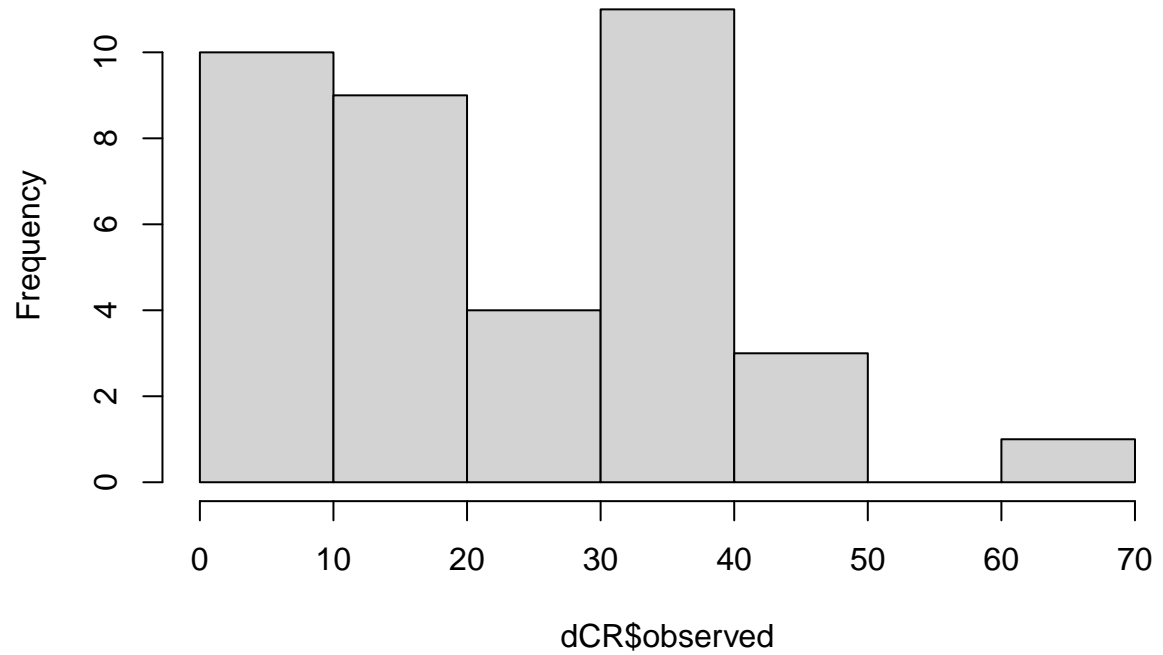
## $uniformity
##
## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.117, p-value = 0.7631
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.65369, p-value = 0.152
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 30, p-value = 1
## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.0000000 0.1157033
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
## 0

```

```
hist(dCR$observed)
```



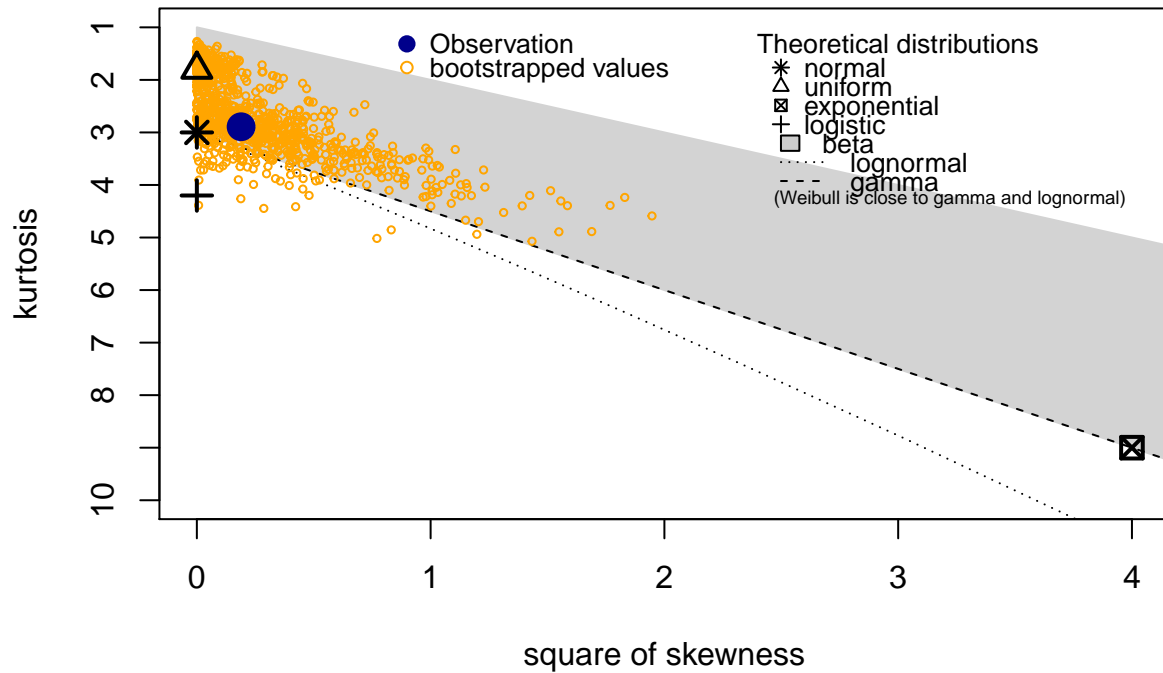
**Histogram of dCR\$observed**



Observed richness

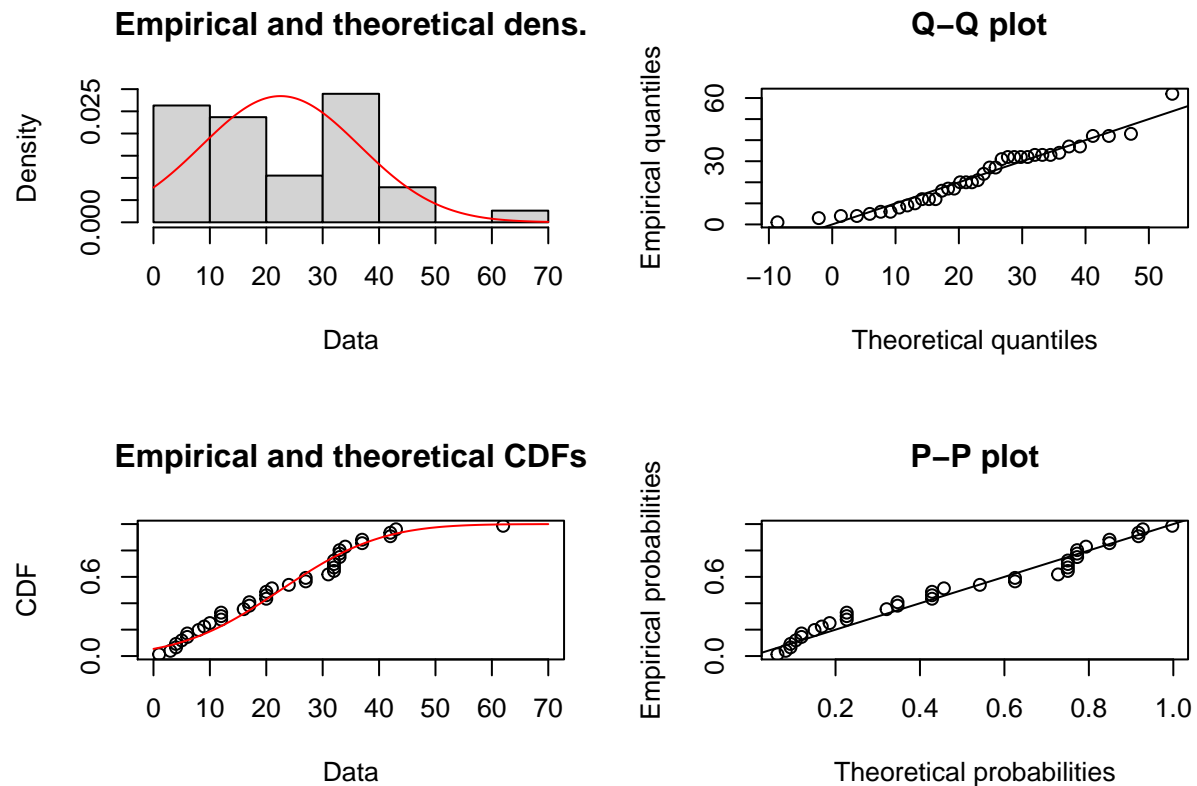
```
descdist(dCR$observed, boot = 1000)
```

## Cullen and Frey graph



```
## summary statistics
## -----
## min: 1   max: 62
## median: 20.5
## mean: 22.52632
## estimated sd: 14.21977
## estimated skewness: 0.4350977
## estimated kurtosis: 2.891807
```

```
fit.norm <- fitdist(dCR$observed, distr = "norm", method = "mme") # looks good
plot(fit.norm)
```



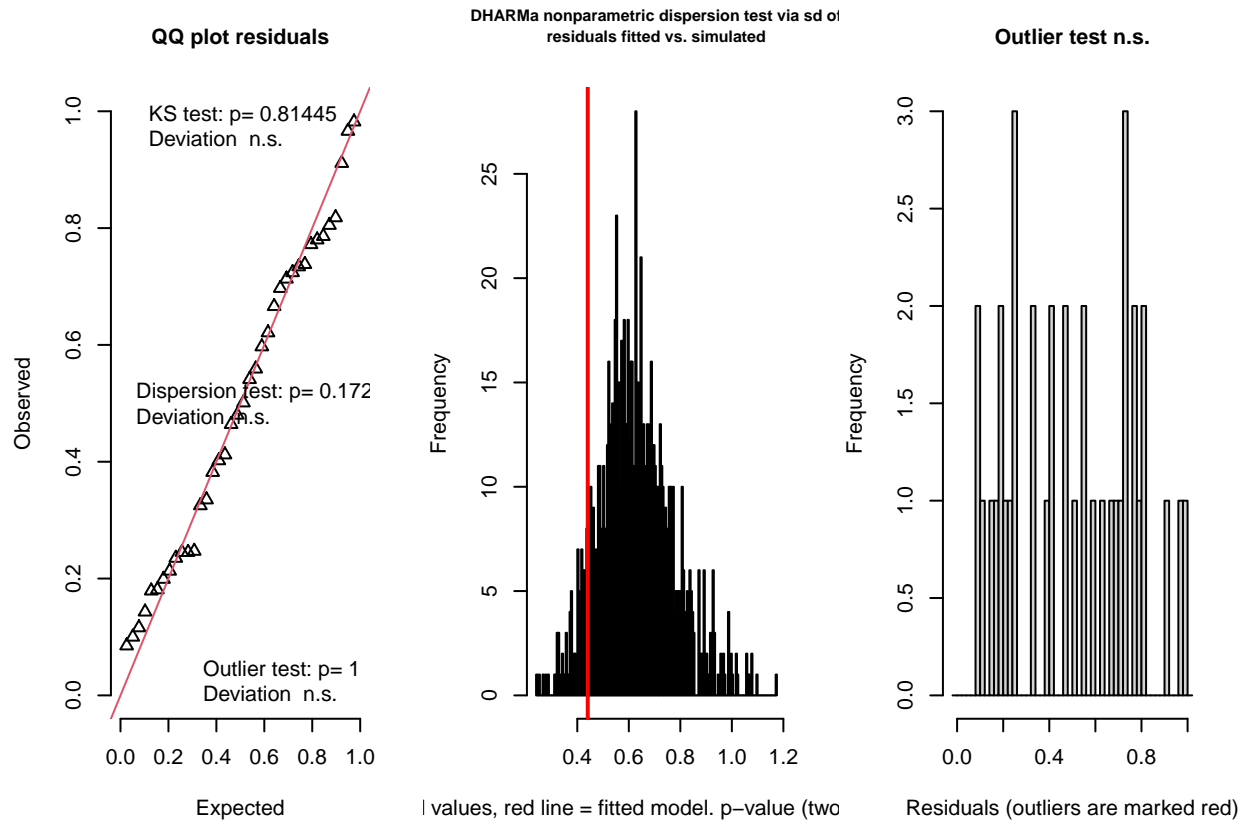
run model

```
dCR$Treatment <- as.factor(dCR$Treatment)
dCR <- within(dCR, Treatment <- relevel(Treatment, ref = "control"))

ob1 <- glm(observed ~ Treatment + Linage, data = dCR)
Anova(ob1, type = "II")
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: observed
##          LR Chisq Df Pr(>Chisq)
## Treatment   0.6393  1   0.42395
## Linage    21.9446 10   0.01539 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
res_ob1 <- simulateResiduals(ob1, n = 1000)
testResiduals(res_ob1)
```



```
## $uniformity
##
## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.10305, p-value = 0.8144
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.70198, p-value = 0.172
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 38, p-value = 1
```

```
## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.00000000 0.09251276
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
## 0
```

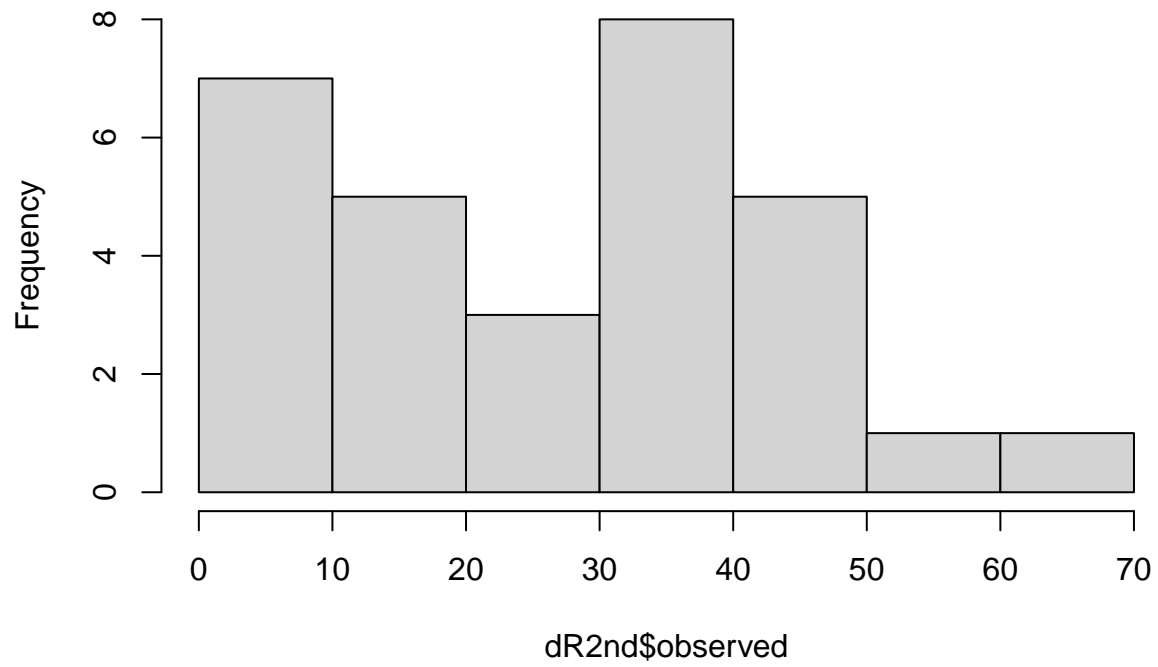
```
## $uniformity
##
## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.10305, p-value = 0.8144
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.70198, p-value = 0.172
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 38, p-value = 1
## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.00000000 0.09251276
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
## 0
```

choose normal distribution

test with subset 2

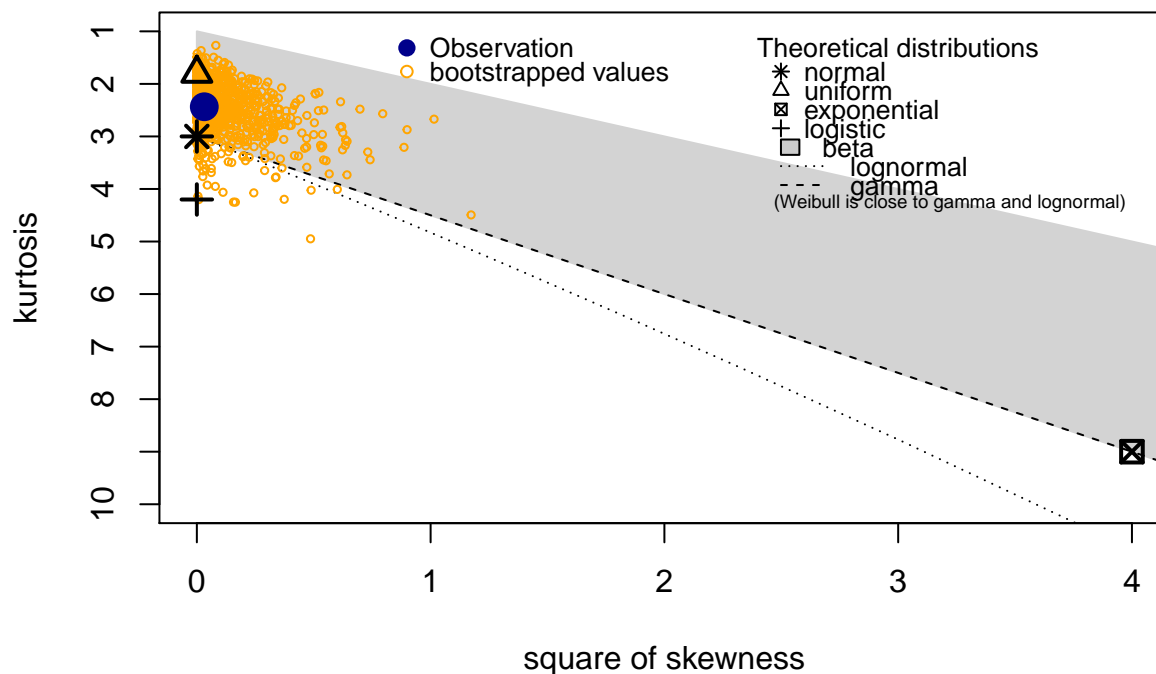
```
hist(dR2nd$observed)
```

**Histogram of dR2nd\$observed**



```
descdist(dR2nd$observed, boot = 1000)
```

## Cullen and Frey graph

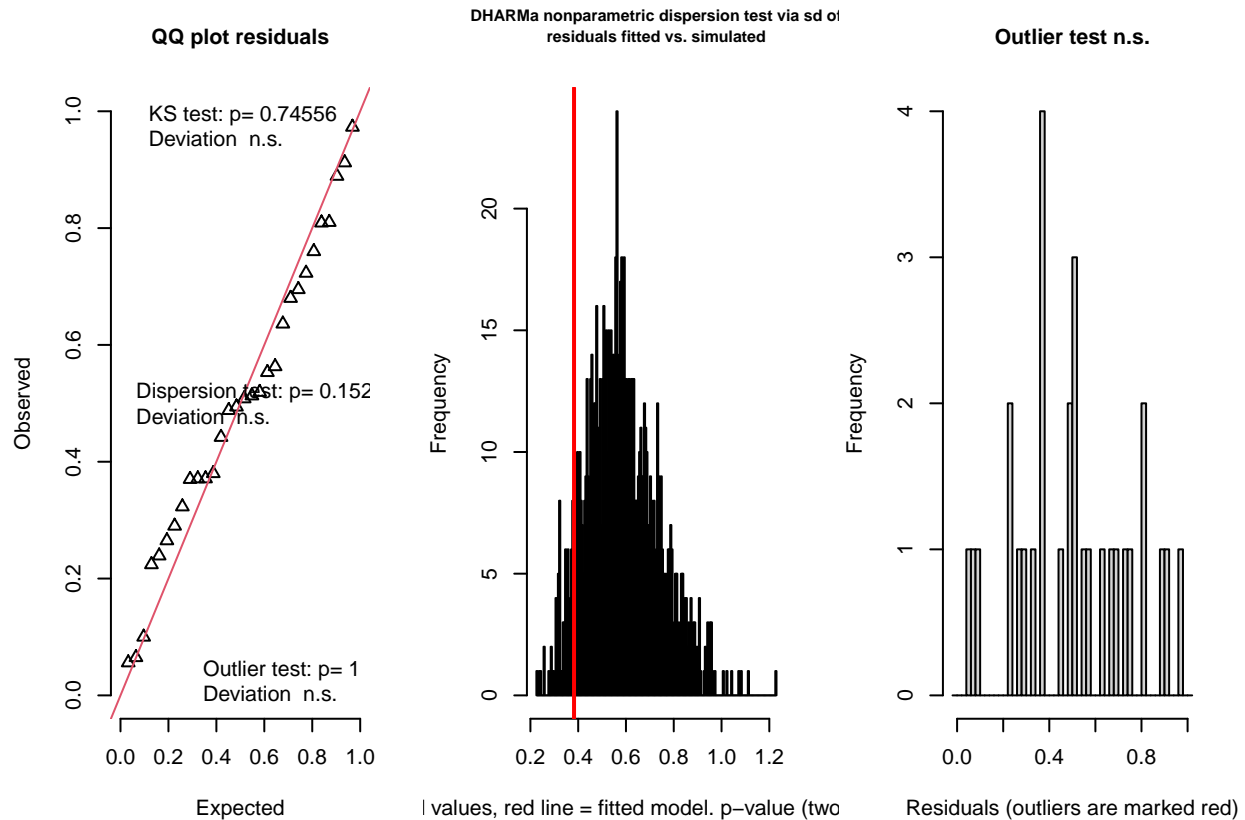


```
## summary statistics
## -----
## min: 1    max: 62
## median: 29
## mean: 27.46667
## estimated sd: 16.49605
## estimated skewness: 0.1784444
## estimated kurtosis: 2.434842
```

```
ob2 <- glm(observed ~ Treatment + Linage, data = dR2nd)
Anova(ob2, type = "II")
```

```
## Analysis of Deviance Table (Type II tests)
##
## Response: observed
##          LR Chisq Df Pr(>Chisq)
## Treatment   3.1731  1   0.07486 .
## Linage     18.8360  9   0.02662 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
res_ob2 <- simulateResiduals(ob2, n = 1000)
testResiduals(res_ob2)
```



```
## $uniformity
##
## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.124, p-value = 0.7456
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.65369, p-value = 0.152
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 30, p-value = 1
```



```
## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.0000000 0.1157033
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
## 0

## $uniformity
##
## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.124, p-value = 0.7456
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.65369, p-value = 0.152
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 30, p-value = 1
## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.0000000 0.1157033
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
## 0
```

## Beta Diversity

### Beta diversity and microbiome divergence

Beta diversity quantifies dissimilarity in community composition between samples. Dissimilarity can be also quantified by distance or divergence. These measures have a broad use in statistical data analysis.

