

# W1. Phenotype data

Bora Kim

March 4, 2021

This is meant for sharing dataset and script of the publication “Strigolactone\_structural specificity in microbiome recruitment in rice, 2021”. This markdown contains the process of phenotypic data: Biomass and levels of strigolactones (SLs) from 16 genotypes of rice grown two natural soils for 31 days.

## 1. Getting started

Load required dataset (samples names on row, phenotypic observation on column).

```
load("W1_phenotype_data_image.Rdata")
meta[1:5,1:5]
```

##	Genotype	Soil	Compartment	Soil_compartment	Replicate
## R_A1_e	IAC165	Field	Root	Fi_RT	1
## R_A1_r	IAC165	Field	Rhizosphere	Fi_RS	1
## R_A10_e	IAC165	Forest	Root	Fo_RT	5
## R_A10_r	IAC165	Forest	Rhizosphere	Fo_RS	5
## R_A2_e	IAC165	Field	Root	Fi_RT	2

Load required packages for analysis

```
library(ggplot2)
library(FSA)
library(rcompanion)
library(dplyr)
library(ggfortify)
library(tibble)
library(reshape2)
```

## 2. Comparing biomass (Fig S1)

To handle data easier, grouping information and biomass were subsetting from initial meta data

```
bm<-subset(meta[,c(8,1:4)],Compartment=="Root")
bm$Genotype2<-factor(bm$Genotype,
                      c("IAC165", "IAC1246", "GWD", "Dullo", "Bina", "Sonk", "Kinko",
                        "TN1", "Bhas", "SC", "Shiokari", "d14", "d3", "d10", "d17", "d27")) #change order of gen
```

### 2.1. Kruskal-Wallis test

Test the biomass by soil type and genotype in each soil using non-parametric Kruskal-Wallis test together with P value correction using Benjamini-Hochberg method.

```
bm_kw<-kruskal.test(bm$total_fresh_biomass_g_per_plant~bm$Soil, data=bm)
bm_kw
```

```
##
## Kruskal-Wallis rank sum test
##
## data: bm$total_fresh_biomass_g_per_plant by bm$Soil
## Kruskal-Wallis chi-squared = 103.27, df = 1, p-value < 2.2e-16
```

Create a result table combining all results

```
bm_kw_res<-as.data.frame(matrix(NA, 1, 4))

bm_kw_res[,1]<-bm_kw$statistic
bm_kw_res[,2]<-bm_kw$parameter
bm_kw_res[,3]<-bm_kw$p.value
bm_kw_res[,4]<-p.adjust(bm_kw$p.value,method="BH" )

colnames(bm_kw_res)<-c("chi_squared", "Df", "P value", "adjusted P value")
rownames(bm_kw_res)<-c("Total biomass_by_soil")

bm_kw_res
```

```
##                chi_squared Df      P value adjusted P value
## Total biomass_by_soil    103.2744  1 2.91797e-24      2.91797e-24
```

### 2.3. Duun test

Compare biomass among genotypes in each soil, therefore subset the biomass from each soil

```
fobm_dunn<-subset(bm, Soil=="Forest")
PT<-dunnTest(fobm_dunn$total_fresh_biomass_g_per_plant~Genotype, data=fobm_dunn, method = "bh")
fobm_dunn_res<-PT$res
fobm_dunn_res_letter<-cldList(comparison = fobm_dunn_res$Comparison,p.value = fobm_dunn_res$P.adj,thresh=0.05)
fobm_dunn_res[1:5,1:4]
```

```
##      Comparison      Z      P.unadj      P.adj
## 1 Bhas - Bina -1.88095303 0.05997831 0.13841148
## 2 Bhas - d10  0.24803776 0.80410519 0.92781368
## 3 Bina - d10  2.12899079 0.03325502 0.08675223
## 4 Bhas - d14  0.06889938 0.94506971 0.96930227
## 5 Bina - d14  1.94985241 0.05119371 0.12286491
```

```
fobm_dunn_res_letter
```

```
##      Group Letter MonoLetter
## 1      Bhas      ab      ab
## 2      Bina  abcde      abcde
## 3        d1       a       a
## 4       d14      ab      ab
## 5       d17      abc      abc
## 6       d27      ab      ab
## 7        d3       a       a
## 8      Dullo  abcde      abcde
## 9       GWD  abcd      abcd
## 10 IAC1246      e       e
## 11 IAC165      e       e
## 12 Kinko     cde      cde
## 13      SC     cde      cde
## 14 Shiokari    de      de
```

```
## 15      Sonk      bcde      bcde
## 16      TN1      abcde      abcde

fibm_dunn<-subset(bm, Soil=="Forest")
PT<-dunnTest(fibm_dunn$total_fresh_biomass_g_per_plant~Genotype, data=fibm_dunn, method = "bh")
fibm_dunn_res<-PT$res
fibm_dunn_res_letter<-cldList(comparison = fibm_dunn_res$Comparison,p.value = fibm_dunn_res$P.adj,thresh)
fibm_dunn_res[1:5,1:4]
```