```
rel.CR <- CR %>%
  transform_sample_counts(function(x) {x/sum(x)} )

rel.R2nd <- R2nd %>%
  transform_sample_counts(function(x) {x/sum(x)} )
```

**relative Abundance plots** agglomerate data to Species level, transform to rel. abundance, melt long format and sort data frame alph. by Species

```
Fungi_Species1 <- CR %>%
  tax_glom(taxrank = "Species") %>%
  transform_sample_counts(function(x) {x/sum(x)} ) %>%
  psmelt() %>%
  arrange(Species)

Fungi_Species2 <- R2nd %>%
  tax_glom(taxrank = "Species") %>%
  transform_sample_counts(function(x) {x/sum(x)} ) %>%
  psmelt() %>%
  arrange(Species)
```

find the mean and standard deviation by Species

```
rem <- subset(Fungi_Species1, Treatment == "removal")
mean.rem <- tapply(rem$Abundance, rem$Species, mean)
mean.rem
```

##	Acremonium_biseptum	Acremonium_spc
##	4.822926e-02	1.825963e-04
##	Ascomycota_spc_spc	Aspergillus_costiformis
##	1.921889e-03	2.792152e-05
##	Chaetomiaceae_spc	Chaetomium_globosum
##	1.527656e-03	5.751256e-01
##	Chaetomium_spc	Diatrypella_vulgaris
##	5.217036e-03	7.370756e-05
##	Eurotiales_spc	Eurotiomycetes_spc
##	1.585447e-04	6.946107e-05
##	Hypocreales_spc	Ophiostomataceae_spc
##	3.086386e-04	1.346598e-04
##	Ophiostomatales_spc	Paecilomyces_spc
##	1.862119e-04	4.264759e-05
##	Penicillium_commune	Penicillium_spc
##	7.113496e-03	2.412351e-03
##	Raffaelea_canadensis	Raffaelea_spc
##	3.755341e-02	1.845842e-03
##	Raffaelea_sulphurea	Sordariales_spc
##	3.076030e-01	3.200399e-04
##	Sordariomycetes_spc	Sporothrix_stenoceras
##	8.133021e-03	1.045215e-05
##	Talaromyces_rugulosus	Thermoascaceae_spc
##	1.273286e-03	8.691522e-05
##	Xenopolyscytalum_sp._I12F_02273	Xylariales_spc
##	3.227091e-04	1.196866e-04

```
SD.rem <- tapply(rem$Abundance, rem$Species, sd)
SD.rem
```

##	Acremonium_biseptum	Acremonium_spc
##	2.129341e-01	7.380513e-04

##	Ascomycota_spc_spc	Aspergillus_costiformis
##	7.016012e-03	1.248688e-04
##	Chaetomiaceae_spc	Chaetomium_globosum
##	2.777907e-03	3.863477e-01
##	Chaetomium_spc	Diatrypella_vulgaris
##	5.606403e-03	3.109255e-04
##	Eurotiales_spc	Eurotiomycetes_spc
##	5.080569e-04	1.680653e-04
##	Hypocreales_spc	Ophiostomataceae_spc
##	1.303065e-03	3.373166e-04
##	Ophiostomatales_spc	Paecilomyces_spc
##	3.095223e-04	1.685520e-04
##	Penicillium_commune	Penicillium_spc
##	3.178475e-02	1.076897e-02
##	Raffaelea_canadensis	Raffaelea_spc
##	5.985818e-02	4.362601e-03
##	Raffaelea_sulphurea	Sordariales_spc
##	3.306086e-01	4.252054e-04
##	Sordariomycetes_spc	Sporothrix_stenoceras
##	8.708500e-03	3.295761e-05
##	Talaromyces_rugulosus	Thermoascaceae_spc
##	4.413233e-03	2.700954e-04
##	Xenopolyscytalum_sp._I12F_02273	Xylariales_spc
##	7.445484e-04	5.352549e-04

```
sec <- subset(Fungi_Species2, Treatment == "2nd-foundation")
mean.sec <- tapply(sec$Abundance, sec$Species, mean)
mean.sec
```

##	Acremonium_biseptum	Acremonium_spc
##	0.0236606934	0.0039535863
##	Ascomycota_spc_spc	Aspergillus_costiformis
##	0.0032163024	0.0000000000
##	Chaetomiaceae_spc	Chaetomium_globosum
##	0.0073566780	0.5073275038
##	Chaetomium_spc	Diatrypella_vulgaris
##	0.0085913959	0.0000000000
##	Eurotiales_spc	Eurotiomycetes_spc
##	0.0007557944	0.0000877943
##	Hypocreales_spc	Ophiostomataceae_spc
##	0.0032611238	0.0005139207
##	Ophiostomatales_spc	Paecilomyces_spc
##	0.0007235568	0.0000000000
##	Penicillium_commune	Penicillium_spc
##	0.0316822915	0.0072273551
##	Raffaelea_canadensis	Raffaelea_spc
##	0.1175589366	0.0072412047
##	Raffaelea_sulphurea	Sordariales_spc
##	0.2531221210	0.0016430442
##	Sordariomycetes_spc	Sporothrix_stenoceras
##	0.0207058085	0.0000000000
##	Talaromyces_rugulosus	Thermoascaceae_spc
##	0.0000000000	0.0000000000
##	Xenopolyscytalum_sp._I12F_02273	Xylariales_spc

```
##                0.0000000000                0.0013708885
```

```
SD.sec <- tapply(sec$Abundance, sec$Species, sd)
SD.sec
```

```
##      Acremonium_biseptum      Acremonium_spc
##      0.0784736423      0.0131125623
##      Ascomycota_spc_spc      Aspergillus_costiformis
##      0.0089808210      0.0000000000
##      Chaetomiaceae_spc      Chaetomium_globosum
##      0.0091540276      0.3533535207
##      Chaetomium_spc      Diatrypella_vulgaris
##      0.0099731689      0.0000000000
##      Eurotiales_spc      Eurotiomycetes_spc
##      0.0025066865      0.0002911808
##      Hypocreales_spc      Ophiostomataceae_spc
##      0.0108159242      0.0013543344
##      Ophiostomatales_spc      Paecilomyces_spc
##      0.0011593941      0.0000000000
##      Penicillium_commune      Penicillium_spc
##      0.1050782734      0.0239371110
##      Raffaelea_canadensis      Raffaelea_spc
##      0.1306026315      0.0142054109
##      Raffaelea_sulphurea      Sordariales_spc
##      0.2771804794      0.0029798185
##      Sordariomycetes_spc      Sporothrix_stenoceras
##      0.0243502551      0.0000000000
##      Talaromyces_rugulosus      Thermoascaceae_spc
##      0.0000000000      0.0000000000
##      Xenopolyscytalum_sp._I12F_02273      Xylariales_spc
##      0.0000000000      0.0045467228
```

```
control <- subset(Fungi_Species1, Treatment == "control")
mean.control <- tapply(control$Abundance, control$Species, mean)
mean.control
```

```
##      Acremonium_biseptum      Acremonium_spc
##      2.018503e-03      6.968641e-06
##      Ascomycota_spc_spc      Aspergillus_costiformis
##      6.965760e-05      8.988316e-05
##      Chaetomiaceae_spc      Chaetomium_globosum
##      4.590560e-04      3.796378e-01
##      Chaetomium_spc      Diatrypella_vulgaris
##      2.400139e-03      0.000000e+00
##      Eurotiales_spc      Eurotiomycetes_spc
##      0.000000e+00      0.000000e+00
##      Hypocreales_spc      Ophiostomataceae_spc
##      0.000000e+00      2.804989e-04
##      Ophiostomatales_spc      Paecilomyces_spc
##      2.988574e-04      0.000000e+00
##      Penicillium_commune      Penicillium_spc
##      1.328225e-05      0.000000e+00
##      Raffaelea_canadensis      Raffaelea_spc
```

```
##          1.104248e-01          3.631699e-03
##          Raffaelea_sulphurea          Sordariales_spc
##          4.948844e-01          1.094521e-04
##          Sordariomycetes_spc          Sporothrix_stenoceras
##          5.383031e-03          0.000000e+00
##          Talaromyces_rugulosus          Thermoascaceae_spc
##          2.899075e-04          0.000000e+00
## Xenopolyscytalum_sp._I12F_02273          Xylariales_spc
##          2.028727e-06          0.000000e+00
```

```
SD.control <- tapply(control$Abundance, control$Species, sd)
SD.control
```

```
##          Acremonium_biseptum          Acremonium_spc
##          9.027020e-03          3.116471e-05
##          Ascomycota_spc_spc          Aspergillus_costiformis
##          1.752356e-04          2.666390e-04
##          Chaetomiaceae_spc          Chaetomium_globosum
##          7.011706e-04          2.980927e-01
##          Chaetomium_spc          Diatrypella_vulgaris
##          2.929313e-03          0.000000e+00
##          Eurotiales_spc          Eurotiomycetes_spc
##          0.000000e+00          0.000000e+00
##          Hypocreales_spc          Ophiostomataceae_spc
##          0.000000e+00          1.103085e-03
##          Ophiostomatales_spc          Paecilomyces_spc
##          8.186757e-04          0.000000e+00
##          Penicillium_commune          Penicillium_spc
##          3.777937e-05          0.000000e+00
##          Raffaelea_canadensis          Raffaelea_spc
##          1.134036e-01          5.389495e-03
##          Raffaelea_sulphurea          Sordariales_spc
##          2.584721e-01          2.370017e-04
##          Sordariomycetes_spc          Sporothrix_stenoceras
##          9.740509e-03          0.000000e+00
##          Talaromyces_rugulosus          Thermoascaceae_spc
##          1.026525e-03          0.000000e+00
## Xenopolyscytalum_sp._I12F_02273          Xylariales_spc
##          9.072742e-06          0.000000e+00
```

edit object for plotting

```
#control vs. removal
Fungi_Species1$Species<-as.character(Fungi_Species1$Species)
Fungi_Species1$Species[Fungi_Species1$Abundance<0.05]<-"Others"
Fungi_Species1$Class<-as.character(Fungi_Species1$Class)
Fungi_Species1$Class[Fungi_Species1$Abundance<0.05]<-"Others"
Fungi_Species1$Species<-factor(Fungi_Species1$Species)
Fungi_Species1<-droplevels(Fungi_Species1)
Fungi_Species1$Treatment<-factor(Fungi_Species1$Treatment,levels=c("control", "removal"))
Fungi_Species1$Species <- factor(Fungi_Species1$Species, levels = c("Others", "Sordariomycetes_spc", "P

#removal vs. 2nd attempt
```









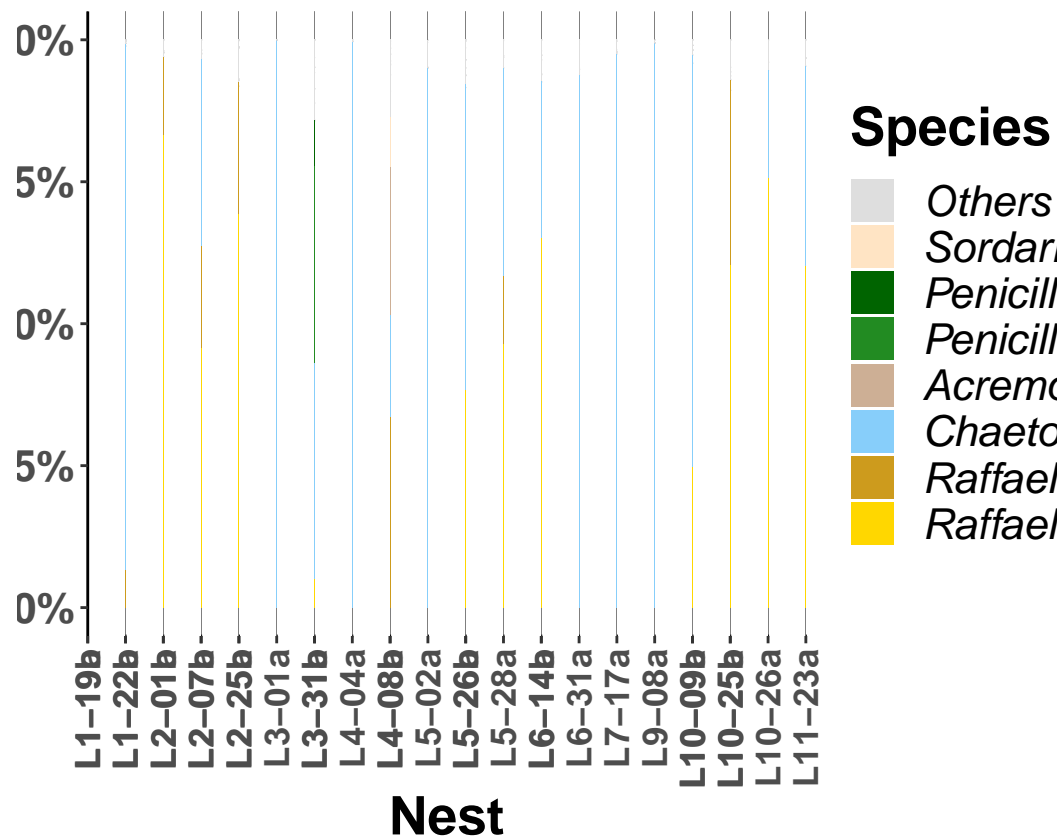
```

theme(legend.text = element_text(face = "italic"))+
theme(strip.text.x = element_blank())

relAb_fun2 <- g4 + theme(panel.spacing.x = unit(0.5, "cm"))

relAb_fun2

```



plot removal vs. 2<sup>nd</sup>-attempt

plot relative abundance with strains

```

R2nd_melt <- R2nd %>%
  transform_sample_counts(function(x) {x/sum(x)} ) %>%
  psmelt() %>%
  arrange(Species)

```

```

## Warning in psmelt(.): The sample variables:
## Sample
## have been renamed to:
## sample_Sample
## to avoid conflicts with special phyloseq plot attribute names.

```

```

R2nd_melt$best_hit<-as.character(R2nd_melt$best_hit)
R2nd_melt$best_hit[R2nd_melt$Abundance<0.05]<- "Others"

```

```

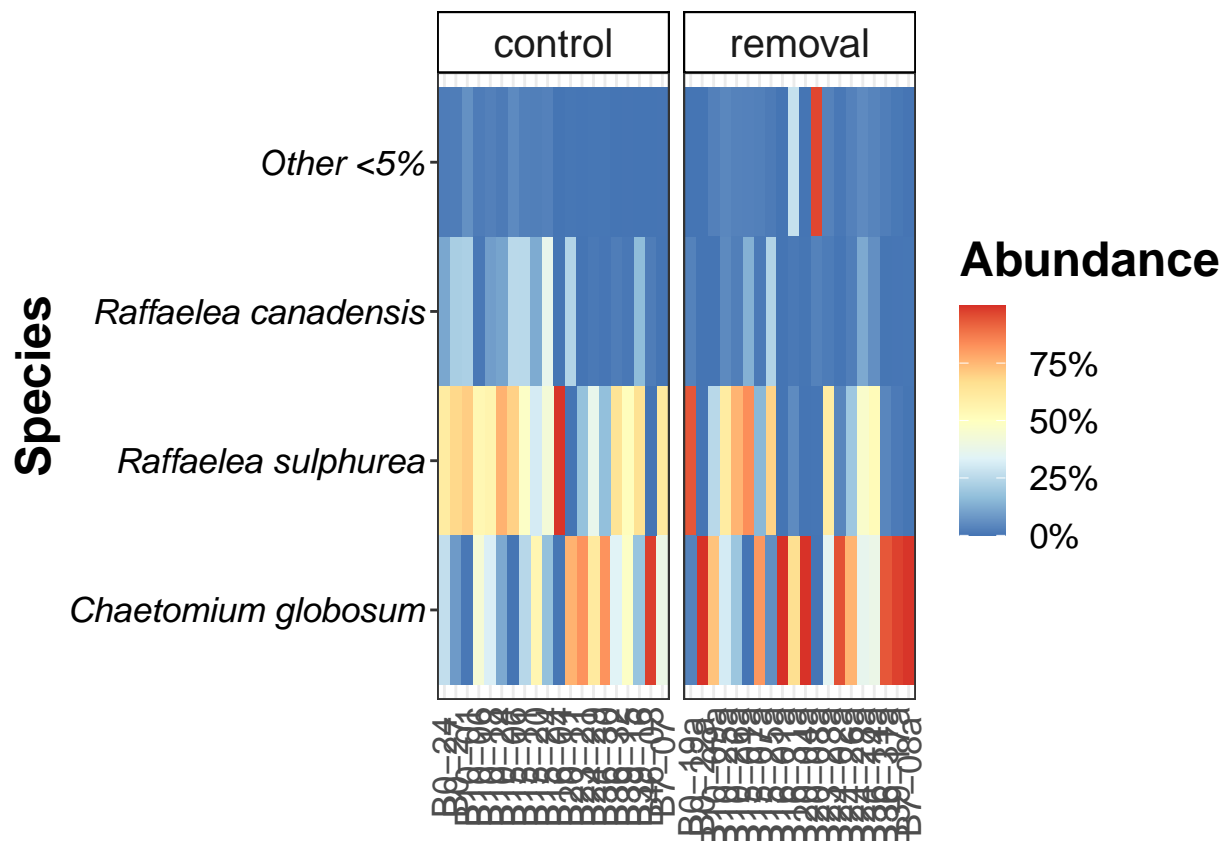
R2nd_OTU <- R2nd_melt
R2nd_OTU$Nest <- factor(R2nd_OTU$Nest, levels = c("B0-19", "B0-22", "B15-01", "B15-07", "B15-25", "B16-01", "B16-02", "B16-03", "B16-04", "B16-05", "B16-06", "B16-07", "B16-08", "B16-09", "B16-10", "B16-11", "B16-12", "B16-13", "B16-14", "B16-15", "B16-16", "B16-17", "B16-18", "B16-19", "B16-20", "B16-21", "B16-22", "B16-23", "B16-24", "B16-25", "B16-26", "B16-27", "B16-28", "B16-29", "B16-30", "B16-31", "B16-32", "B16-33", "B16-34", "B16-35", "B16-36", "B16-37", "B16-38", "B16-39", "B16-40", "B16-41", "B16-42", "B16-43", "B16-44", "B16-45", "B16-46", "B16-47", "B16-48", "B16-49", "B16-50", "B16-51", "B16-52", "B16-53", "B16-54", "B16-55", "B16-56", "B16-57", "B16-58", "B16-59", "B16-60", "B16-61", "B16-62", "B16-63", "B16-64", "B16-65", "B16-66", "B16-67", "B16-68", "B16-69", "B16-70", "B16-71", "B16-72", "B16-73", "B16-74", "B16-75", "B16-76", "B16-77", "B16-78", "B16-79", "B16-80", "B16-81", "B16-82", "B16-83", "B16-84", "B16-85", "B16-86", "B16-87", "B16-88", "B16-89", "B16-90", "B16-91", "B16-92", "B16-93", "B16-94", "B16-95", "B16-96", "B16-97", "B16-98", "B16-99", "B16-100"))

```

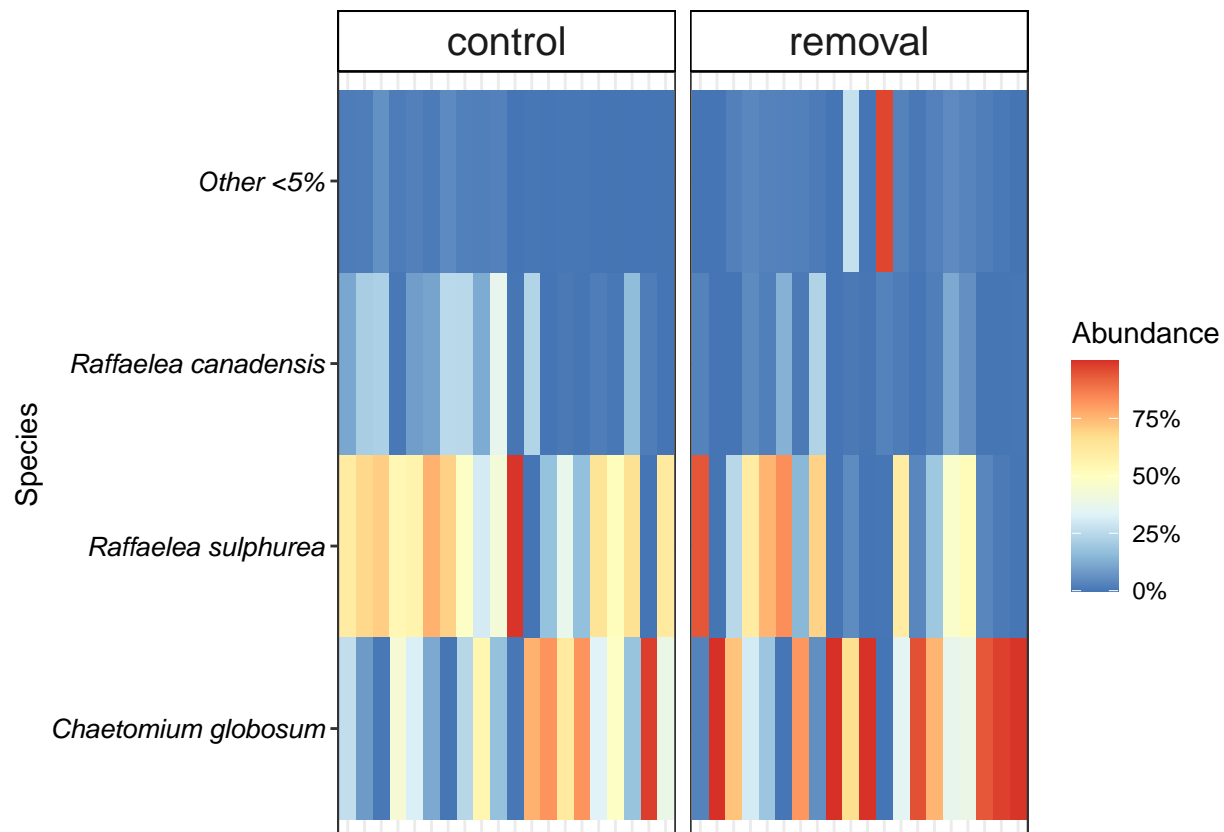


```
# Clean the facet label box
p.heat <- p.heat + theme(legend.key = element_blank(),
                        strip.background = element_rect(colour="black", fill="white"))+
                        theme(strip.text = element_text(size = 15))

p.heat + theme(axis.text.x=element_text(angle = 90, size = 13))+
theme(axis.title.y = element_text(size = 18, face = "bold"), axis.title.x = element_blank())+
theme(axis.text.y = element_text(size = 13, face = "italic"))+
theme(legend.text = element_text(size = 13))+
theme(legend.title = element_text(size = 18, face = "bold"))
```



p.heat



plot control vs. 2<sup>nd</sup> attempt (>5% abundance)

```
pseq2 <- aggregate_rare(rel.R2nd, level = "Species", detection = 5/100, prevalence = 10/100)

pseqh2 <- pseq2 %>%
  psmelt()

pseqh2$Treatment<-factor(pseqh2$Treatment,levels=c("removal", "2nd-foundation"))
pseqh2$Species<-factor(pseqh2$Species,levels=c("Chaetomium_globosum", "Raffaelea_sulphurea", "Raffaelea_
canadensis", "Other"))

p.heat <- ggplot(pseqh, aes(x = Sample, y = Species)) + geom_tile(aes(fill = Abundance))

# Change color
p.heat <- p.heat + scale_fill_distiller("Abundance", palette = "RdYlBu", labels = scales::percent) + th

# Make bacterial names italics
p.heat <- p.heat + theme(axis.text.y = element_text(colour = 'black',
  size = 10,
  face = 'italic'))

# Make separate samples based on main variable
p.heat <- p.heat + facet_grid(~Treatment,
  scales = "free") + rremove("x.text")

#Clean the x-axis
p.heat <- p.heat + theme(axis.title.x=element_blank(),
  axis.text.x=element_blank(),
```

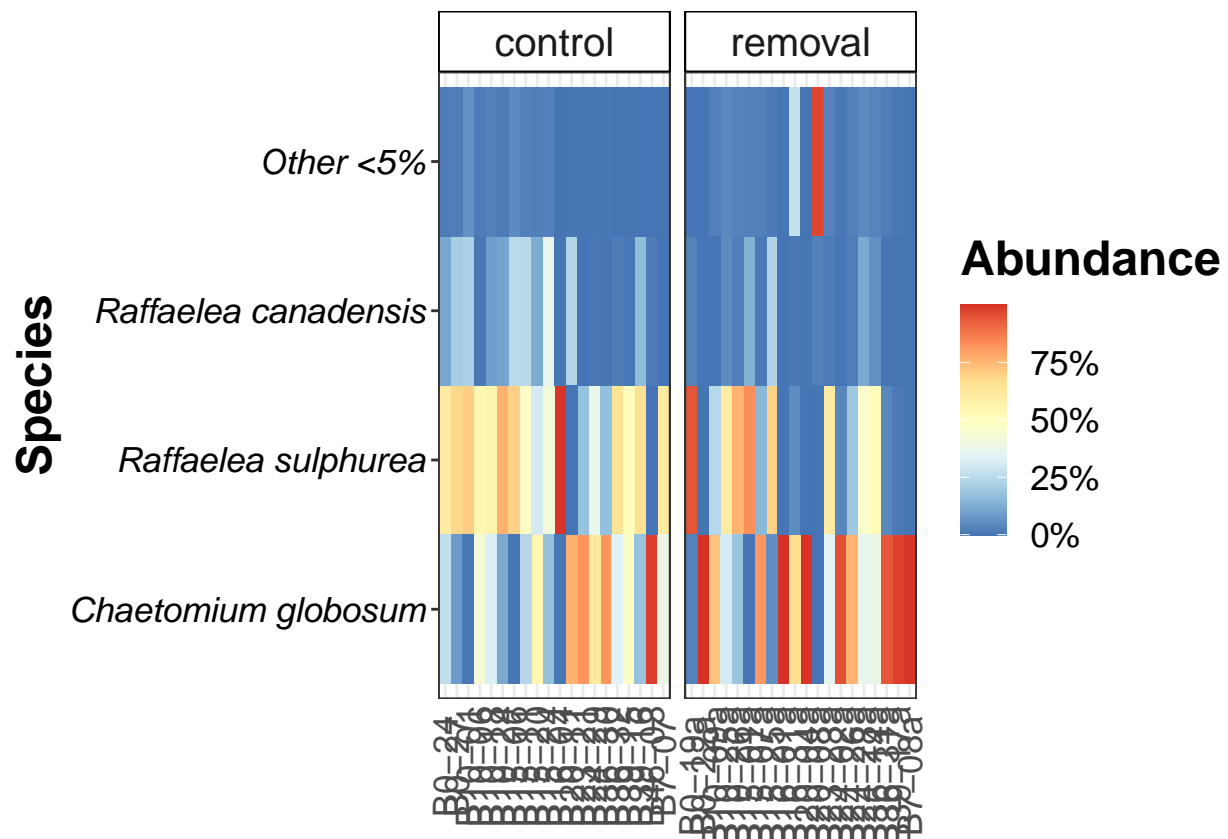
```

axis.ticks.x=element_blank())

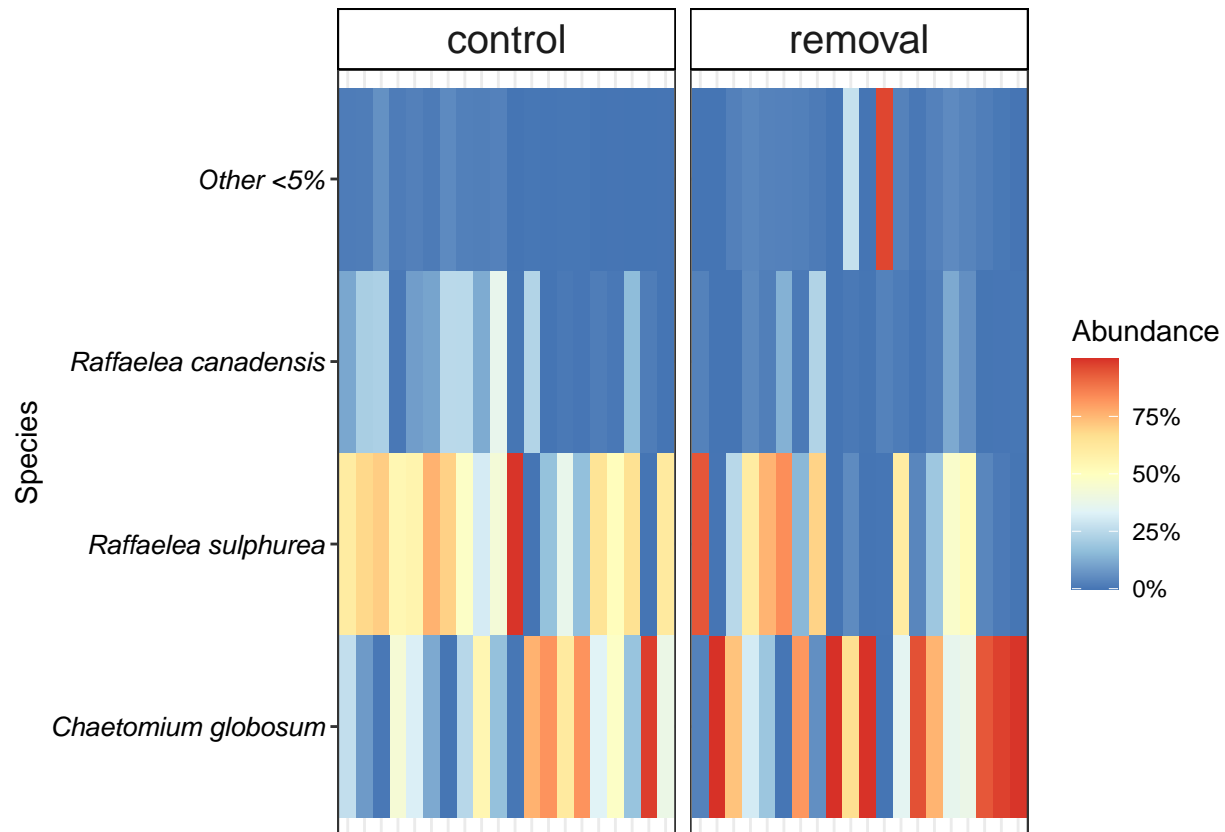
# Clean the facet label box
p.heat <- p.heat + theme(legend.key = element_blank(),
                        strip.background = element_rect(colour="black", fill="white"))+
                        theme(strip.text = element_text(size = 15))

p.heat + theme(axis.text.x=element_text(angle = 90, size = 13))+
  theme(axis.title.y = element_text(size = 18, face = "bold"), axis.title.x = element_blank())+
  theme(axis.text.y = element_text(size = 13, face = "italic"))+
  theme(legend.text = element_text(size = 13))+
  theme(legend.title = element_text(size = 18, face = "bold"))

```



p.heat



Core taxa abundance plot define color bar

```
colpal_fun <- c("lightskyblue", "gold", "goldenrod3", "forestgreen", "peachpuff3", "darkolivegreen3")
```

plot core taxa of control vs. removal with relative abundance

relative Abundance

```
ps4 <- prune_taxa(taxa_sums(rel.CR) > 0, rel.CR)
ps4 <- tax_glom(ps4, taxrank = 'Species')
psOrd4 = subset_taxa(ps4, Species == "Chaetomium_globosum" | Species == "Raffaelea_sulphurea" | Species == "Raffaelea_canadensis")
psctr = subset_samples(psOrd4, Treatment=="control")
#Melt and plot
melt3<-psmelt(psOrd4)
melt4<-psmelt(psctr)
melt3$Treament <- factor(melt3$Treatment, levels = c("control","removal"))
melt3$Species <- factor(melt3$Species, levels = c("Chaetomium_globosum", "Raffaelea_sulphurea", "Raffaelea_canadensis"))
melt4$Treament <- factor(melt4$Treatment, levels = c("control","removal"))
melt4$Species <- factor(melt4$Species, levels = c("Chaetomium_globosum", "Raffaelea_sulphurea", "Raffaelea_canadensis"))
a_mean <- melt4 %>%
  group_by(Species) %>%
  summarize(mean_val = mean(Abundance))
print(a_mean)
```

```
## # A tibble: 3 x 2
```

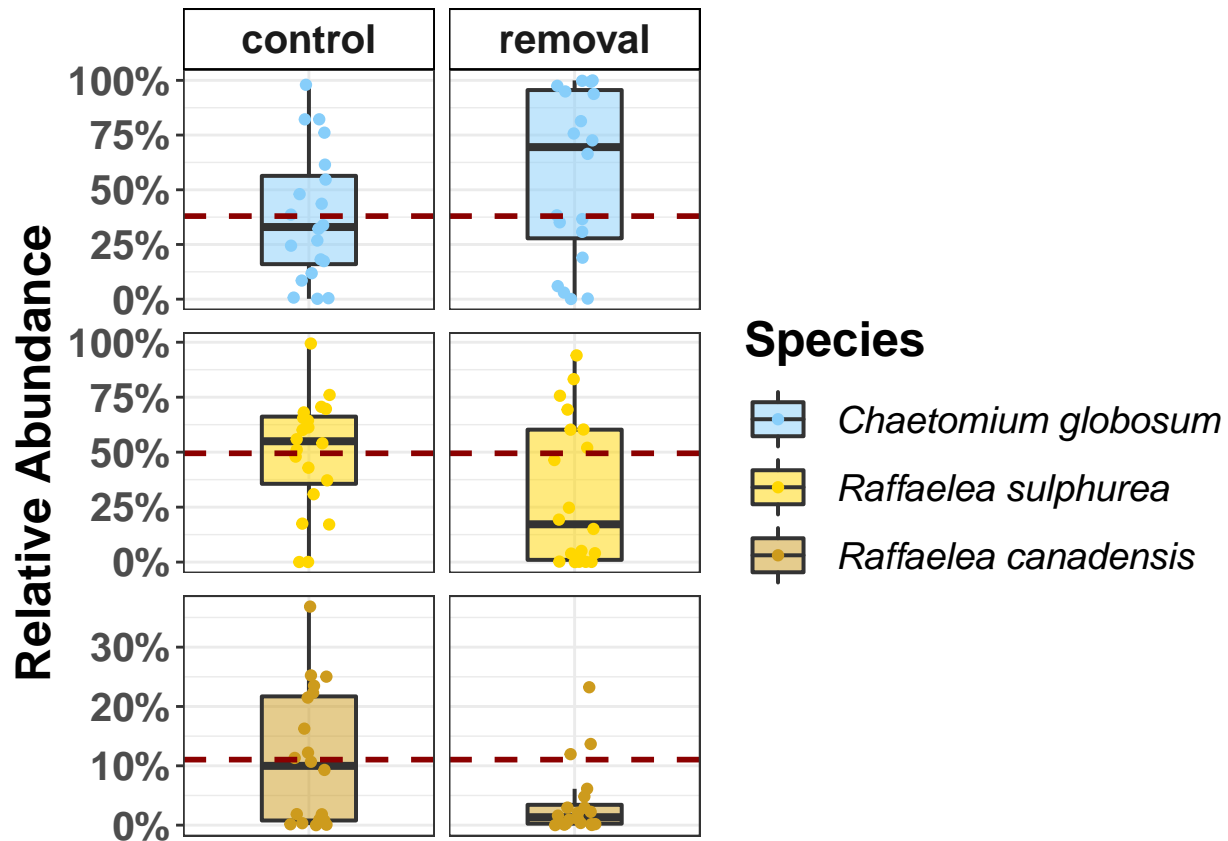
##	Species	mean_val
##	<fct>	<dbl>
## 1	Chaetomium globosum	0.380
## 2	Raffaelea sulphurea	0.495
## 3	Raffaelea canadensis	0.110

```

p2<-ggplot(data = melt3, aes(x = Treatment, y = Abundance)) +
  geom_boxplot(aes(fill=Species),alpha=0.5,lwd=0.7, position = position_dodge(width = 0.3), width=0.45,
  scale_fill_manual(values = colpal_fun, name = "Species")+
  scale_color_manual(values = colpal_fun, name = "Species")+
  labs(x = "", y = "Abundance\n")+
  facet_grid(Species~fct_relevel(Treatment, "control", "removal"), scales = "free")+theme_bw()
p2<-p2+ theme(legend.position="right")+ylab("Relative Abundance")
p2<-p2+ theme(legend.text=element_text(size=14, face = "italic"))+
  theme(legend.key = element_rect(color = NA, fill = NA),legend.key.size = unit(0.9, "cm"))+
  theme(legend.title = element_text(size = 18, face = "bold"))+
  scale_y_continuous(labels=percent_format())
abu2<-p2 + theme(strip.background =element_rect(fill="white", color="black"))+
  theme(strip.text.x = element_text(size = 15, face = "bold"))
Treatment2<-abu2+theme(axis.title.y = element_text(size=18, face="bold"))+theme(axis.text.y = element_t
  theme(axis.text.x = element_text(size=18, angle=90,face="bold"))+
  theme(axis.title.x = element_blank(), axis.text.x=element_blank(),
    axis.ticks.x=element_blank())+
  theme(strip.text.y = element_blank())
relAbcoreCR <- Treatment2 + theme(panel.spacing.y = unit(0.3, "cm"))

relAbcoreCR

```



plot core taxa of removal vs. 2<sup>nd</sup> attempt with relative abundance

```
Treats <- c('removal' = "removal", '2nd-foundation' = "2nd-att.")
ps3 <- prune_taxa(taxa_sums(rel.R2nd) > 0, rel.R2nd)
ps3 <- tax_glom(ps3, taxrank = 'Species')
psOrd3 = subset_taxa(ps3, Species == "Chaetomium_globosum" | Species == "Raffaelea_sulphurea" | Species == "Raffaelea_canadensis")
psctr = subset_samples(psOrd3, Treatment == "removal")
#Melt and plot
melt<-psmelt(psOrd3)
melt2<-psmelt(psctr)
melt$Treatment <- factor(melt$Treatment, levels = c("removal", "2nd-foundation"))
melt$Species <- factor(melt$Species, levels = c("Chaetomium_globosum", "Raffaelea_sulphurea", "Raffaelea_canadensis"))
melt2$Treatment <- factor(melt2$Treatment, levels = c("removal", "2nd-foundation"))
melt2$Species <- factor(melt2$Species, levels = c("Chaetomium_globosum", "Raffaelea_sulphurea", "Raffaelea_canadensis"))
a_mean <- melt2 %>%
  group_by(Species) %>%
  summarize(mean_val = mean(Abundance))
print(a_mean)
```

```
## # A tibble: 3 x 2
##   Species          mean_val
##   <fct>          <dbl>
## 1 Chaetomium globosum 0.575
## 2 Raffaelea sulphurea 0.308
## 3 Raffaelea canadensis 0.0376
```

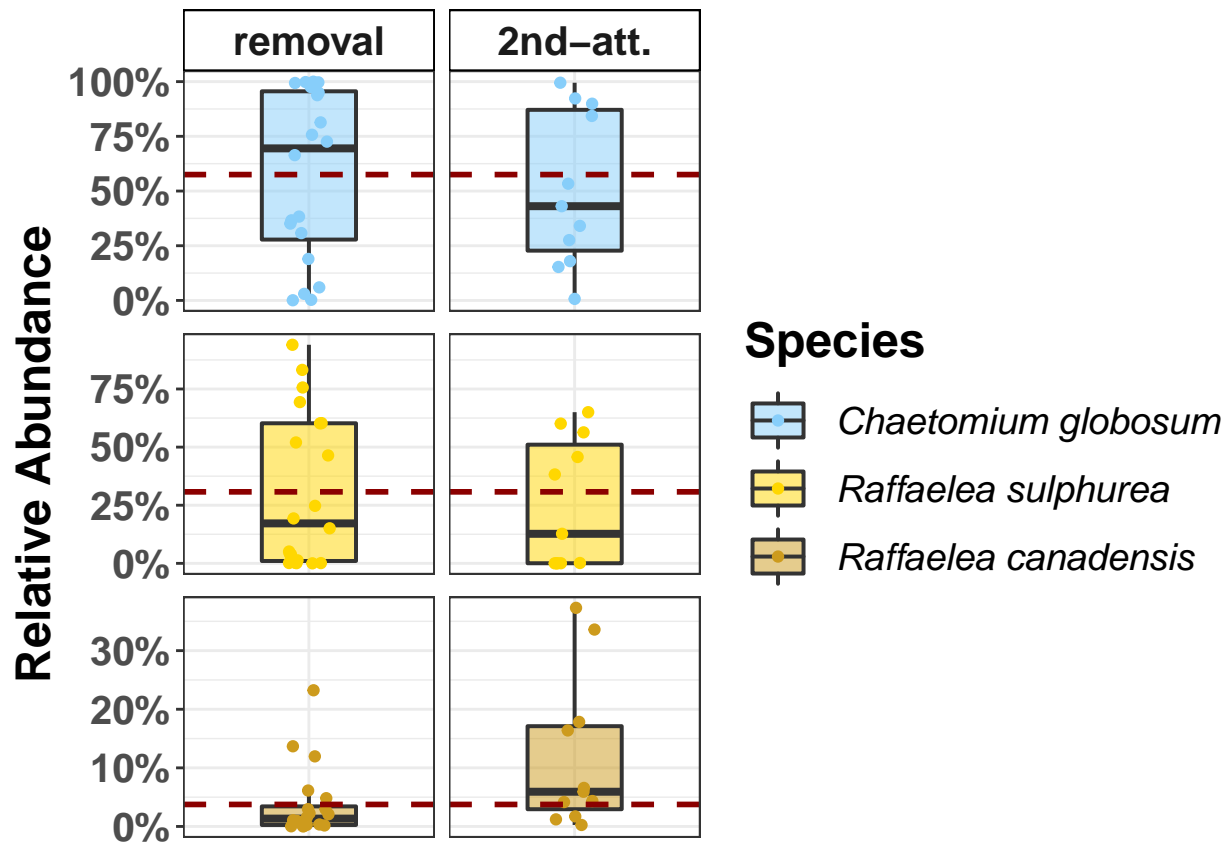


```

p<-ggplot(data = melt, aes(x = Treatment, y = Abundance)) +
  geom_boxplot(aes(fill=Species),alpha=0.5,lwd=0.7, position = position_dodge(width = 0.3), width=0.45,
  scale_fill_manual(values = colpal_fun, name = "Species")+
  scale_color_manual(values = colpal_fun, name = "Species")+
  labs(x = "", y = "Abundance\n")+
  facet_grid(Species~fct_relevel(Treatment, "removal", "2nd-foundation"), labeller = as_labeller(Treats
  theme_bw()
p<-p+ theme(legend.position="right")+
  ylab("Relative Abundance")
p<-p+ theme(legend.text=element_text(size=14, face = "italic"))+theme(legend.key = element_rect(color =
  scale_y_continuous(labels=percent_format())
abu<-p + theme(strip.background =element_rect(fill="white", color="black"))+
  theme(strip.text.x = element_text(size = 15, face = "bold"))
TreatmentCR<-abu+theme(axis.title.y = element_text(size=18, face="bold"))+
  theme(axis.text.y = element_text(size=15, face="bold")) +
  theme(axis.text.x = element_text(size=18, angle=90,face="bold"))+
  theme(axis.title.x = element_blank(), axis.text.x=element_blank(),
    axis.ticks.x=element_blank())+
  theme(strip.text.y = element_blank())
relAbcoreR2nd_fun <- TreatmentCR + theme(panel.spacing.y = unit(0.3, "cm"))

relAbcoreR2nd_fun

```



Ordination analysis

## Sample ordination

```
set.seed(1)
ordi.CR = ordinate(rel.CR, "NMDS", "bray", k=3, trymax=100)
```

## NMDS

```
## Run 0 stress 0.0346926
## Run 1 stress 0.03892109
## Run 2 stress 0.04011849
## Run 3 stress 0.03469555
## ... Procrustes: rmse 0.0006566098  max resid 0.00193732
## ... Similar to previous best
## Run 4 stress 0.03911702
## Run 5 stress 0.03469257
## ... New best solution
## ... Procrustes: rmse 4.163446e-05  max resid 0.0001359446
## ... Similar to previous best
## Run 6 stress 0.03418742
## ... New best solution
## ... Procrustes: rmse 0.03183313  max resid 0.1870848
## Run 7 stress 0.03418735
## ... New best solution
## ... Procrustes: rmse 0.0002360602  max resid 0.0008162302
## ... Similar to previous best
## Run 8 stress 0.03419229
## ... Procrustes: rmse 0.0008910375  max resid 0.004884492
## ... Similar to previous best
## Run 9 stress 0.03907424
## Run 10 stress 0.03902262
## Run 11 stress 0.03469247
## Run 12 stress 0.03849743
## Run 13 stress 0.0346925
## Run 14 stress 0.0388932
## Run 15 stress 0.03469337
## Run 16 stress 0.03418733
## ... New best solution
## ... Procrustes: rmse 0.0001105944  max resid 0.000336442
## ... Similar to previous best
## Run 17 stress 0.03418721
## ... New best solution
## ... Procrustes: rmse 9.364042e-05  max resid 0.0004726928
## ... Similar to previous best
## Run 18 stress 0.03418759
## ... Procrustes: rmse 0.0005333684  max resid 0.002974389
## ... Similar to previous best
## Run 19 stress 0.03469241
## Run 20 stress 0.03469295
## *** Solution reached
```

```
ordi.CR$stress #0.03418721 --> good fit
```

```
## [1] 0.03418721
```

```
set.seed(1)
ordi.R2nd = ordinate(rel.R2nd, "NMDS", "bray", k=3, trymax=100)
```

```
## Run 0 stress 0.03165789
## Run 1 stress 0.03305765
## Run 2 stress 0.03345132
## Run 3 stress 0.03165796
## ... Procrustes: rmse 2.985695e-05 max resid 0.000119893
## ... Similar to previous best
## Run 4 stress 0.03512227
## Run 5 stress 0.03165843
## ... Procrustes: rmse 0.0001765007 max resid 0.0007290808
## ... Similar to previous best
## Run 6 stress 0.03305756
## Run 7 stress 0.03305762
## Run 8 stress 0.03165782
## ... New best solution
## ... Procrustes: rmse 4.148317e-05 max resid 0.0001625033
## ... Similar to previous best
## Run 9 stress 0.03165798
## ... Procrustes: rmse 7.899082e-05 max resid 0.0003264773
## ... Similar to previous best
## Run 10 stress 0.03345129
## Run 11 stress 0.03305763
## Run 12 stress 0.03512172
## Run 13 stress 0.03512318
## Run 14 stress 0.03165778
## ... New best solution
## ... Procrustes: rmse 2.469497e-05 max resid 0.000106505
## ... Similar to previous best
## Run 15 stress 0.03305745
## Run 16 stress 0.03512266
## Run 17 stress 0.03306002
## Run 18 stress 0.03165773
## ... New best solution
## ... Procrustes: rmse 0.0001553175 max resid 0.0004708727
## ... Similar to previous best
## Run 19 stress 0.03305738
## Run 20 stress 0.0334517
## *** Solution reached
```

```
ordi.R2nd$stress #0.03165773
```

```
## [1] 0.03165773
```

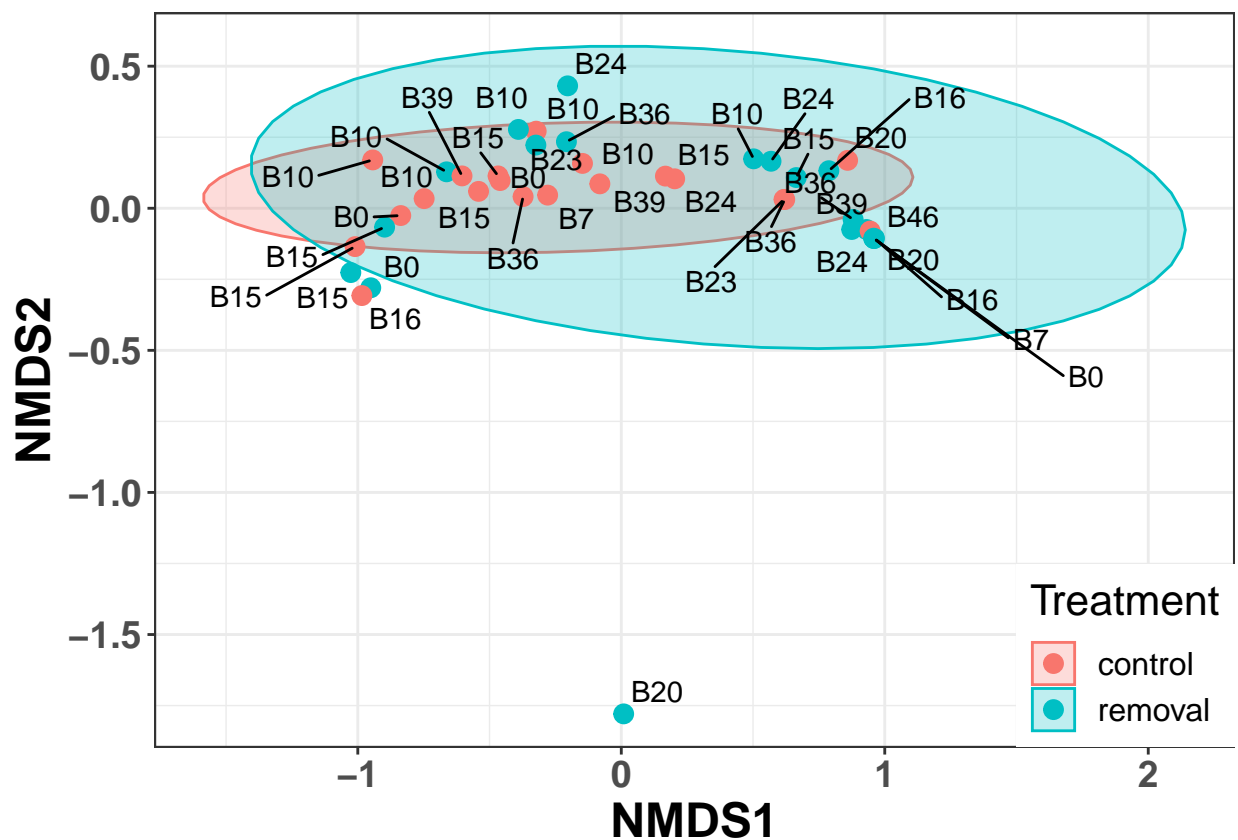
```
plot NMDS
```

```

NMDS_CR_fun <- plot_ordination(rel.CR, ordi.CR, color = "Treatment", axes=c(1,2))+
  stat_ellipse(level = 0.95, geom = "polygon", aes(fill = Treatment), alpha = 0.25)+
  geom_point(size=3, inherit.aes=T)+
  geom_text_repel(aes(label=Linage), color = "black", size = 4, max.overlaps = Inf)+
  theme_bw()+
  theme(legend.justification=c(1,0), legend.position=c(1,0))+
  theme(axis.text = element_text(size = 14, face = "bold"))+
  theme(axis.title = element_text(size = 18, face = "bold"))+
  theme(legend.text = element_text(size = 12))+
  theme(legend.title = element_text(size = 16))+
  theme(title = element_text(size = 18))

```

NMDS\_CR\_fun



```

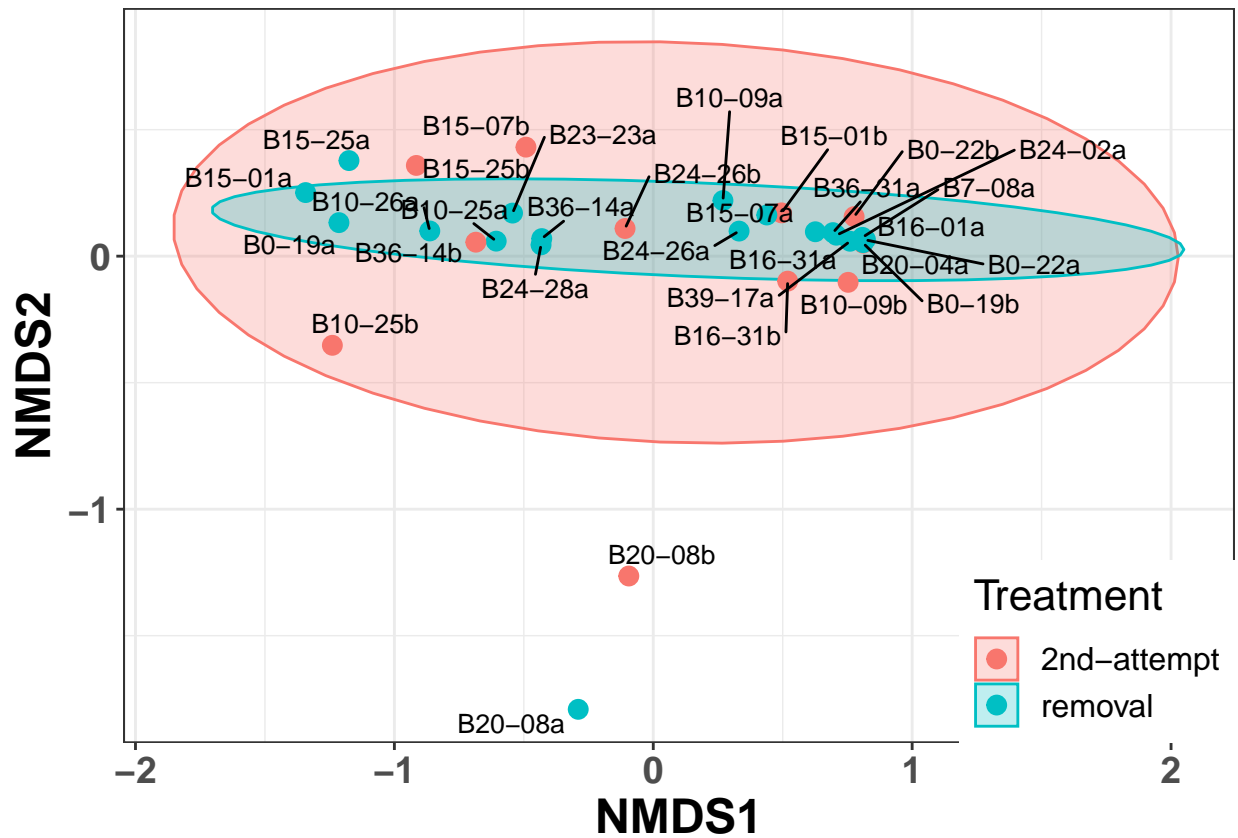
sample_data(rel.R2nd)$Treatment <- as.factor(sample_data(rel.R2nd)$Treatment)
levels(sample_data(rel.R2nd)$Treatment) = c("2nd-attempt", "removal")

NMDS_R2nd_fun <- plot_ordination(rel.R2nd, ordi.R2nd, color = "Treatment", axes=c(1,2))+
  stat_ellipse(level = 0.95, geom = "polygon", aes(fill = Treatment), alpha = 0.25)+
  geom_point(size=3, inherit.aes=T)+
  geom_text_repel(aes(label=Sample), color = "black", size = 3.5, max.overlaps = Inf)+
  theme_bw()+
  theme(legend.justification=c(1,0), legend.position=c(1,0))+
  theme(axis.text = element_text(size = 14, face = "bold"))+
  theme(axis.title = element_text(size = 18, face = "bold"))+

```

```
theme(legend.text = element_text(size = 12))+
theme(legend.title = element_text(size = 16))+
theme(title = element_text(size = 18))
```

NMDS\_R2nd\_fun



### Permanova for community-level multivariate comparisons

PERMANOVA quantifies multivariate community-level differences between groups.