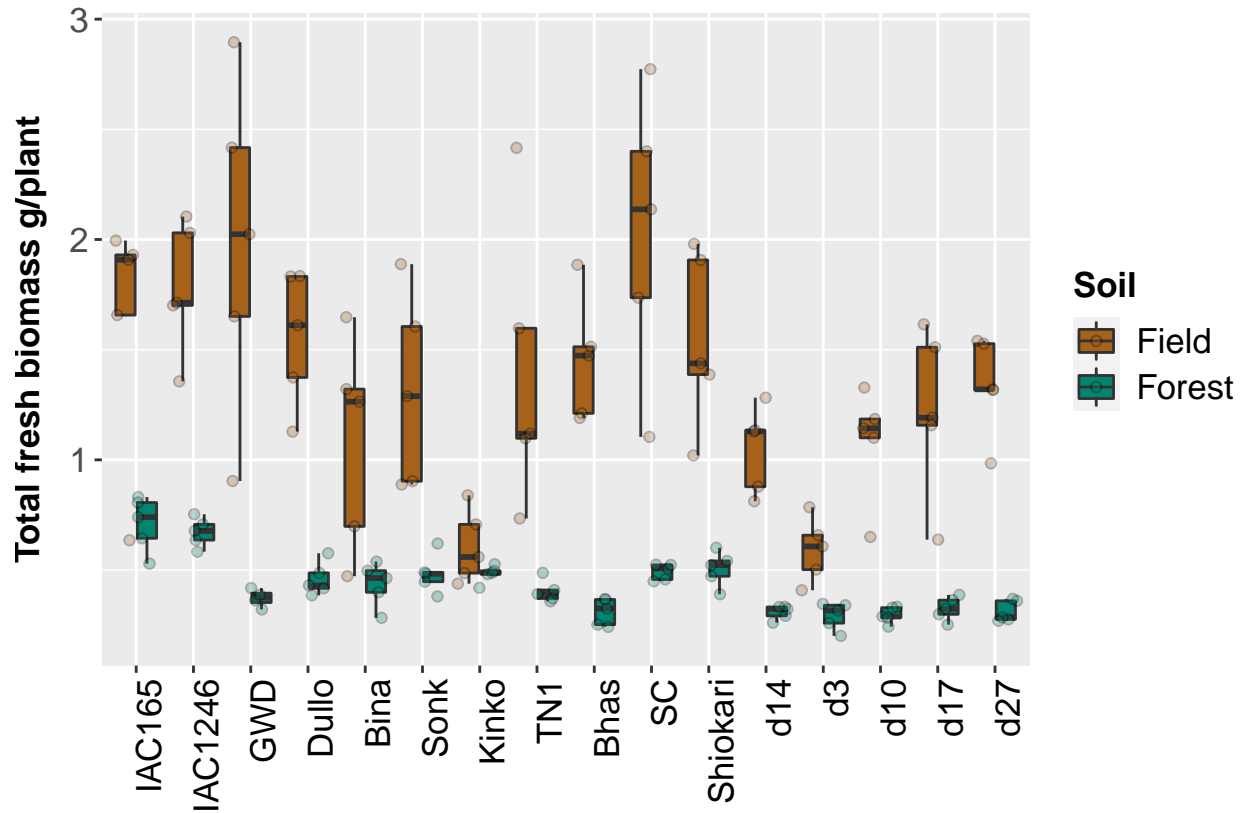
##	Comparison	Z	P.unadj	P.adj
## 1	Bhas - Bina	-1.88095303	0.05997831	0.13841148
## 2	Bhas - d10	0.24803776	0.80410519	0.92781368
## 3	Bina - d10	2.12899079	0.03325502	0.08675223
## 4	Bhas - d14	0.06889938	0.94506971	0.96930227
## 5	Bina - d14	1.94985241	0.05119371	0.12286491

```
fibm_dunn_res_letter
```