```
# Pick relative abundances (compositional) and sample metadata
otu.CR <- abundances(rel.CR)
meta.CR <- meta(rel.CR)

otu.R2nd <- abundances(rel.R2nd)
meta.R2nd <- meta(rel.R2nd)
```

**PERMANOVA significance test for group-level differences** Now let us evaluate whether the Treatment has a significant effect on overall garden microbiota composition. Perform PERMANOVA:

```
# control vs. removal
set.seed(1)
adonis2(distance(rel.CR, method = "bray") ~ Treatment + Linage, data = meta.CR)
```

```
## Permutation test for adonis under reduced model
## Terms added sequentially (first to last)
## Permutation: free
## Number of permutations: 999
##
## adonis2(formula = distance(rel.CR, method = "bray") ~ Treatment + Linage, data = meta.CR)
##           Df SumOfSqs      R2      F Pr(>F)
## Treatment  1   0.5422 0.08329 3.9420  0.032 *
## Linage     10   2.1167 0.32514 1.5389  0.107
## Residual   28   3.8513 0.59157
## Total      39   6.5102 1.00000
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

#removal vs. 2nd attempt
set.seed(1)
adonis2(distance(rel.R2nd, method = "bray") ~ Treatment + Linage, data = meta.R2nd)

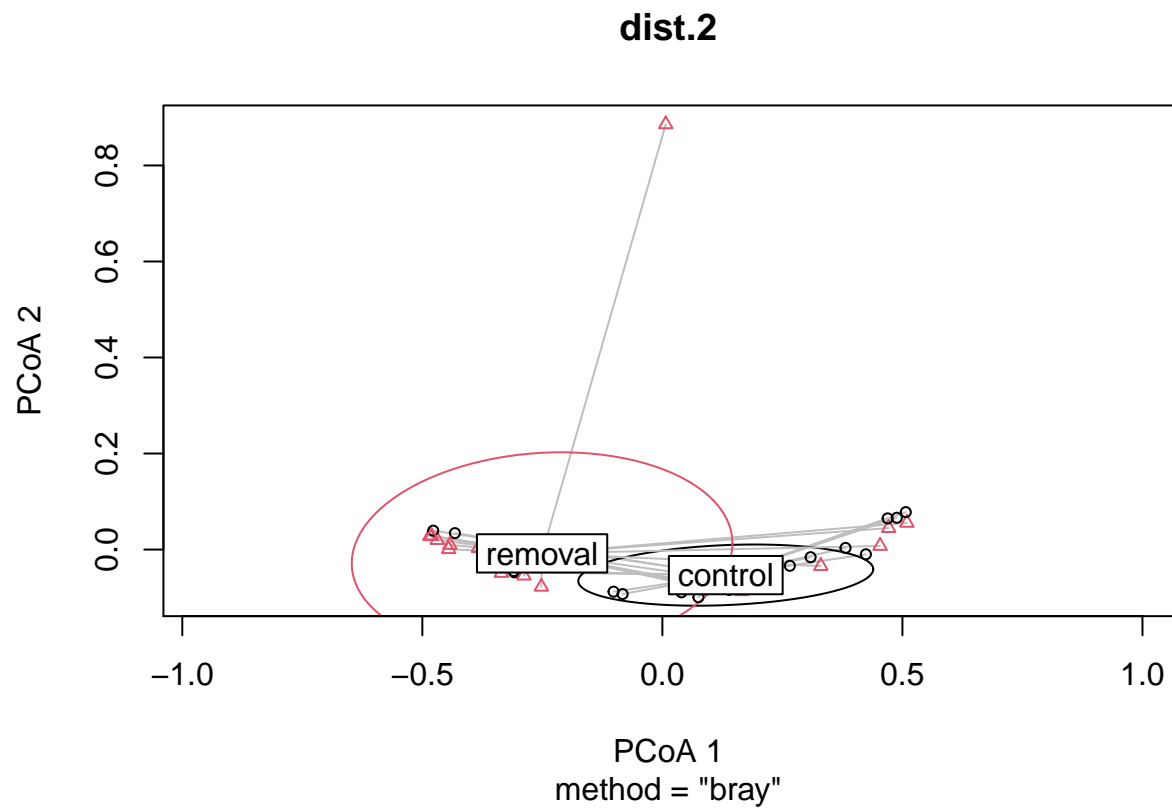
## Permutation test for adonis under reduced model
## Terms added sequentially (first to last)
## Permutation: free
## Number of permutations: 999
##
## adonis2(formula = distance(rel.R2nd, method = "bray") ~ Treatment + Linage, data = meta.R2nd)
##           Df SumOfSqs      R2      F Pr(>F)
## Treatment  1   0.0951 0.01614 0.5438  0.641
## Linage      9   2.2989 0.39029 1.4612  0.132
## Residual   20   3.4962 0.59357
## Total      30   5.8902 1.00000
```

**Checking the homogeneity condition** Check that variance homogeneity assumptions hold (to ensure the reliability of the results):

```
# control vs. removal
dist <- vegdist(t(otu.CR))
dist.2 <- betadisper(dist, meta.CR$Treatment)
anova(dist.2)

## Analysis of Variance Table
##
## Response: Distances
##           Df Sum Sq Mean Sq F value Pr(>F)
## Groups      1 0.08264 0.082640  1.7441 0.1945
## Residuals   38 1.80055 0.047383

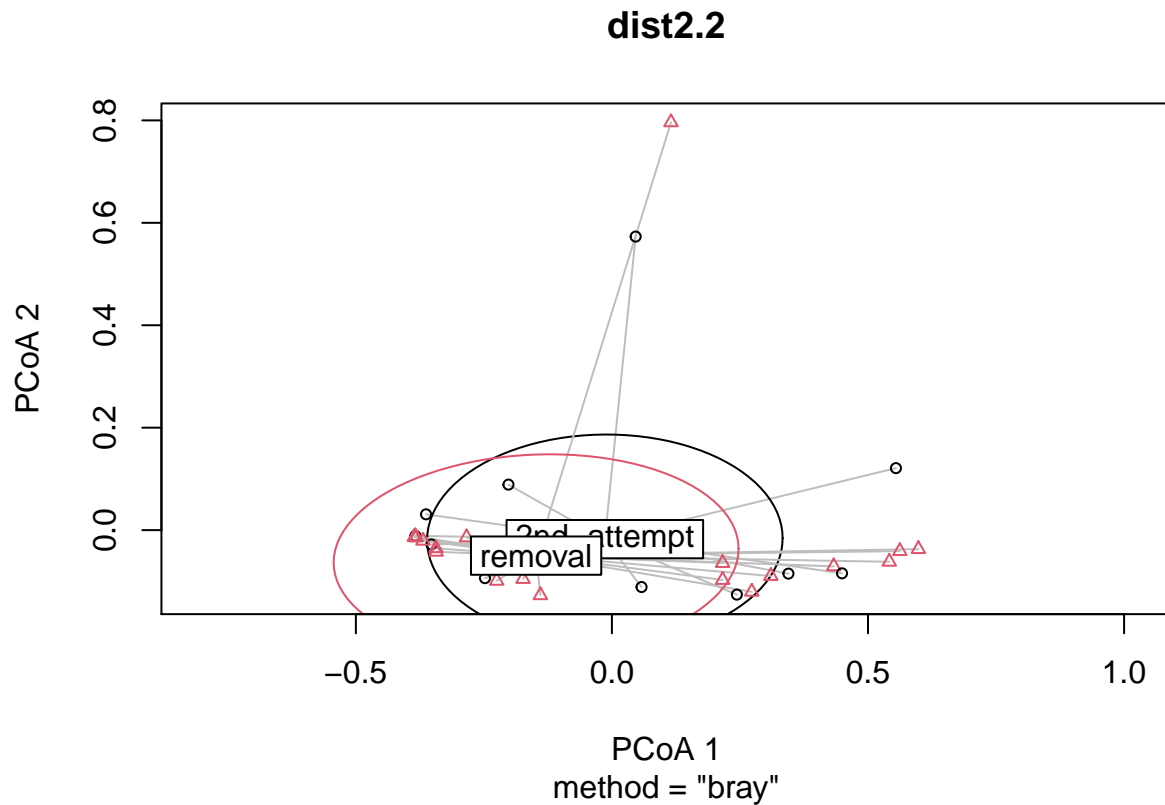
plot(dist.2, hull = FALSE, ellipse = TRUE)
```



```
# removal vs. 2nd attempt
dist2 <- vegdist(t(otu.R2nd))
dist2.2 <- betadisper(dist2, meta.R2nd$Treatment)
anova(dist2.2)
```

```
## Analysis of Variance Table
##
## Response: Distances
##          Df Sum Sq Mean Sq F value Pr(>F)
## Groups    1 0.00751 0.007511  0.1497 0.7017
## Residuals 29 1.45538 0.050185
```

```
plot(dist2.2, hull = FALSE, ellipse = TRUE)
```



### closer look at most abundant taxa

subset datasets of both combinations to most abundant Geni (each Chaetomiaceae or *Raffaelea*) (rel. abundances)

```
CG_CR <- subset_taxa(rel.CR, Species == "Chaetomium_globosum")
CG_CR <- CG_CR %>%
  tax_glom(taxrank = "Species") %>%
  psmelt()

CG_R2nd <- subset_taxa(rel.R2nd, Species == "Chaetomium_globosum")
CG_R2nd <- CG_R2nd %>%
  tax_glom(taxrank = "Species") %>%
  psmelt()

RS_CR <- subset_taxa(rel.CR, Species == "Raffaelea_sulphurea")
RS_CR <- RS_CR %>%
  tax_glom(taxrank = "Species") %>%
  psmelt()

RS_R2nd <- subset_taxa(rel.R2nd, Species == "Raffaelea_sulphurea")
RS_R2nd <- RS_R2nd %>%
  tax_glom(taxrank = "Species") %>%
  psmelt()
```



```

RC_CR <- subset_taxa(rel.CR, Species == "Raffaelea_canadensis")
RC_CR <- RC_CR %>%
  tax_glom(taxrank = "Species") %>%
  psmelt()

RC_R2nd <- subset_taxa(rel.R2nd, Species == "Raffaelea_canadensis")
RC_R2nd <- RC_R2nd %>%
  tax_glom(taxrank = "Species") %>%
  psmelt()

```

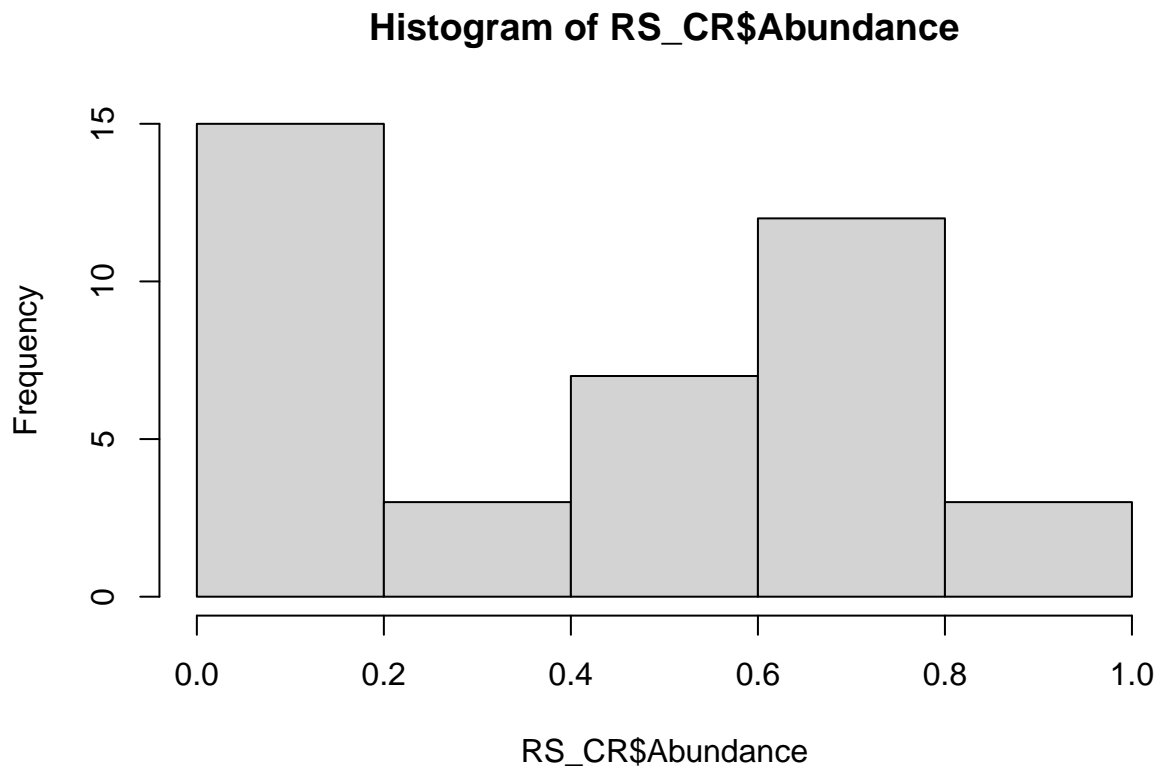
dataset *Raffaelea sulphurea*

```

RS_CR$Treatment <- as.factor(RS_CR$Treatment)
RS_CR <- within(RS_CR, Treatment <- relevel(Treatment, ref = "control"))

#control vs. removal
hist(RS_CR$Abundance)

```

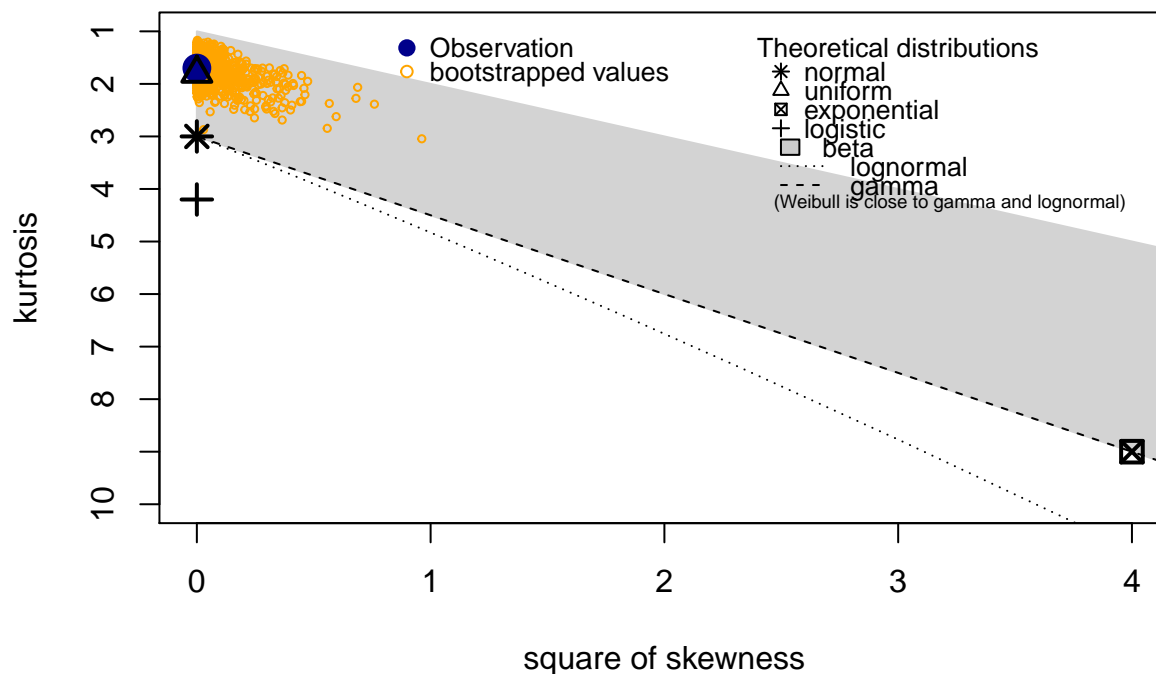


```

descdist(RS_CR$Abundance, boot = 1000)

```

## Cullen and Frey graph



```
## summary statistics
## -----
## min: 0.0001473839   max: 0.9941497
## median: 0.4719555
## mean: 0.4012437
## estimated sd: 0.3078807
## estimated skewness: 0.01282538
## estimated kurtosis: 1.698684
```

```
logistic <- function(p) log(p / (1-p) +0.01)
```

```
RS.mod <- lm(logistic(Abundance) ~ Treatment + Linage, data = RS_CR)
summary(RS.mod)
```

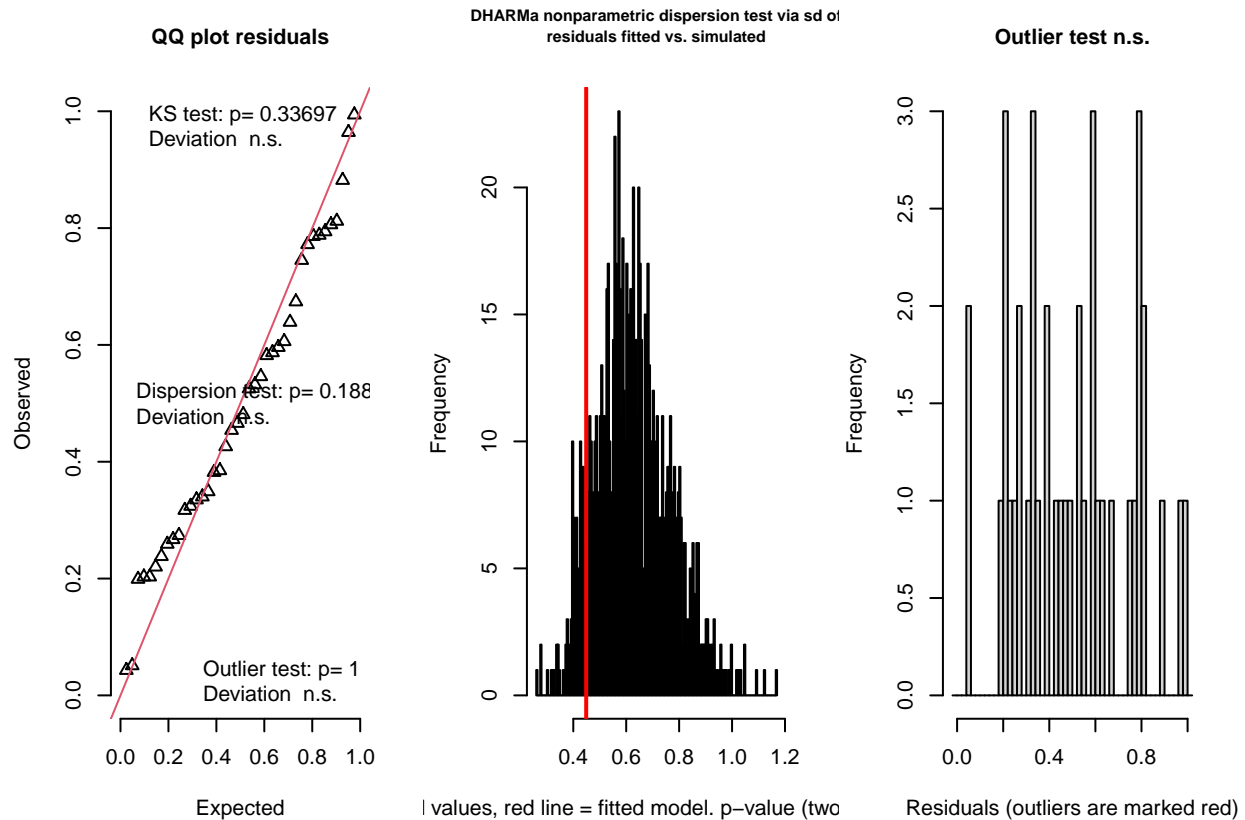
```
##
## Call:
## lm(formula = logistic(Abundance) ~ Treatment + Linage, data = RS_CR)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.6243 -1.1048 -0.0834  0.9287  4.8945
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.5922     1.0502   0.564   0.5773
```

```
## Treatmentremoval -1.4670      0.6584 -2.228  0.0341 *
## LinageB10      0.4553      1.2510  0.364  0.7186
## LinageB15      0.1023      1.2510  0.082  0.9354
## LinageB16     -0.3513      1.5273 -0.230  0.8198
## LinageB20     -4.1102      1.5273 -2.691  0.0119 *
## LinageB23     -0.4007      1.7273 -0.232  0.8182
## LinageB24     -0.7373      1.4199 -0.519  0.6077
## LinageB36     -0.8169      1.4104 -0.579  0.5671
## LinageB39     -1.1260      1.5273 -0.737  0.4671
## LinageB46     -5.1286      2.2541 -2.275  0.0307 *
## LinageB7      -1.8596      1.7273 -1.077  0.2909
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.995 on 28 degrees of freedom
## Multiple R-squared:  0.4629, Adjusted R-squared:  0.2519
## F-statistic: 2.194 on 11 and 28 DF, p-value: 0.04585
```

```
Anova(RS.mod, type = "II")
```

```
## Anova Table (Type II tests)
##
## Response: logistic(Abundance)
##      Sum Sq Df F value    Pr(>F)
## Treatment  19.752  1  4.9650 0.03407 *
## Linage    74.132 10  1.8634 0.09476 .
## Residuals 111.391 28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
res_RS <- simulateResiduals(fittedModel=RS.mod, n = 1000)
testResiduals(res_RS)
```



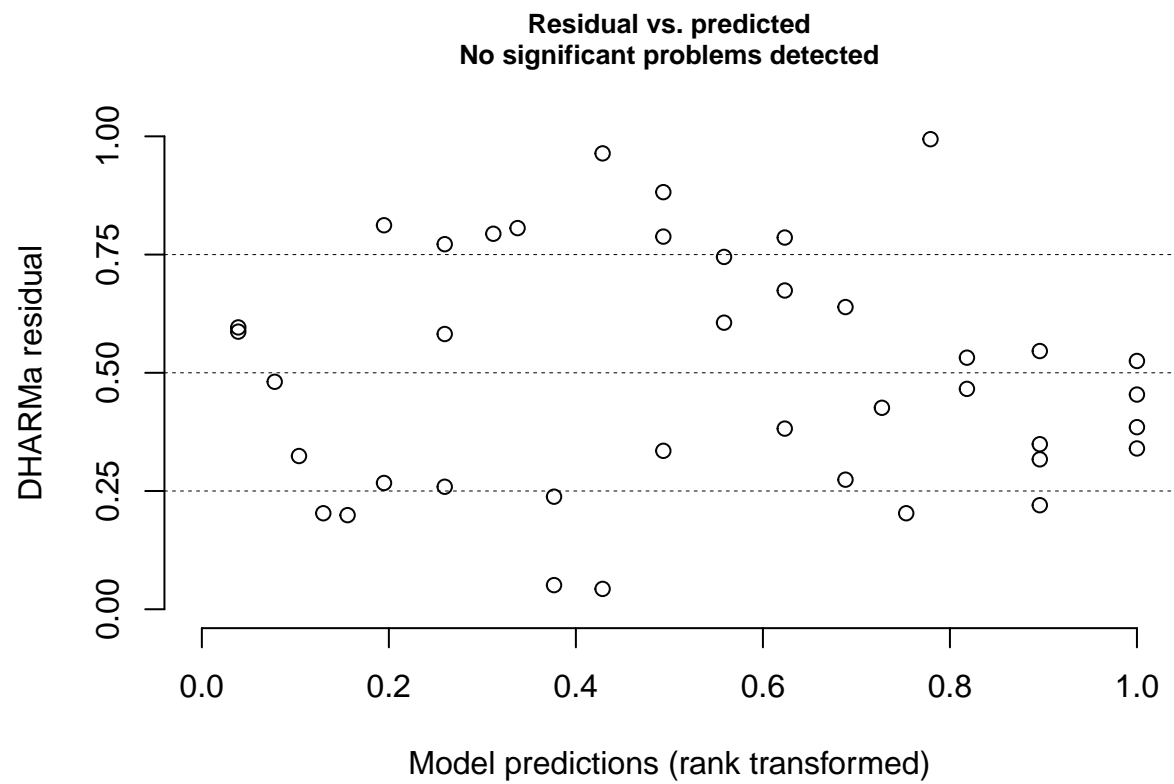
```
## $uniformity
##
## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.149, p-value = 0.337
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.71597, p-value = 0.188
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 40, p-value = 1
```

```
## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.0000000 0.0880973
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
## 0
```

```
## $uniformity
##
## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.149, p-value = 0.337
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.71597, p-value = 0.188
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 40, p-value = 1
## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.0000000 0.0880973
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
## 0
```

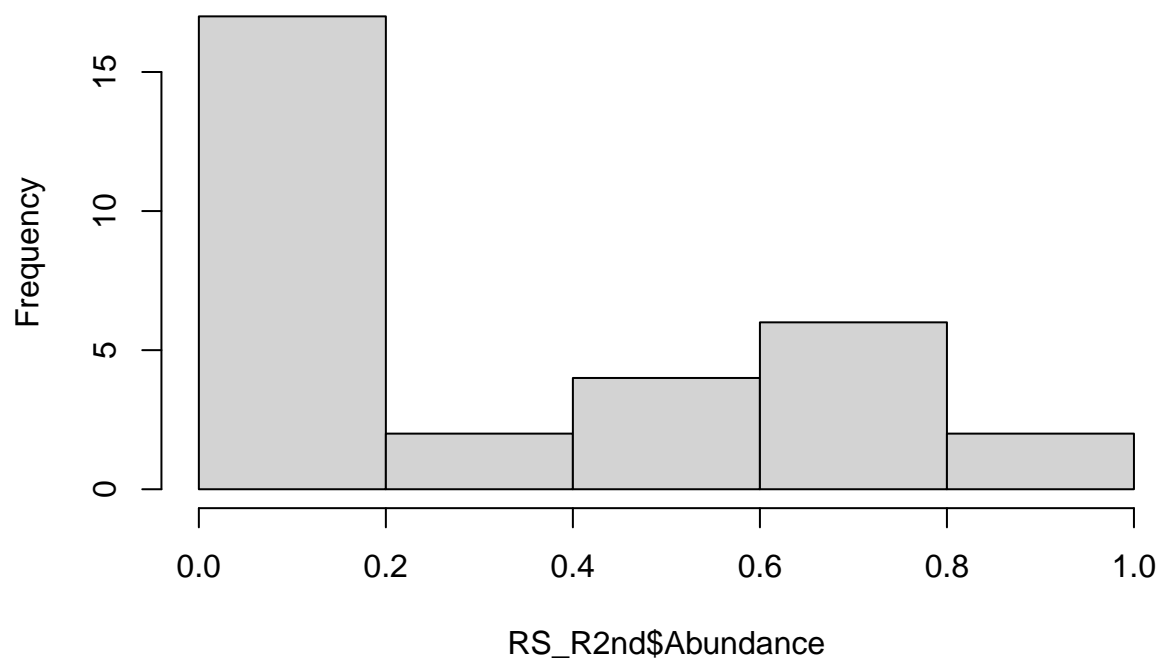
```
plotResiduals(res_RS)
```

```
## Unable to calculate quantile regression for quantile 0.25. Possibly to few (unique) data points / pr
## Unable to calculate quantile regression for quantile 0.5. Possibly to few (unique) data points / pre
## Unable to calculate quantile regression for quantile 0.75. Possibly to few (unique) data points / pr
```



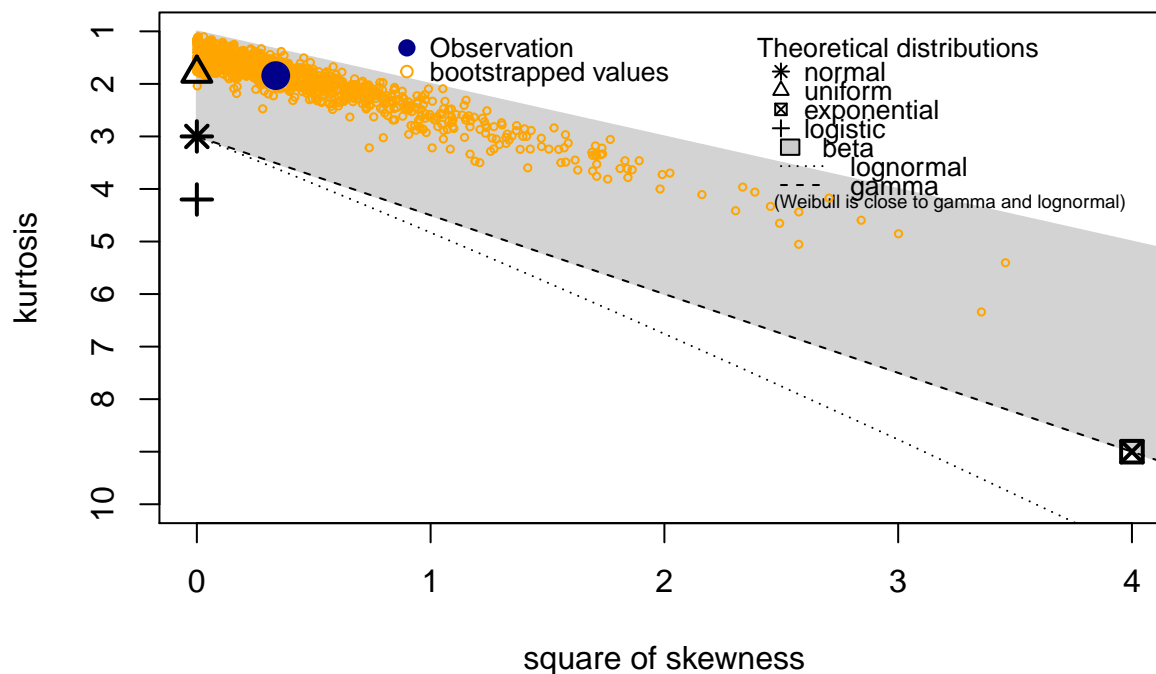
```
#removal vs. 2nd attempt  
hist(RS_R2nd$Abundance)
```

**Histogram of RS\_R2nd\$Abundance**



```
descdist(RS_R2nd$Abundance, boot = 1000)
```

## Cullen and Frey graph



```
## summary statistics
## -----
## min: 0.0001473839 max: 0.9401502
## median: 0.1507833
## mean: 0.2882711
## estimated sd: 0.3090897
## estimated skewness: 0.5810069
## estimated kurtosis: 1.842227
```

```
RS.mod2 <- lm(logistic(Abundance) ~ Treatment + Linage, data = RS_R2nd)
summary(RS.mod2)
```

```
##
## Call:
## lm(formula = logistic(Abundance) ~ Treatment + Linage, data = RS_R2nd)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -3.2509 -1.1071  0.0000  0.8656  5.0054
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    -3.1968     1.0690  -2.991  0.00723 **
## Treatmentremoval  0.9462     0.7833   1.208  0.24112
## LinageB10       1.8599     1.3367   1.391  0.17939
```

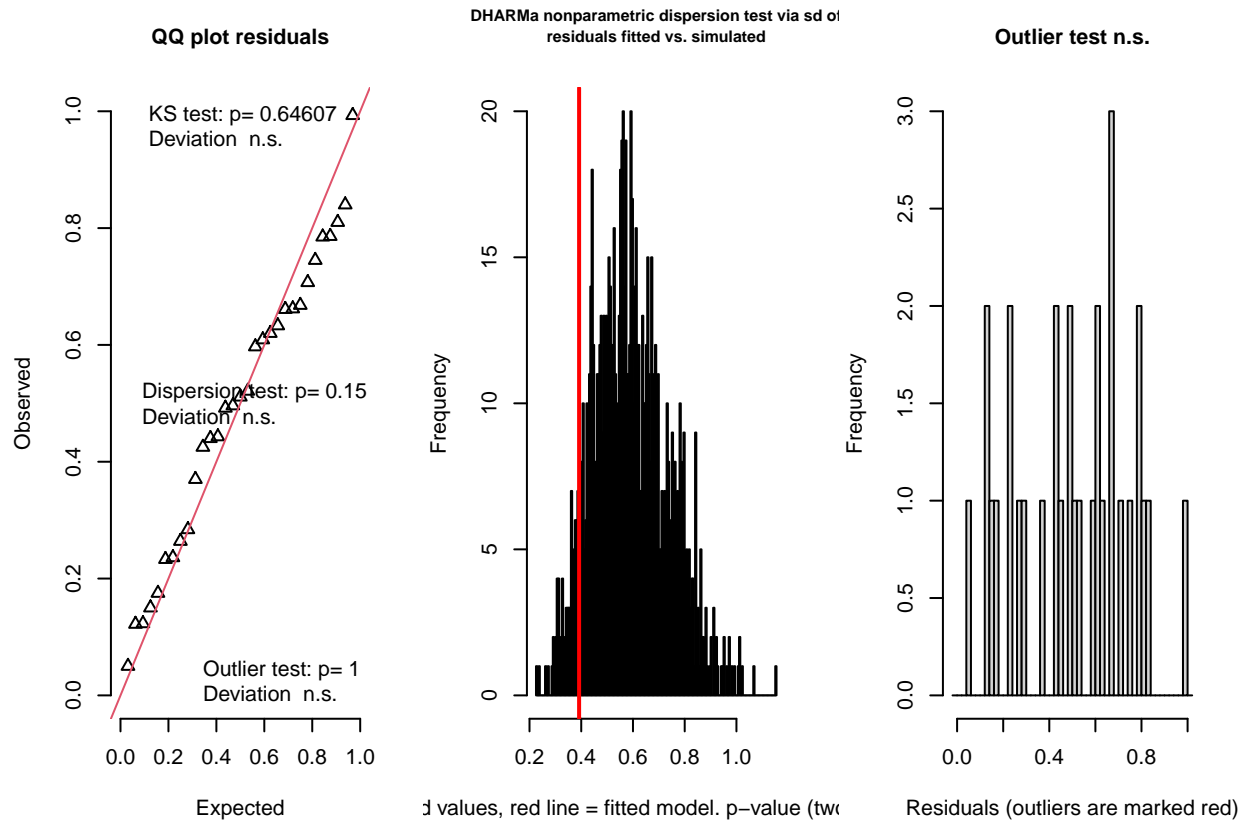


```
## LinageB15      2.5816      1.2841      2.011      0.05806 .
## LinageB16     -1.3710      1.5249     -0.899      0.37933
## LinageB20     -1.9304      1.5249     -1.266      0.22010
## LinageB23      2.6719      2.2583      1.183      0.25062
## LinageB24      1.2528      1.4202      0.882      0.38817
## LinageB36      1.8078      1.5249      1.185      0.24972
## LinageB39     -1.5180      2.2583     -0.672      0.50914
## LinageB7      -2.2154      2.2583     -0.981      0.33830
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.989 on 20 degrees of freedom
## Multiple R-squared:  0.5235, Adjusted R-squared:  0.2853
## F-statistic: 2.197 on 10 and 20 DF, p-value: 0.06429
```

```
Anova(RS.mod2, type = "II")
```

```
## Anova Table (Type II tests)
##
## Response: logistic(Abundance)
##      Sum Sq Df F value  Pr(>F)
## Treatment  5.775  1  1.4594 0.24112
## Linage    84.589  9  2.3751 0.05143 .
## Residuals 79.142 20
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
res_RS2 <- simulateResiduals(fittedModel=RS.mod2, n = 1000)
testResiduals(res_RS2)
```



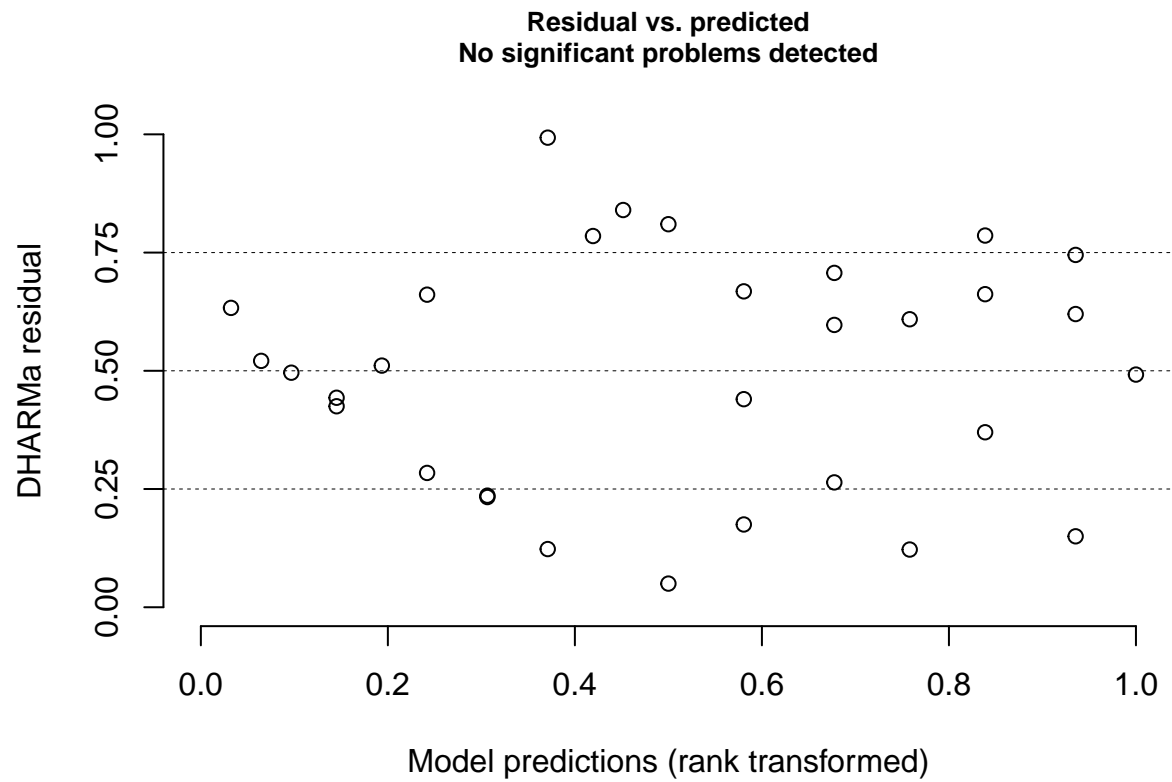
```
## $uniformity
##
## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.12774, p-value = 0.6461
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.66599, p-value = 0.15
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 31, p-value = 1
```

```
## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.0000000 0.1121887
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
## 0
```

```
## $uniformity
##
## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.12774, p-value = 0.6461
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.66599, p-value = 0.15
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 31, p-value = 1
## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.0000000 0.1121887
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
## 0
```

```
plotResiduals(res_RS2)
```

```
## Unable to calculate quantile regression for quantile 0.25. Possibly to few (unique) data points / pr
## Unable to calculate quantile regression for quantile 0.5. Possibly to few (unique) data points / pre
## Unable to calculate quantile regression for quantile 0.75. Possibly to few (unique) data points / pr
```

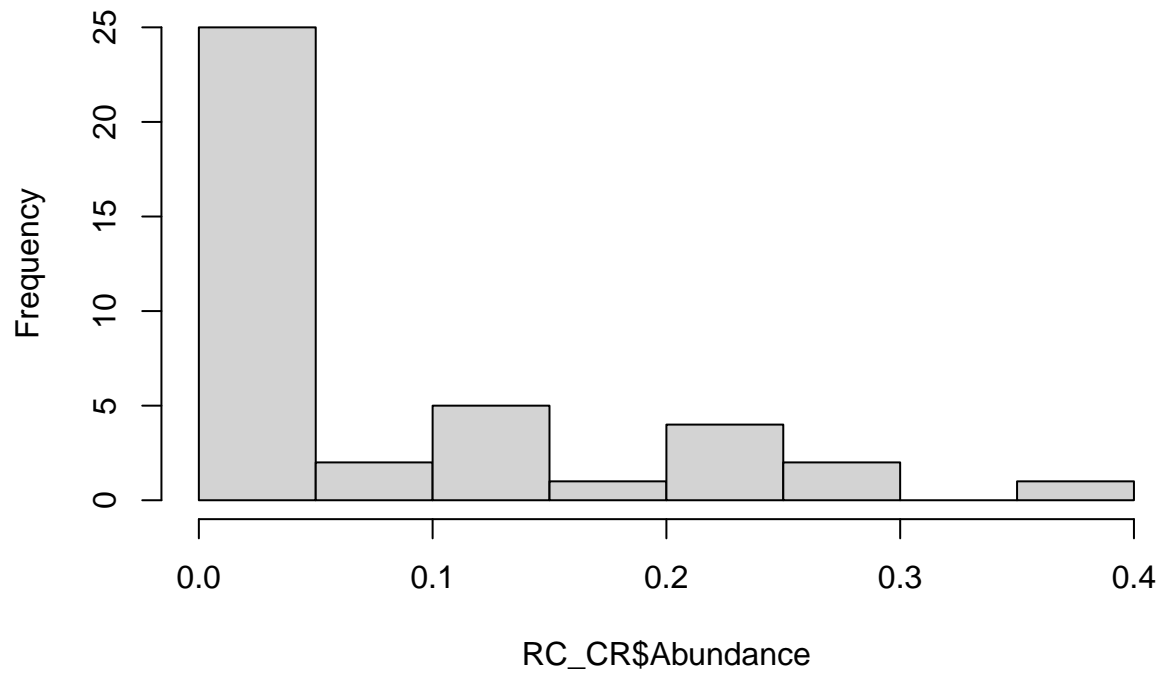


dataset *Raffaelea canadensis*

```
RC_CR$Treatment <- as.factor(RC_CR$Treatment)
RC_CR <- within(RC_CR, Treatment <- relevel(Treatment, ref = "control"))

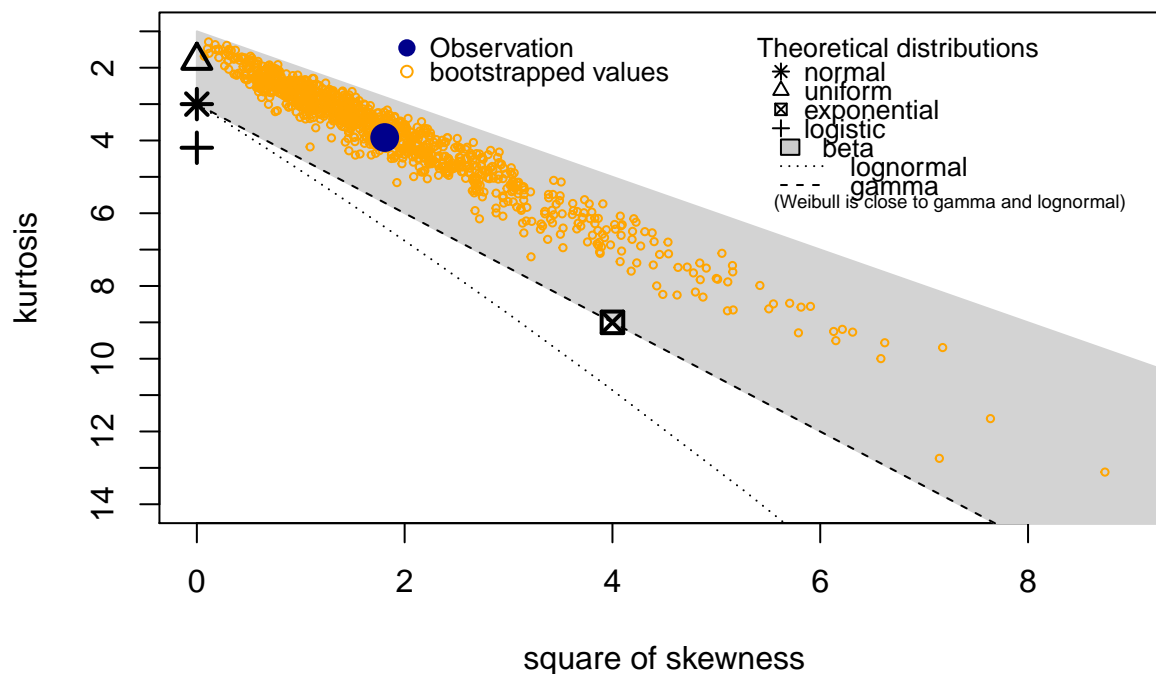
#control vs. removal
hist(RC_CR$Abundance)
```

**Histogram of RC\_CR\$Abundance**



```
descdist(RC_CR$Abundance, boot = 1000)
```

## Cullen and Frey graph



```
## summary statistics
## -----
## min: 0    max: 0.3683126
## median: 0.01966896
## mean: 0.0739891
## estimated sd: 0.09681155
## estimated skewness: 1.344583
## estimated kurtosis: 3.919206
```

```
RC_CR.mod <- lm(logistic(Abundance) ~ Treatment + Linage, data = RC_CR)
summary(RC_CR.mod)
```

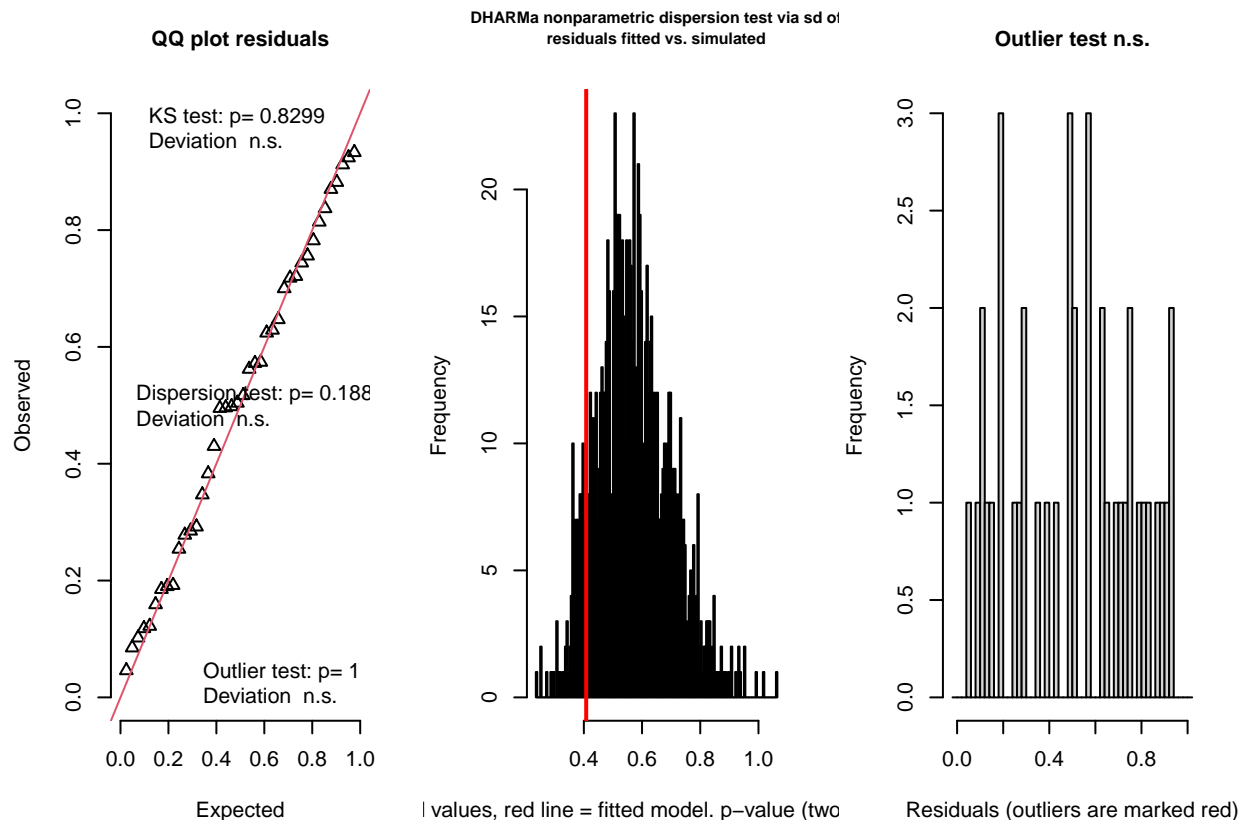
```
##
## Call:
## lm(formula = logistic(Abundance) ~ Treatment + Linage, data = RC_CR)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.78787 -0.68776  0.00395  0.58529  1.69934
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -2.34061    0.56519  -4.141 0.000287 ***
## Treatmentremoval -0.80637    0.35431  -2.276 0.030700 *
## LinageB10      -0.21455    0.67328  -0.319 0.752341
```