##	Group	Letter	MonoLetter
## 1	Bhas	ab	ab
## 2	Bina	abcde	abcde
## 3	d1	a	a
## 4	d14	ab	ab
## 5	d17	abc	abc
## 6	d27	ab	ab
## 7	d3	a	a
## 8	Dullo	abcde	abcde
## 9	GWD	abcd	abcd
## 10	IAC1246	e	e
## 11	IAC165	e	e
## 12	Kinko	cde	cde
## 13	SC	cde	cde
## 14	Shiokari	de	de
## 15	Sonk	bcde	bcde
## 16	TN1	abcde	abcde

## 2.4. Plot biomass

```
ggplot(bm) +
  geom_boxplot(aes(x=Genotype2, y=total_fresh_biomass_g_per_plant, fill=Soil), outlier.colour = NA)+
  labs(x="", y = "Total fresh biomass g/plant") + scale_fill_manual(values=c("#a6611a","#018571"))+
  geom_point(aes(x=Genotype2, y=total_fresh_biomass_g_per_plant, fill=Soil), alpha = 0.3, shape = 21,
             position = position_jitterdodge())+
  theme(axis.text.x = element_text(size = 13, colour = "black",angle=90, hjust=1),
        axis.text.y = element_text(size = 13),
        axis.title.y = element_text(face = "bold", size = 13, vjust = 3),
        legend.text = element_text(size = 13), legend.title = element_text(size = 13,face = "bold"))
```



### 3. Comparing strigolactone level (figure 1)

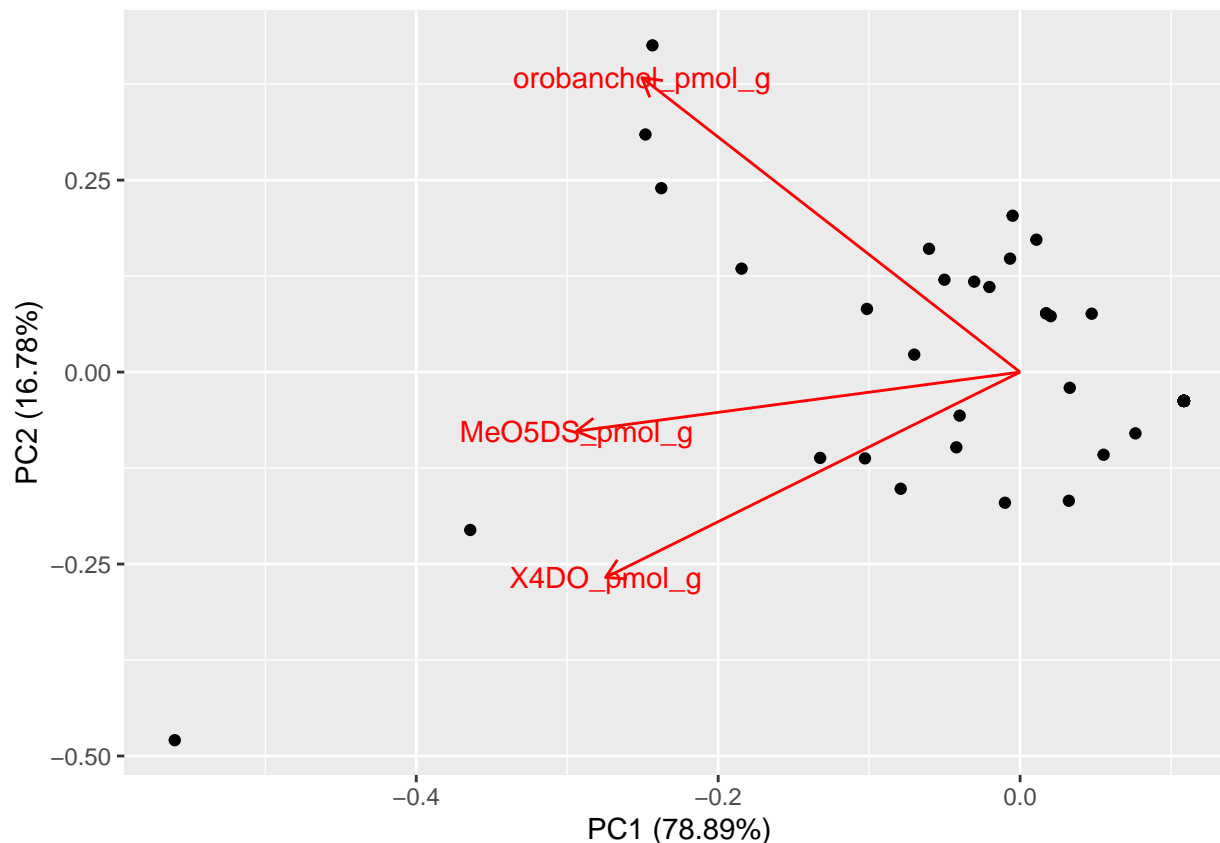
As SLs were not detected in field soil, further SL-related studies are only samples obtained from forest soil

```
foSLs<-subset(meta, Soil_compartment=="Fo_RS"&SL_analysis=="yes")
```