```
## LinageB15      1.06255    0.67328    1.578 0.125759
## LinageB16     -1.45631    0.82196   -1.772 0.087320 .
## LinageB20     -0.05887    0.82196   -0.072 0.943417
## LinageB23     -1.34091    0.92961   -1.442 0.160269
## LinageB24     -0.47751    0.76417   -0.625 0.537111
## LinageB36     -0.98234    0.75902   -1.294 0.206162
## LinageB39     -0.69249    0.82196   -0.842 0.406652
## LinageB46     -1.20338    1.21312   -0.992 0.329710
## LinageB7      -1.70636    0.92961   -1.836 0.077064 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.073 on 28 degrees of freedom
## Multiple R-squared:  0.5171, Adjusted R-squared:  0.3274
## F-statistic: 2.726 on 11 and 28 DF,  p-value: 0.01576
```

```
Anova(RC_CR.mod, type = "II")
```

```
## Anova Table (Type II tests)
##
## Response: logistic(Abundance)
##      Sum Sq Df F value    Pr(>F)
## Treatment  5.968  1  5.1797 0.03070 *
## Linage    26.848 10  2.3301 0.03805 *
## Residuals 32.262 28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
res_RC_CR <- simulateResiduals(fittedModel=RC_CR.mod, n = 1000)
testResiduals(res_RC_CR)
```



```
## $uniformity
##
## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.095, p-value = 0.8299
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.71597, p-value = 0.188
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 40, p-value = 1
```

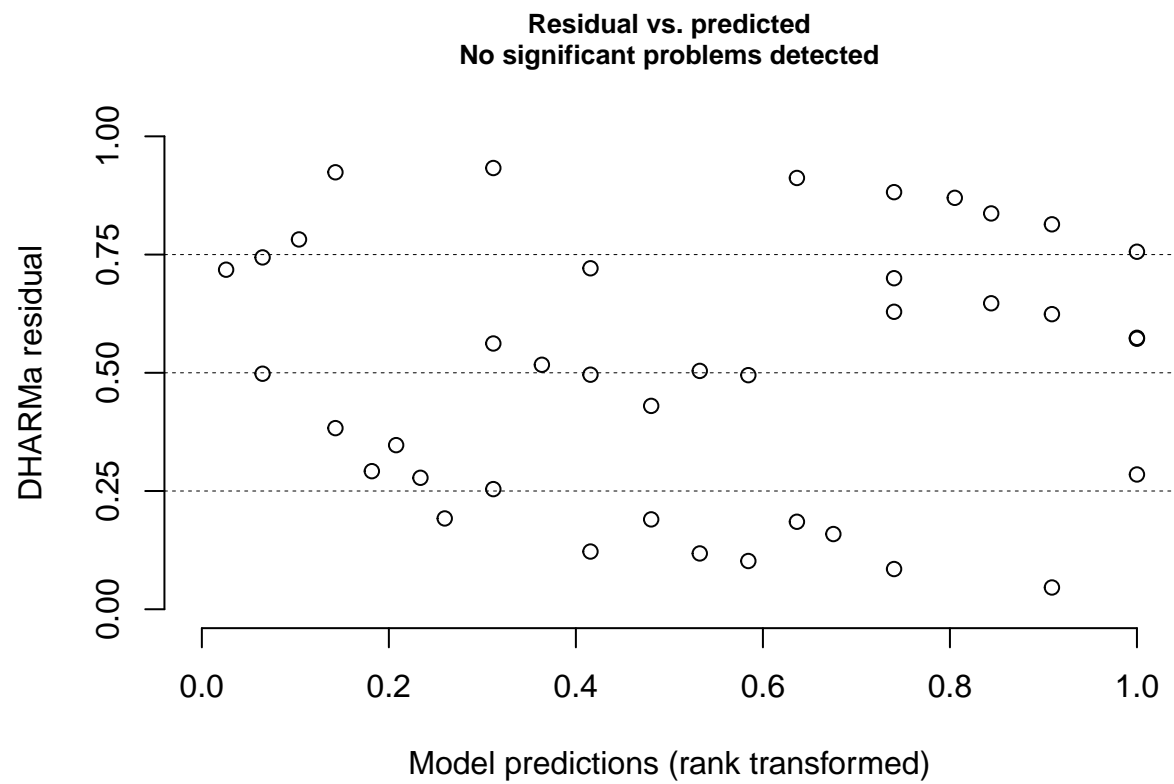


```
## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.0000000 0.0880973
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
## 0
```

```
## $uniformity
##
## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.095, p-value = 0.8299
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.71597, p-value = 0.188
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 40, p-value = 1
## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.0000000 0.0880973
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
## 0
```

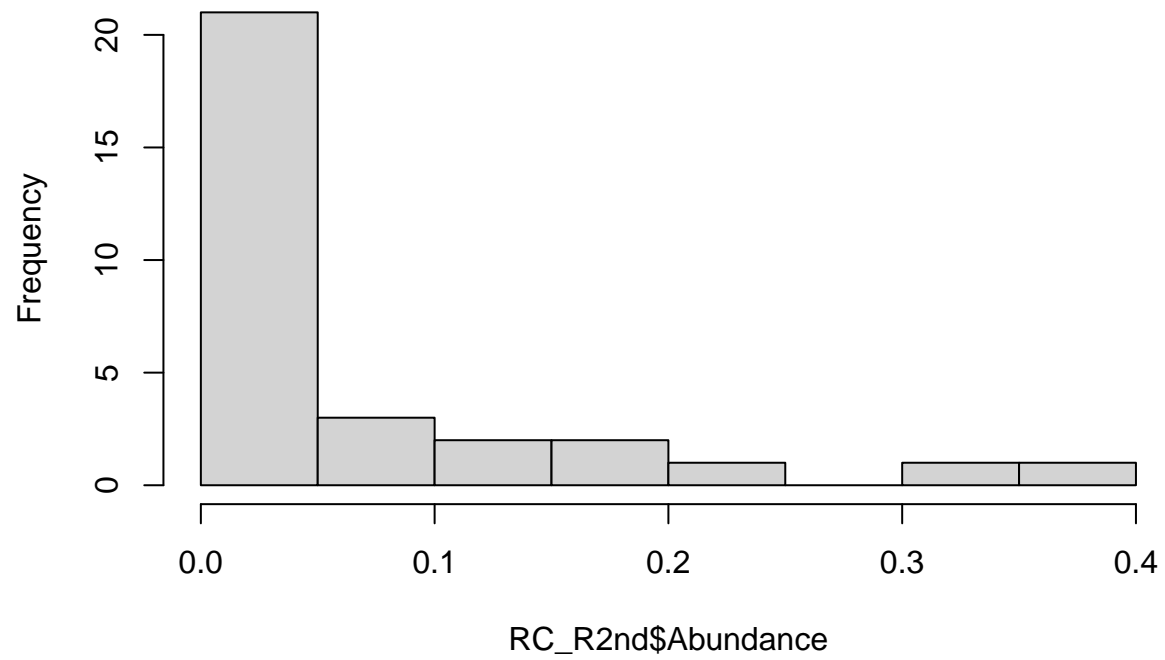
```
plotResiduals(res_RC_CR)
```

```
## Unable to calculate quantile regression for quantile 0.25. Possibly to few (unique) data points / pr
## Unable to calculate quantile regression for quantile 0.5. Possibly to few (unique) data points / pre
## Unable to calculate quantile regression for quantile 0.75. Possibly to few (unique) data points / pr
```



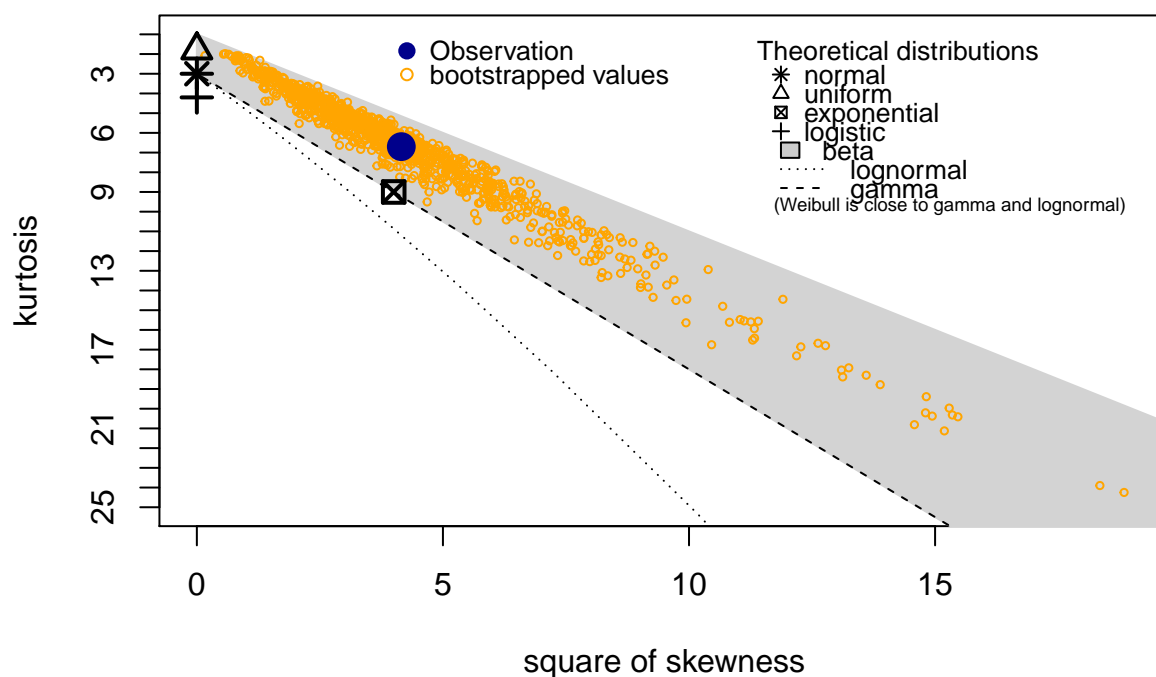
```
#removal vs. 2nd attempt  
hist(RC_R2nd$Abundance)
```

**Histogram of RC\_R2nd\$Abundance**



```
descdist(RC_R2nd$Abundance, boot = 1000)
```

## Cullen and Frey graph



```
## summary statistics
## -----
## min: 9.825596e-05 max: 0.3730086
## median: 0.02241167
## mean: 0.06594247
## estimated sd: 0.09730933
## estimated skewness: 2.038278
## estimated kurtosis: 6.698162
```

```
RC_R2nd.mod <- lm(logistic(Abundance) ~ Treatment + Linage, data = RC_R2nd)
summary(RC_R2nd.mod)
```

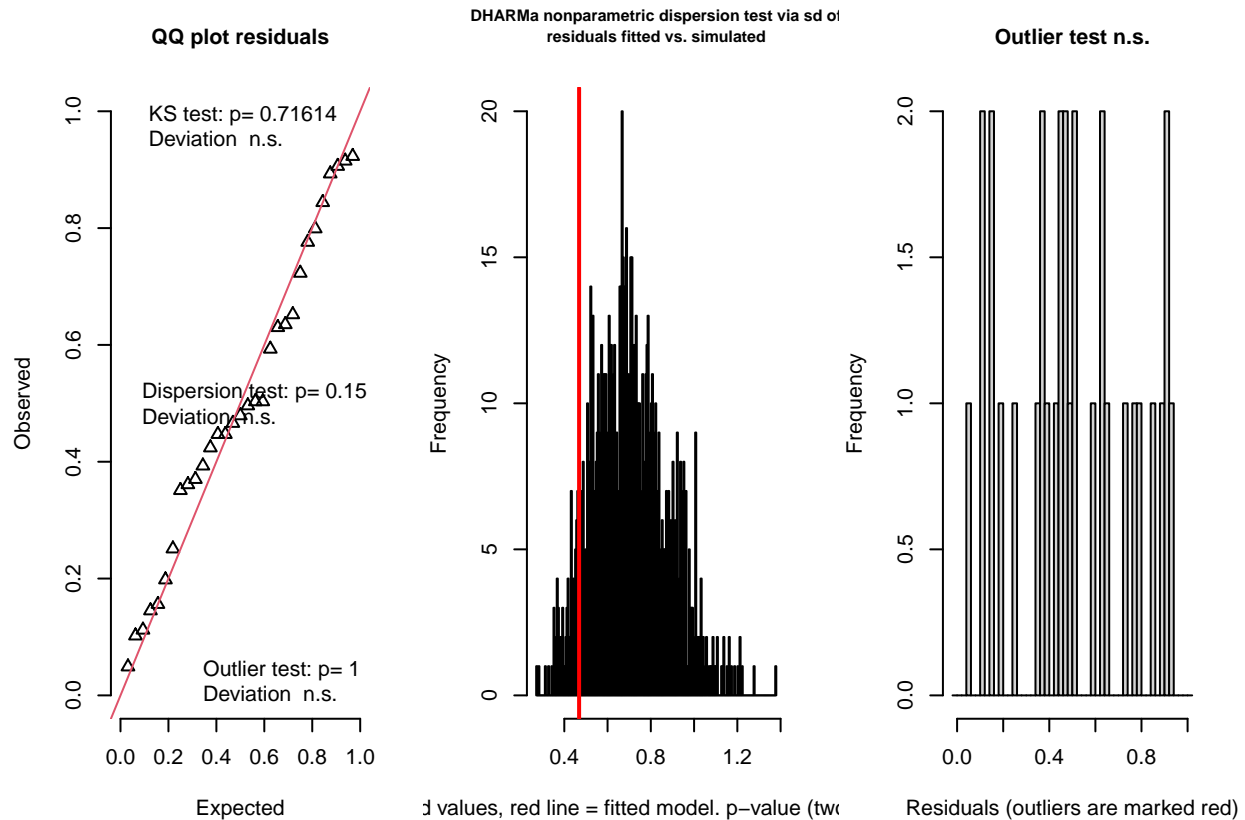
```
##
## Call:
## lm(formula = logistic(Abundance) ~ Treatment + Linage, data = RC_R2nd)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.99379 -0.41802 -0.06997  0.54212  1.71614
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -3.1885     0.6183  -5.157 4.79e-05 ***
## Treatmentremoval -0.9391     0.4530  -2.073  0.0513 .
## LinageB10       0.9697     0.7731   1.254  0.2242
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## LinageB15      1.3776      0.7427      1.855      0.0784 .
## LinageB16     -0.2286      0.8820     -0.259      0.7981
## LinageB20      1.0399      0.8820      1.179      0.2522
## LinageB23      0.4908      1.3061      0.376      0.7110
## LinageB24      0.7313      0.8214      0.890      0.3839
## LinageB36      0.4945      0.8820      0.561      0.5812
## LinageB39     -0.1815      1.3061     -0.139      0.8909
## LinageB7      -0.1675      1.3061     -0.128      0.8992
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.151 on 20 degrees of freedom
## Multiple R-squared:  0.4011, Adjusted R-squared:  0.1016
## F-statistic: 1.339 on 10 and 20 DF, p-value: 0.2765
```

```
Anova(RC_R2nd.mod, type = "II")
```

```
## Anova Table (Type II tests)
##
## Response: logistic(Abundance)
##      Sum Sq Df F value    Pr(>F)
## Treatment  5.6882  1  4.2971 0.05131 .
## Linage     9.4990  9  0.7973 0.62301
## Residuals 26.4747 20
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
res_RC_R2nd <- simulateResiduals(fittedModel=RC_R2nd.mod, n = 1000)
testResiduals(res_RC_R2nd)
```



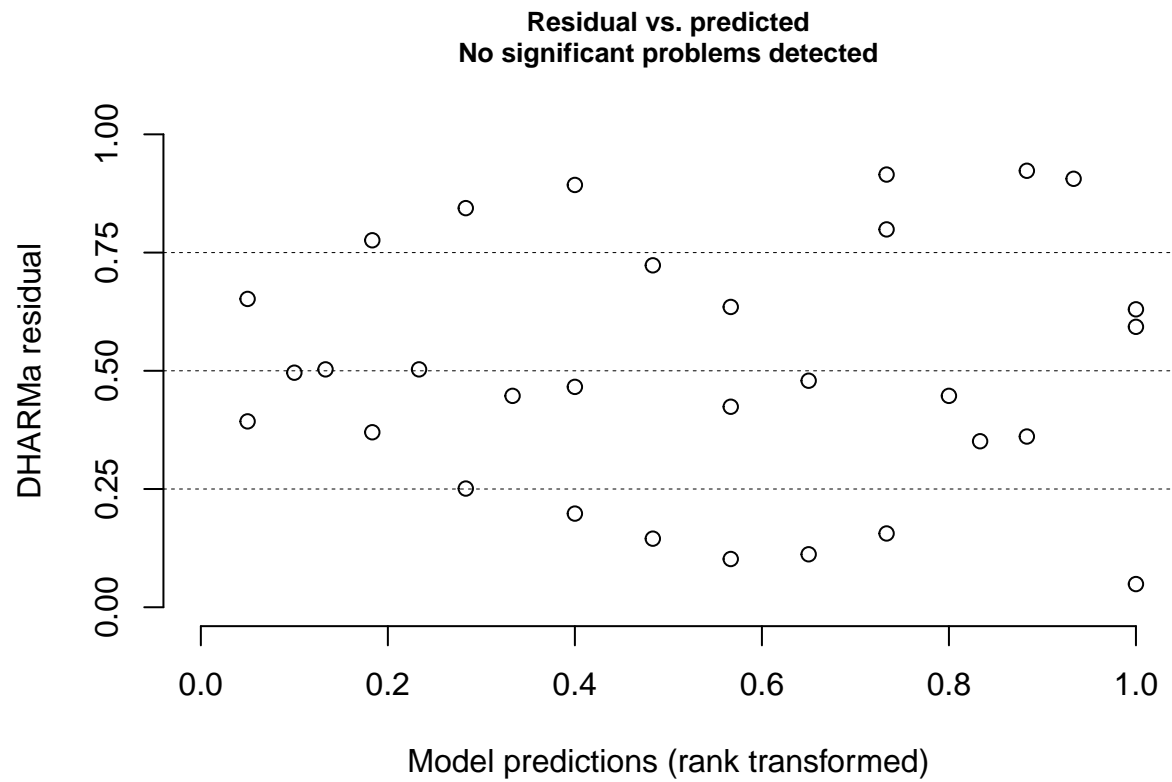
```
## $uniformity
##
## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.12519, p-value = 0.7161
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.66599, p-value = 0.15
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 31, p-value = 1
```

```
## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.0000000 0.1121887
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
## 0
```

```
## $uniformity
##
## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.12519, p-value = 0.7161
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.66599, p-value = 0.15
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 31, p-value = 1
## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.0000000 0.1121887
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
## 0
```

```
plotResiduals(res_RC_R2nd)
```

```
## Unable to calculate quantile regression for quantile 0.25. Possibly to few (unique) data points / pr
## Unable to calculate quantile regression for quantile 0.5. Possibly to few (unique) data points / pre
## Unable to calculate quantile regression for quantile 0.75. Possibly to few (unique) data points / pr
```



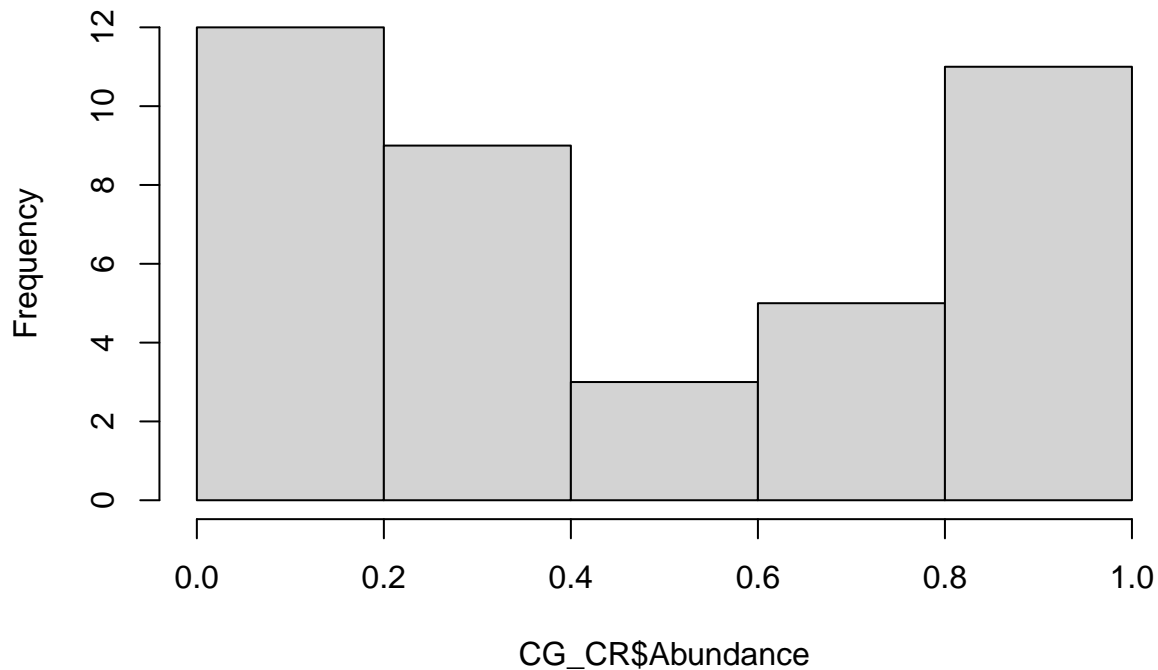
dataset *Chaetomium globosum*

```
CG_CR$Treatment <- as.factor(CG_CR$Treatment)
CG_CR <- within(CG_CR, Treatment <- relevel(Treatment, ref = "control"))

#control vs. removal
hist(CG_CR$Abundance)
```



## Histogram of CG\_CR\$Abundance



```
CG_CR.mod <- lm(logistic(Abundance) ~ Treatment + Linage, data = CG_CR)
summary(CG_CR.mod)
```

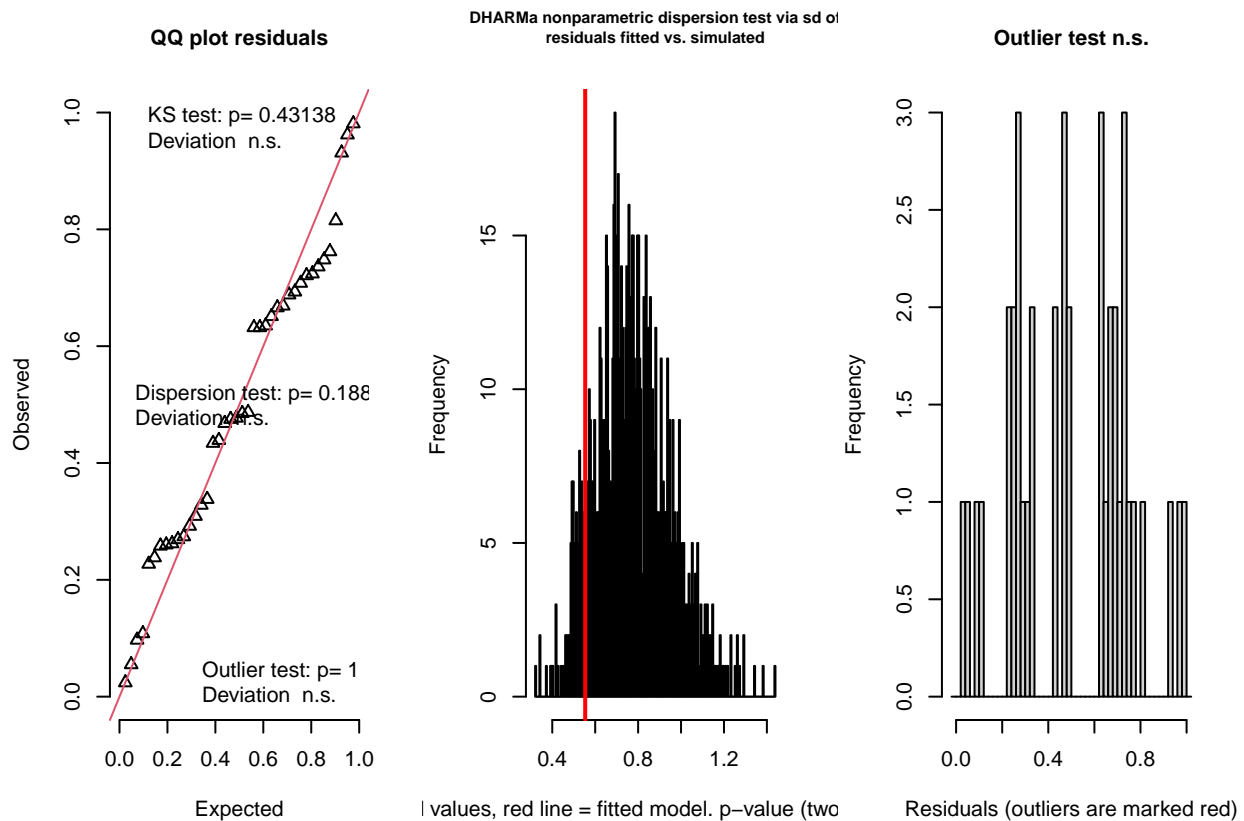
```
##
## Call:
## lm(formula = logistic(Abundance) ~ Treatment + Linage, data = CG_CR)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.9174 -1.8652 -0.0126  1.5953  6.1238
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.9525     1.5687  -0.607   0.5486
## Treatmentremoval  1.8051     0.9834   1.836   0.0771 .
## LinageB10      -0.9916     1.8687  -0.531   0.5999
## LinageB15     -1.5876     1.8687  -0.850   0.4027
## LinageB16       1.3350     2.2814   0.585   0.5631
## LinageB20       0.5364     2.2814   0.235   0.8158
## LinageB23       0.5182     2.5801   0.201   0.8423
## LinageB24       0.6024     2.1210   0.284   0.7785
## LinageB36       0.8342     2.1067   0.396   0.6951
## LinageB39       1.0642     2.2814   0.466   0.6445
## LinageB46       4.8388     3.3670   1.437   0.1618
## LinageB7        2.3669     2.5801   0.917   0.3668
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.979 on 28 degrees of freedom
## Multiple R-squared:  0.2958, Adjusted R-squared:  0.01916
## F-statistic: 1.069 on 11 and 28 DF,  p-value: 0.4186
```

```
Anova(CG_CR.mod, type = "II")
```

```
## Anova Table (Type II tests)
##
## Response: logistic(Abundance)
##           Sum Sq Df F value    Pr(>F)
## Treatment  29.908  1   3.3695 0.07706 .
## Linage     72.260 10   0.8141 0.61777
## Residuals 248.530 28
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
res.CG_CR.mod <- simulateResiduals(CG_CR.mod, n = 1000)
testResiduals(res.CG_CR.mod)
```



```
## $uniformity
##
```

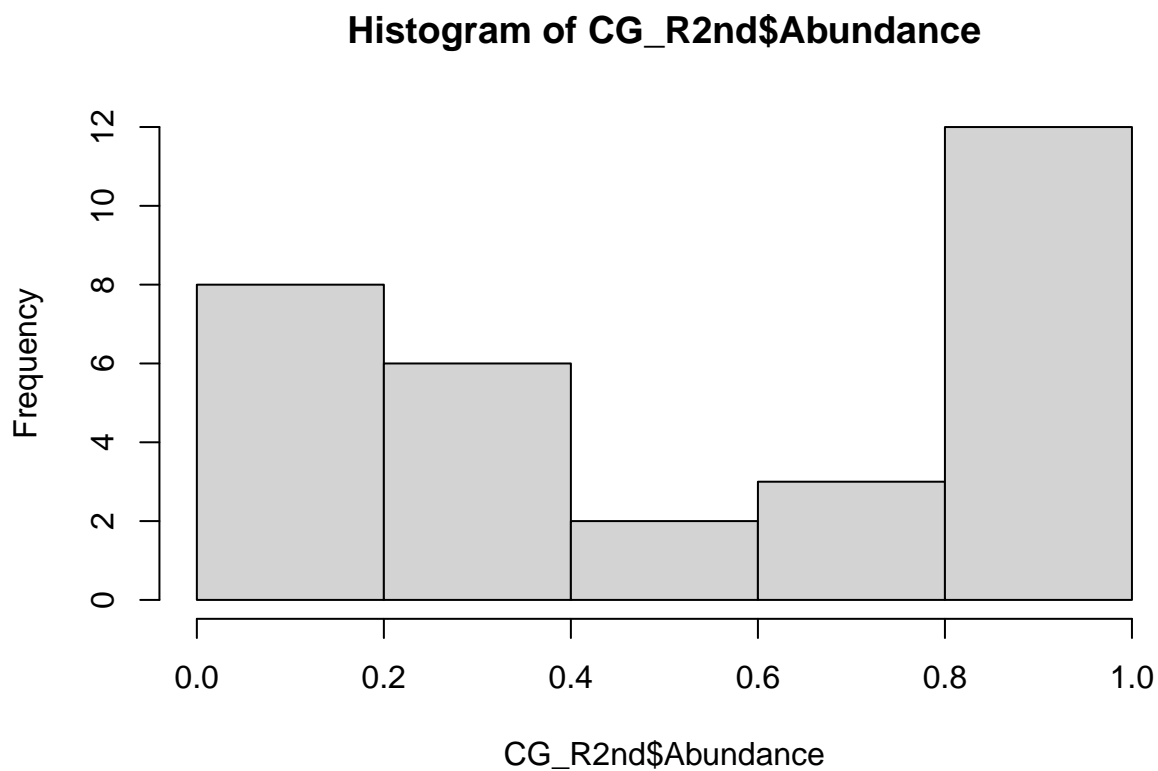
```

## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.138, p-value = 0.4314
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.71597, p-value = 0.188
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 40, p-value = 1
## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.0000000 0.0880973
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
##
##
## $uniformity
##
## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.138, p-value = 0.4314
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.71597, p-value = 0.188
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations

```

```
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 40, p-value = 1
## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.0000000 0.0880973
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
## 0
```

```
#removal vs. 2nd attempt
hist(CG_R2nd$Abundance)
```



```
CG_R2nd.mod <- lm(logistic(Abundance) ~ Treatment + Linage, data = CG_R2nd)
summary(CG_R2nd.mod)
```

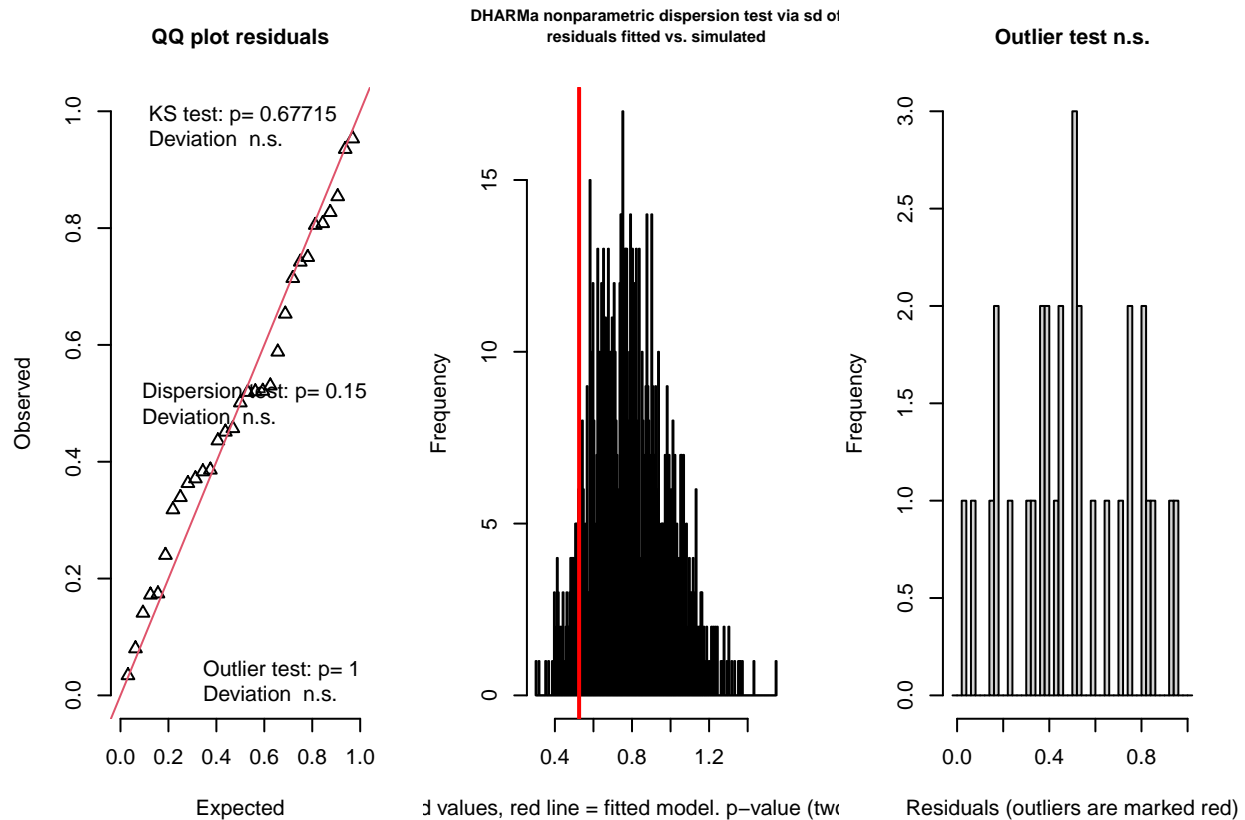
```
##
## Call:
## lm(formula = logistic(Abundance) ~ Treatment + Linage, data = CG_R2nd)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.1054 -1.3363 -0.0285  2.0304  5.6895
##
## Coefficients:
```

```
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)      2.5165      1.7758   1.417  0.1719
## Treatmentremoval  0.3940      1.3012   0.303  0.7652
## LinageB10        -3.3741      2.2207  -1.519  0.1443
## LinageB15        -3.7326      2.1332  -1.750  0.0955 .
## LinageB16         0.1315      2.5333   0.052  0.9591
## LinageB20        -2.8706      2.5333  -1.133  0.2706
## LinageB23        -3.5060      3.7516  -0.935  0.3612
## LinageB24        -1.8909      2.3593  -0.801  0.4323
## LinageB36        -2.3398      2.5333  -0.924  0.3667
## LinageB39         0.7595      3.7516   0.202  0.8416
## LinageB7         2.1701      3.7516   0.578  0.5694
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.305 on 20 degrees of freedom
## Multiple R-squared:  0.2923, Adjusted R-squared:  -0.06151
## F-statistic: 0.8262 on 10 and 20 DF,  p-value: 0.6092
```

```
Anova(CG_R2nd.mod, type = "II")
```

```
## Anova Table (Type II tests)
##
## Response: logistic(Abundance)
##           Sum Sq Df F value Pr(>F)
## Treatment  1.001  1  0.0917 0.7652
## Linage     85.990  9  0.8748 0.5623
## Residuals 218.424 20
```

```
res.CG_R2nd.mod <- simulateResiduals(CG_R2nd.mod, n = 1000)
testResiduals(res.CG_R2nd.mod)
```



```
## $uniformity
##
## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.12445, p-value = 0.6771
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.66599, p-value = 0.15
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 31, p-value = 1
```

```

## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.0000000 0.1121887
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
## 0

## $uniformity
##
## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.12445, p-value = 0.6771
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.66599, p-value = 0.15
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 31, p-value = 1
## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.0000000 0.1121887
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
## 0

```

## testing difference of ‘successfull’ vs. ‘failed’ nests

create subset of samples from ‘removal’ group

```

REM <- subset_samples(rel.R2nd, Treatment == "removal")
REM <- prune_taxa(taxa_sums(REM) > 0, REM)

sample_data(REM)

```

```

##           Sample   Nest Linage Treatment age_sampling_d. Group ColSums
## B0-19a    B0-19a  B0-19    B0    removal           31 Fungi    8655
## B0-22a    B0-22a  B0-22    B0    removal           31 Fungi   19669

```

```
## B10-09a B10-09a B10-09      B10      removal      30 Fungi      26662
## B10-25a B10-25a B10-25      B10      removal      30 Fungi      24345
## B10-26a B10-26a B10-26      B10      removal      30 Fungi      26861
## B15-01a B15-01a B15-01      B15      removal      33 Fungi      12127
## B15-07a B15-07a B15-07      B15      removal      29 Fungi      19916
## B15-25a B15-25a B15-25      B15      removal      28 Fungi       8107
## B16-01a B16-01a B16-01      B16      removal      32 Fungi      20355
## B16-31a B16-31a B16-31      B16      removal      30 Fungi      15800
## B20-04a B20-04a B20-04      B20      removal      35 Fungi      34274
## B20-08a B20-08a B20-08      B20      removal      31 Fungi      27572
## B23-23a B23-23a B23-23      B23      removal      32 Fungi      26875
## B24-02a B24-02a B24-02      B24      removal      32 Fungi       1087
## B24-26a B24-26a B24-26      B24      removal      32 Fungi      16017
## B24-28a B24-28a B24-28      B24      removal      32 Fungi      19882
## B36-14a B36-14a B36-14      B36      removal      19 Fungi       8615
## B36-31a B36-31a B36-31      B36      removal      32 Fungi      17279
## B39-17a B39-17a B39-17      B39      removal      31 Fungi      12233
## B7-08a   B7-08a   B7-08       B7       removal      33 Fungi       4694
```

```
suc = c("successfull", "successfull", "successfull", "successfull", "failed", "successfull", "successfull")
sample_data(REM)$success <- suc

dfREM <- REM %>%
  tax_glom(taxrank = "Species") %>%
  psmelt()
```

```
otu.REM <- abundances(REM)
meta.REM <- meta(REM)

set.seed(1)
adonis2(distance(REM, method = "bray") ~ success, data = meta.REM)
```

## Permanova on data REM

```
## Permutation test for adonis under reduced model
## Terms added sequentially (first to last)
## Permutation: free
## Number of permutations: 999
##
## adonis2(formula = distance(REM, method = "bray") ~ success, data = meta.REM)
##           Df SumOfSqs      R2      F Pr(>F)
## success    1   0.3363 0.09082 1.7981  0.164
## Residual  18   3.3662 0.90918
## Total     19   3.7024 1.00000
```

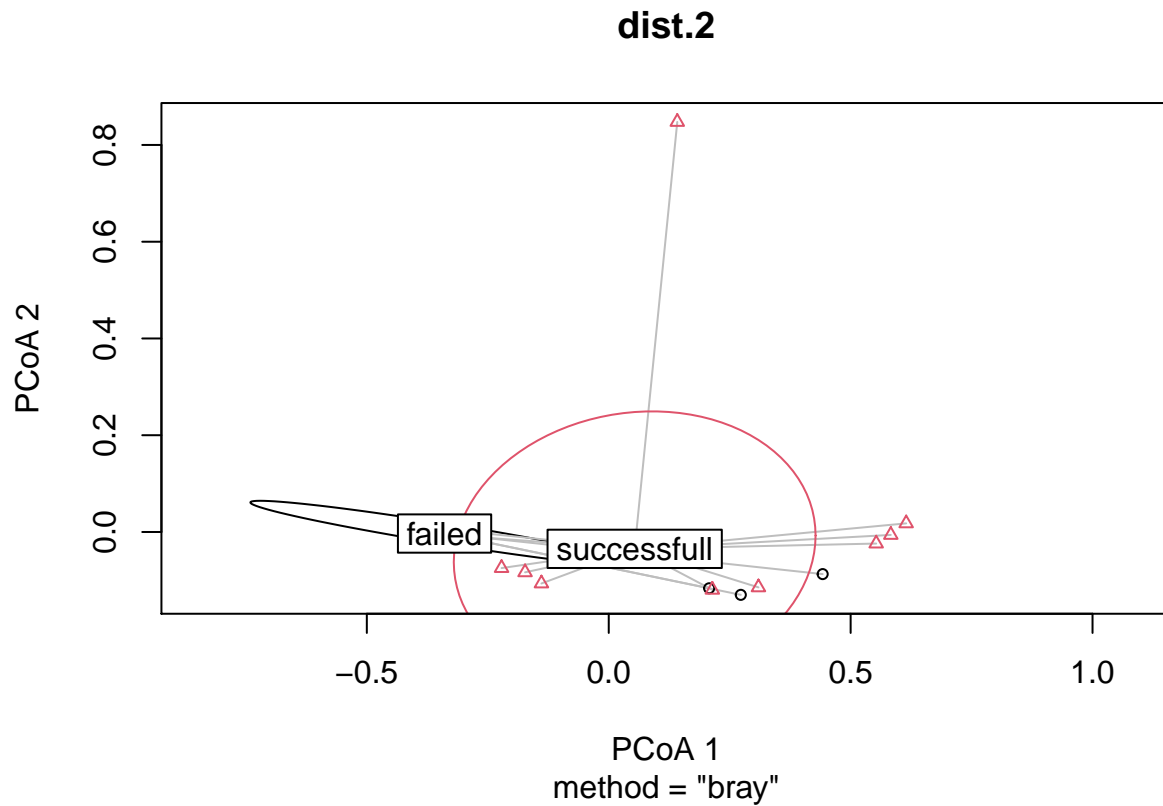
```
dist <- vegdist(t(otu.REM))
dist.2 <- betadisper(dist, meta.REM$success)
anova(dist.2)
```

```
## Analysis of Variance Table
```



```
##
## Response: Distances
##           Df Sum Sq Mean Sq F value Pr(>F)
## Groups    1 0.15332 0.153318   2.109 0.1636
## Residuals 18 1.30858 0.072699
```

```
plot(dist.2, hull = FALSE, ellipse = TRUE)
```



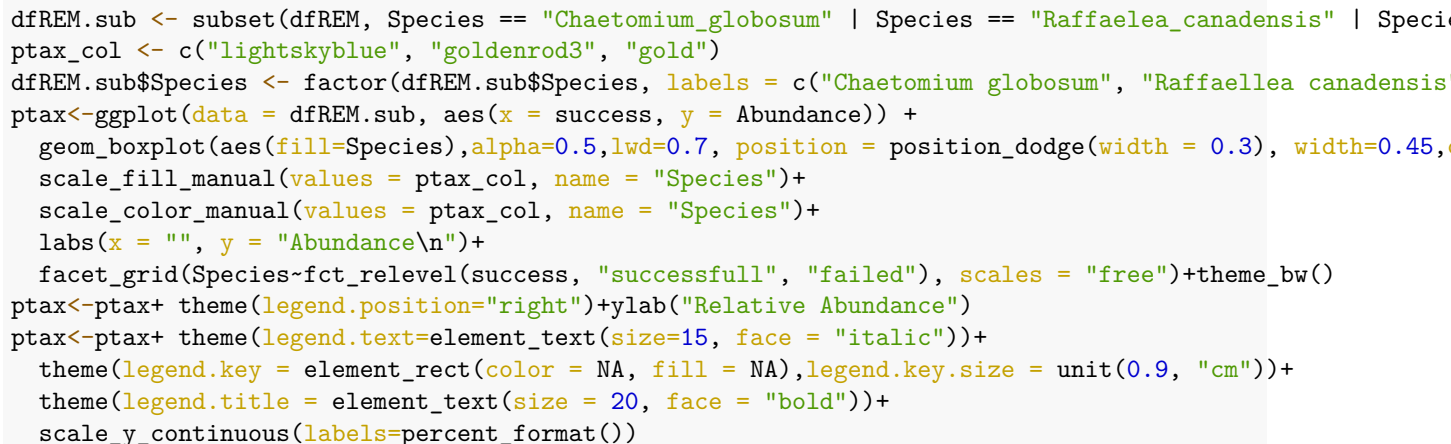
prepare data and plot

```
dfREM.plot <- dfREM
dfREM.plot$Species <- as.character(dfREM.plot$Species)
dfREM.plot$Species[dfREM.plot$Abundance < 0.05] <- "Others"
dfREM.plot$Species <- factor(dfREM.plot$Species, levels = c("Others", "Penicillium_commune", "Acremonium_l"))

Fungi_Success_plot <- ggplot(dfREM.plot, aes(x = Sample, y = Abundance, fill = Species)) +
  geom_bar(stat = "identity", color = NA, position = "fill") +
  scale_fill_manual(values = Plot_colors_s, name = "Species") +
  facet_grid(~success, scales = "free_x", space = "free_x")

FS <- Fungi_Success_plot +
  theme(plot.title = element_text(size = 20, face = "bold")) +
  theme_classic() +
  #gets rid of background
  theme(panel.grid.major = element_line(colour = "grey50")) +
  labs(x = "Treatment", y = "Relative Abundance") +
```

FS

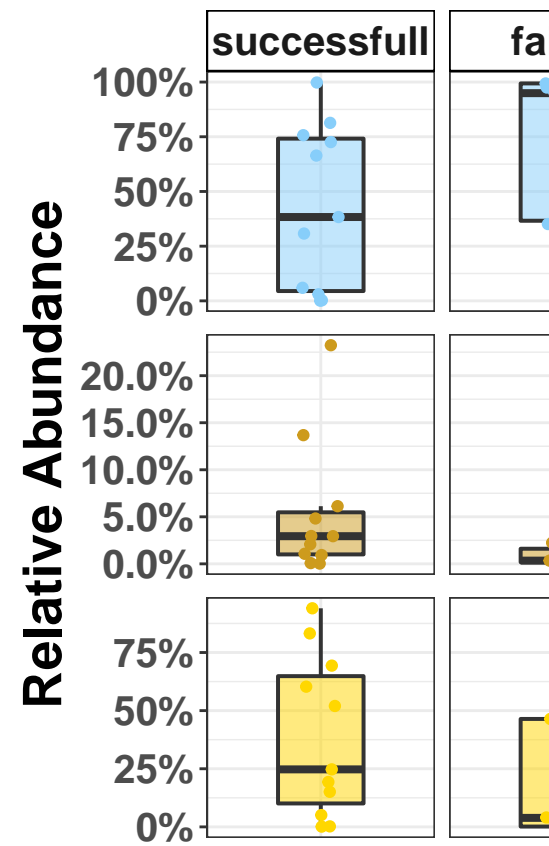


```

abu2<-ptax + theme(strip.background =element_rect(fill="white", color="black"))+
  theme(strip.text.x = element_text(size = 15, face = "bold"))
Success.plot<-abu2+theme(axis.title.y = element_text(size=20, face="bold"))+theme(axis.text.y = element.
  theme(axis.text.x = element_text(size=20, angle=90,face="bold"))+
  theme(axis.title.x = element_blank(), axis.text.x=element_blank(),
    axis.ticks.x=element_blank())+
  theme(strip.text.y = element_blank())
Success.plot_fun <- Success.plot + theme(panel.spacing.y = unit(0.3, "cm"))

Success.plot_fun

```



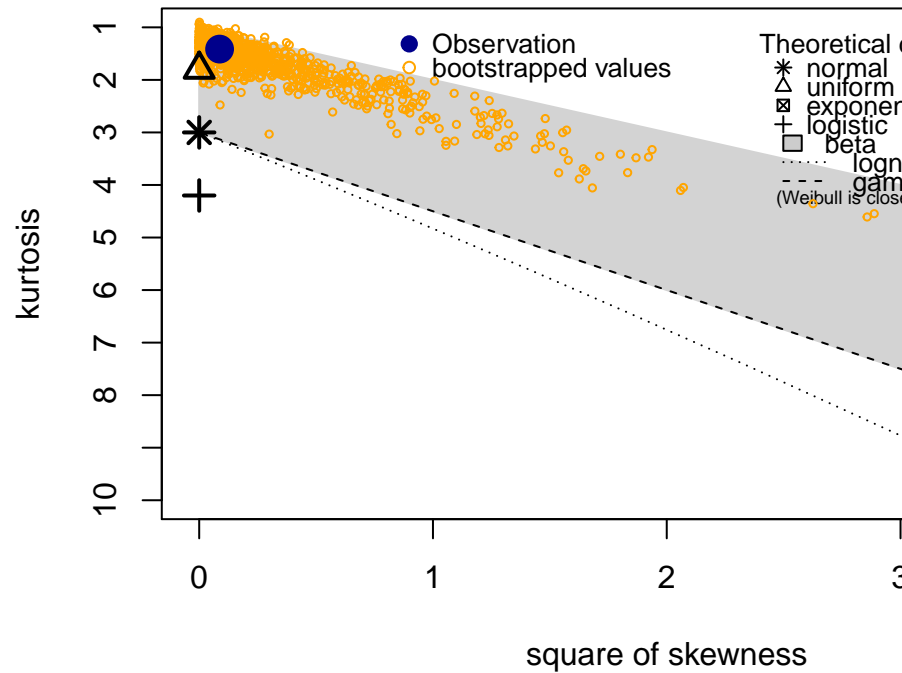
plot core taxa of successful vs. failed nests with relative abundance

```

CGrem <- subset(dfREM, Species == "Chaetomium_globosum")
descdist(CGrem$Abundance, boot = 1000)

```

## Cullen and Frey graph



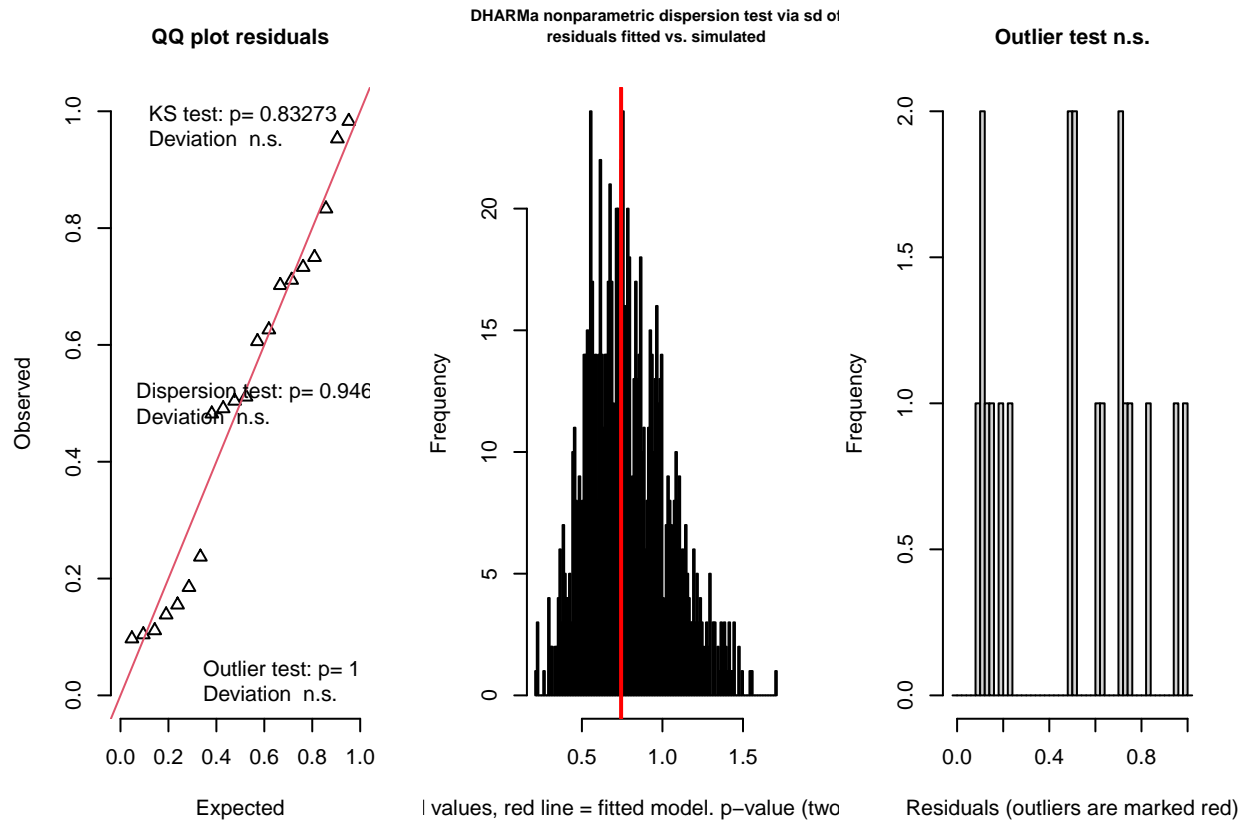
test if taxa different in these two groups

```
## summary statistics
## -----
## min: 0.0007979109 max: 0.9997544
## median: 0.6949201
## mean: 0.5751256
## estimated sd: 0.3863477
## estimated skewness: -0.2953997
## estimated kurtosis: 1.414034
```

```
CGrem.mod <- lm(logistic(Abundance) ~ success, data = CGrem)
Anova(CGrem.mod, type = "II")
```

```
## Anova Table (Type II tests)
##
## Response: logistic(Abundance)
##          Sum Sq Df F value    Pr(>F)
## success    56.252  1  5.5052 0.03061 *
## Residuals 183.925 18
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
res_CGrem.mod <- simulateResiduals(CGrem.mod, n = 1000)
testResiduals(res_CGrem.mod)
```



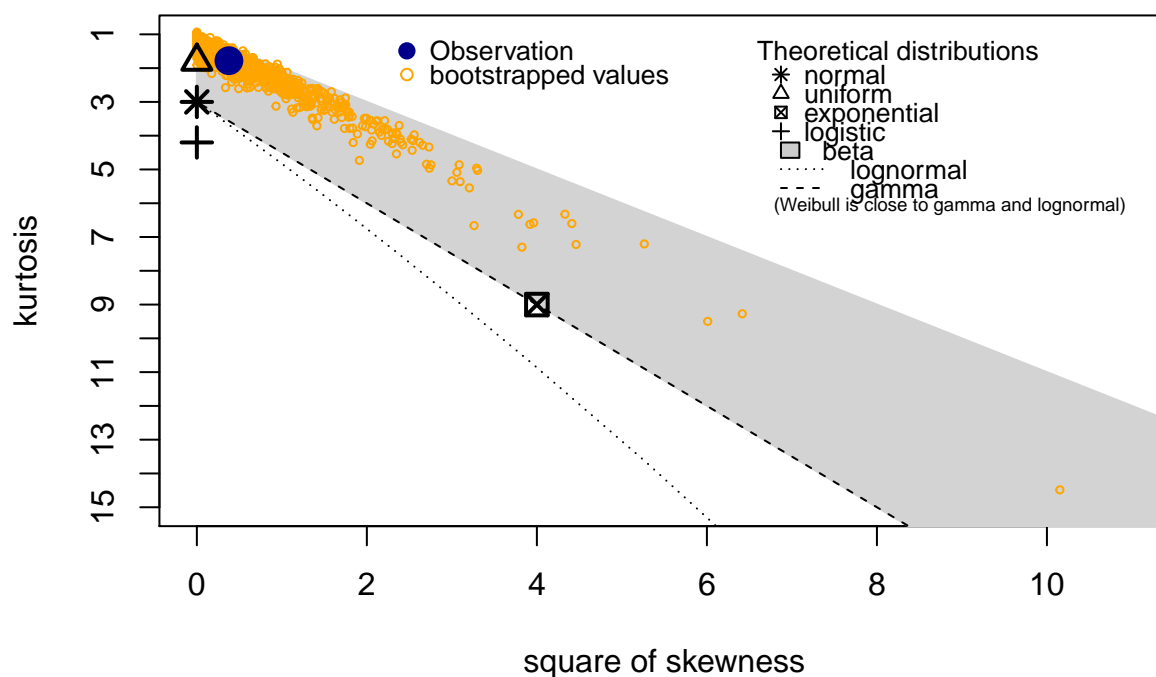
```
## $uniformity
##
## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.132, p-value = 0.8327
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.94547, p-value = 0.946
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 20, p-value = 1
```

```
## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.0000000 0.1684335
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
## 0
```

```
## $uniformity
##
## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.132, p-value = 0.8327
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.94547, p-value = 0.946
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 20, p-value = 1
## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.0000000 0.1684335
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
## 0
```

```
RSrem <- subset(dfREM, Species == "Raffaelea_sulphurea")
descdist(RSrem$Abundance, boot = 1000)
```

## Cullen and Frey graph

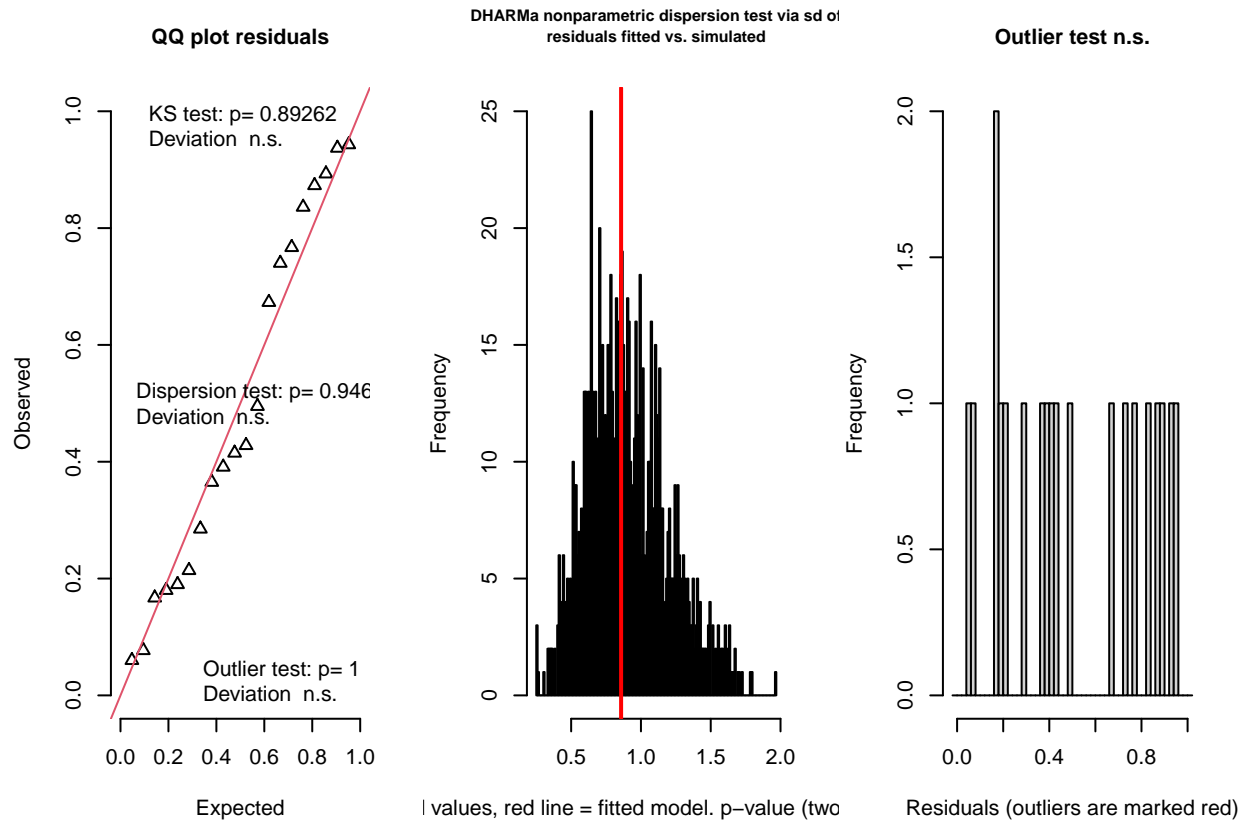


```
## summary statistics
## -----
## min: 0.0001473839   max: 0.9401502
## median: 0.1719765
## mean: 0.307603
## estimated sd: 0.3306086
## estimated skewness: 0.6144493
## estimated kurtosis: 1.77946
```

```
RSrem.mod <- lm(logistic(Abundance) ~ success, data = RSrem)
Anova(RSrem.mod, type = "II")
```

```
## Anova Table (Type II tests)
##
## Response: logistic(Abundance)
##           Sum Sq Df F value Pr(>F)
## success   11.281  1  2.1023 0.1643
## Residuals  96.585 18
```

```
res_RSrem.mod <- simulateResiduals(RSrem.mod, n = 1000)
testResiduals(res_RSrem.mod)
```



```
## $uniformity
##
## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.122, p-value = 0.8926
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.94547, p-value = 0.946
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 20, p-value = 1
```



```

## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.0000000 0.1684335
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
## 0

## $uniformity
##
## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.122, p-value = 0.8926
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.94547, p-value = 0.946
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 20, p-value = 1
## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.0000000 0.1684335
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
## 0

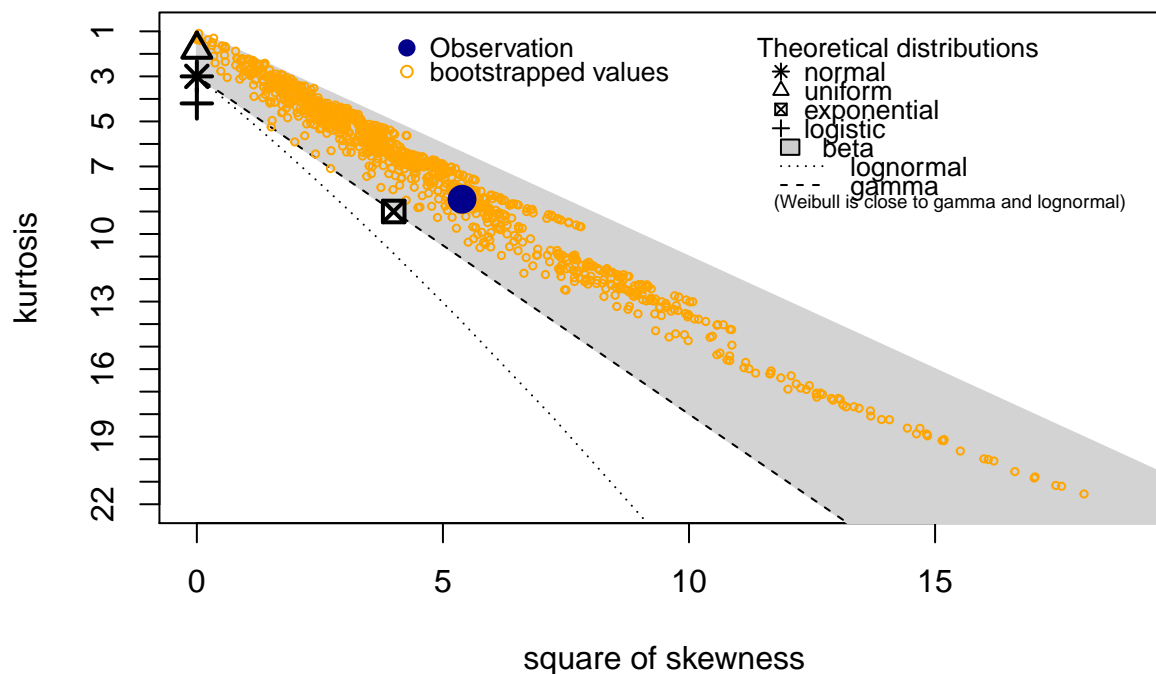
```

```

RCrem <- subset(dfREM, Species == "Raffaelea_canadensis")
descdist(RCrem$Abundance, boot = 1000)

```

## Cullen and Frey graph

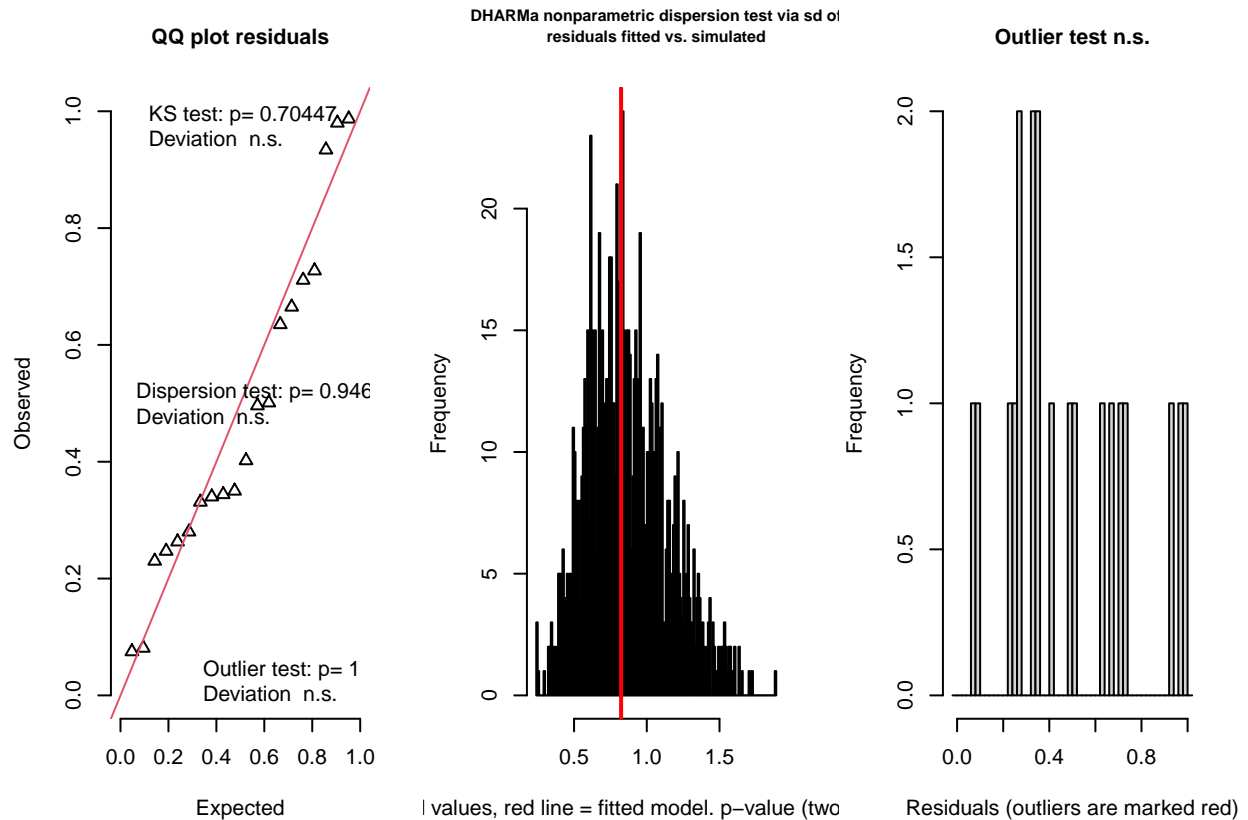


```
## summary statistics
## -----
## min: 9.825596e-05 max: 0.2323918
## median: 0.01338467
## mean: 0.03755341
## estimated sd: 0.05985818
## estimated skewness: 2.321288
## estimated kurtosis: 8.455446
```

```
RCrem.mod <- lm(logistic(Abundance) ~ success, data = RCrem)
Anova(RCrem.mod, type = "II")
```

```
## Anova Table (Type II tests)
##
## Response: logistic(Abundance)
##      Sum Sq Df F value Pr(>F)
## success  2.7337 1  2.8548 0.1083
## Residuals 17.2368 18
```

```
res_RCrem.mod <- simulateResiduals(RCrem.mod, n = 1000)
testResiduals(res_RCrem.mod)
```



```
## $uniformity
##
## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.15, p-value = 0.7045
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.94547, p-value = 0.946
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 20, p-value = 1
```

```

## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.0000000 0.1684335
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
## 0

## $uniformity
##
## One-sample Kolmogorov-Smirnov test
##
## data: simulationOutput$scaledResiduals
## D = 0.15, p-value = 0.7045
## alternative hypothesis: two-sided
##
##
## $dispersion
##
## DHARMA nonparametric dispersion test via sd of residuals fitted vs.
## simulated
##
## data: simulationOutput
## dispersion = 0.94547, p-value = 0.946
## alternative hypothesis: two.sided
##
##
## $outliers
##
## DHARMA outlier test based on exact binomial test with approximate
## expectations
##
## data: simulationOutput
## outliers at both margin(s) = 0, observations = 20, p-value = 1
## alternative hypothesis: true probability of success is not equal to 0.001998002
## 95 percent confidence interval:
## 0.0000000 0.1684335
## sample estimates:
## frequency of outliers (expected: 0.001998001998002 )
## 0

```

## References

- Davis NM, Proctor DiM, Holmes SP, Relman DA, Callahan BJ. 2018. Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. *Microbiome*. 6(1). doi:10.1186/S40168-018-0605-2.
- Delignette-Muller ML, Dutang C. 2015. fitdistrplus: An R package for fitting distributions. *J Stat Softw*. 64(4):1–34. doi:10.18637/jss.v064.i04.
- Dimitris Rizopoulos (2022). GLMMadaptive: Generalized Linear Mixed Models using Adaptive Gaussian Quadrature. R package version 0.8-5. <https://CRAN.R-project.org/package=GLMMadaptive>
- Douglas Bates, Martin Maechler, Ben Bolker, Steve Walker (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1-48. doi:10.18637/jss.v067.i01.

John Fox and Sanford Weisberg (2019). An {R} Companion to Applied Regression, Third Edition. Thousand Oaks CA: Sage. URL: <https://socialsciences.mcmaster.ca/jfox/Books/Companion/>

Gavin L. Simpson (2022). permute: Functions for Generating Restricted Permutations of Data. R package version 0.9-7. <https://CRAN.R-project.org/package=permute>

Hartig F. 2021. DHARMA: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. <https://cran.r-project.org/package=DHARMA>.

Højsgaard, Søren and Halekoh, Ulrich (2022). doBy: Groupwise Statistics, LSmeans, Linear Estimates, Utilities. R package version 4.6.12. <https://CRAN.R-project.org/package=doBy>

Kassambara, Alboukadel (2020). ggpubr: ‘ggplot2’ Based Publication Ready Plots. R package version 0.4.0. <https://CRAN.R-project.org/package=ggpubr>

Kuznetsova A, Brockhoff PB, Christensen RHB (2017). “lmerTest Package: Tests in Linear Mixed Effects Models.” *Journal of Statistical Software*, 82(13), 1-26. doi: 10.18637/jss.v082.i13 (URL: <https://doi.org/10.18637/jss.v082.i13>).

Lahti, Leo et al. microbiome R package. URL: <http://microbiome.github.io>

Lüdtke D (2018). “ggeffects: Tidy Data Frames of Marginal Effects from Regression Models.” *Journal of Open Source Software*, 3(26), 772. doi: 10.21105/joss.00772 (URL: <https://doi.org/10.21105/joss.00772>).

Lüdtke D (2021). *sjPlot: Data Visualization for Statistics in Social Science*. R package version 2.8.10, <URL: <https://CRAN.R-project.org/package=sjPlot>>.

Lüdtke D, Ben-Shachar MS, Patil I, Waggoner P, Makowski D. (2021) performance: An R Package for Assessment, Comparison and Testing of Statistical Models. *Journal of Open Source Software*, 6(60), 3139. doi:10.21105/joss.03139.

Lüdtke et al., (2021). see: An R Package for Visualizing Statistical Models. *Journal of Open Source Software*, 6(64), 3393. <https://doi.org/10.21105/joss.03393>

Mangiafico S. 2021. rcompanion: Functions to Support Extension Education Program Evaluation. <https://cran.r-project.org/package=rcompanion>.

Martinez Arbizu, Pedro (2017). pairwiseAdonis: Pairwise Multilevel Comparison using Adonis. R package version 0.4.

McMurdie PJ, Holmes S. 2013. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One*. 8(4):e61217. doi:10.1371/journal.pone.0061217.

Mollie E. Brooks, Kasper Kristensen, Koen J. van Benthem, Arni Magnusson, Casper W. Berg, Anders Nielsen, Hans J. Skaug, Martin Maechler and Benjamin M. Bolker (2017). glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, 9(2), 378-400.

Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlinn D, Minchin PR, O’Hara RB, Simpson GL, Solymos P, et al. 2020a. vegan: Community Ecology Package. <https://cran.r-project.org/package=vegan>.

Pedersen, Thomas Lin (2020). patchwork: The Composer of Plots. R package version 1.1.1. <https://CRAN.R-project.org/package=patchwork>

R Core Team. 2021. R: A language and environment for statistical computing. <https://www.r-project.org/>.

Russell V. Lenth (2022). emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.7.2. <https://CRAN.R-project.org/package=emmeans>

Sarkar, Deepayan (2008) *Lattice: Multivariate Data Visualization with R*. Springer, New York. ISBN 978-0-387-75968-5

Shetty, Sudarshan and Lahti, Leo (2021). microbiomeutilities: Utilities for Microbiome Analytics. R package version 1.00.16.

Slowikowski, Kamil (2021). `ggrepel`: Automatically Position Non-Overlapping Text Labels with ‘`ggplot2`’. R package version 0.9.1. <https://CRAN.R-project.org/package=ggrepel>

Wickham H. 2016. `ggplot2`: Elegant Graphics for Data Analysis. New York: Springer-Verlag New York.

Wickham et al., (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686, <https://doi.org/10.21105/joss.01686>

Wickham H, Seidel D. 2020. `scales`: Scale Functions for Visualization. <https://cran.r-project.org/package=scales>.

Wickham H, François R, Henry L, Müller K. 2021. `dplyr`: A Grammar of Data Manipulation. <https://cran.r-project.org/package=dplyr>.

Hadley Wickham (2021). `forcats`: Tools for Working with Categorical Variables (Factors). R package version 0.5.1. <https://CRAN.R-project.org/package=forcats>

Wilke, Claus O. (2020). `cowplot`: Streamlined Plot Theme and Plot Annotations for ‘`ggplot2`’. R package version 1.1.1. <https://CRAN.R-project.org/package=cowplot>

Yihui Xie (2021). `knitr`: A General-Purpose Package for Dynamic Report Generation in R. R package version 1.37.