#### 3.1. Principle component analysis on three endogenous SLs

```
foSLs<-subset(meta, Soil_compartment=="Fo_RS"&SL_analysis=="yes")
foSLs<-subset(meta, Soil_compartment=="Fo_RS"&SL_analysis=="yes")
foSLs_pca<-prcomp(foSLs[11:13], scale. = T)
```

```
autoplot(foSLs_pca, data = foSLs_pca, loadings = TRUE,
         loadings.label = TRUE, loadings.label.size = 4)
```



### 3.2. Dunn test

First, remove some genotypes due to zeros in all replicates (non-detectable) which could make problem during Dunn test.

```
foSLs_dunn<-foSLs %>% filter(!Genotype%in%c('Kinko','TN1','Bhas','SC','d10','d17','d27'))
foSLs_dunn<-foSLs_dunn[,c(11:13,1)]
foSLs_dunn[1:5,1:4]
```

	orobanchol_pmol_g	X4DO_pmol_g	MeO5DS_pmol_g	Genotype
## R_A7_r	0.1201817	5.554455	0.9425109	IAC165
## R_A8_r	0.2420453	9.055589	16.2683221	IAC165
## R_A9_r	0.3023398	5.792403	15.6206970	IAC165
## R_B7_r	0.1252308	0.000000	4.7847343	IAC1246
## R_B8_r	0.2688506	8.246183	16.3267069	IAC1246

Perform Dunn test using loop. As level of MeO5DS was not significantly different among genotype, it was excluded.

```
indices=2
Z<-as.data.frame(matrix(NA, 36, indices)) #results list =36
P.unadj<-as.data.frame(matrix(NA, 36, indices)) #results list =36
P.adj<-as.data.frame(matrix(NA, 36, indices)) #results list =36
foSLs_dunn_letter<-as.data.frame(matrix(NA, 9, indices)) #results list =9

for(i in 1:indices) {
  PT<-dunnTest(foSLs_dunn[,i]~Genotype, data=foSLs_dunn, method = "bh")
  Z[,i]<-PT$res$Z
}
```

```

P.unadj[,i]<-PT$res$P.unadj
P.adj[,i]<-PT$res$P.adj
PT2<-PT$res
cl<-cldList(comparison = PT2$Comparison, p.value = PT2$P.adj,threshold = 0.05)
foSLs_dunn_letter[,i]<-cl$Letter
}

```

```

foSLs_dunn_res<-cbind(Z,P.unadj, P.adj)
rownames(foSLs_dunn_res)<- PT$res$Comparison
colnames(foSLs_dunn_res)<-c("Z_orobanchol","Z_4D0",
                           "unadjusted.P_orobanchol","unadjusted.P_4D0",
                           "adjusted.P_orobanchol","adjusted.P_4D0")
rownames(foSLs_dunn_letter) <- cl$Group
colnames(foSLs_dunn_letter) <- colnames(foSLs_dunn[1:indices])

```

```
foSLs_dunn_res[1:5, 1:6]
```

```

##           Z_orobanchol      Z_4D0 unadjusted.P_orobanchol unadjusted.P_4D0
## Bina - d14      -0.43210273 -2.44556403             0.66566675      0.014462576
## Bina - d3       1.46434814 -1.24675813             0.14309885      0.212486195
## d14 - d3       1.89645087  1.19880590             0.05790045      0.230603428
## Bina - Dullo   -0.07514975 -0.09650201             0.94009556      0.923121875
## d14 - Dullo    0.40795579  2.63772170             0.68330613      0.008346505
##           adjusted.P_orobanchol adjusted.P_4D0
## Bina - d14           0.7988001      0.1041306
## Bina - d3            0.3030329      0.4249724
## d14 - d3            0.2084416      0.4150862
## Bina - Dullo        0.9400956      0.9774232
## d14 - Dullo        0.7935168      0.1001581

```

```
foSLs_dunn_letter
```

```

##           orobanchol_pmol_g X4D0_pmol_g
## Bina                ab             ab
## d14                 ab             a
## d3                  ab             ab
## Dullo               ab             ab
## GWD                  a             ab
## IAC1246              b             ab
## IAC165               b             ab
## Shiokari            ab             a
## Sonk                ab             b

```

### 3.3. Plot SLs

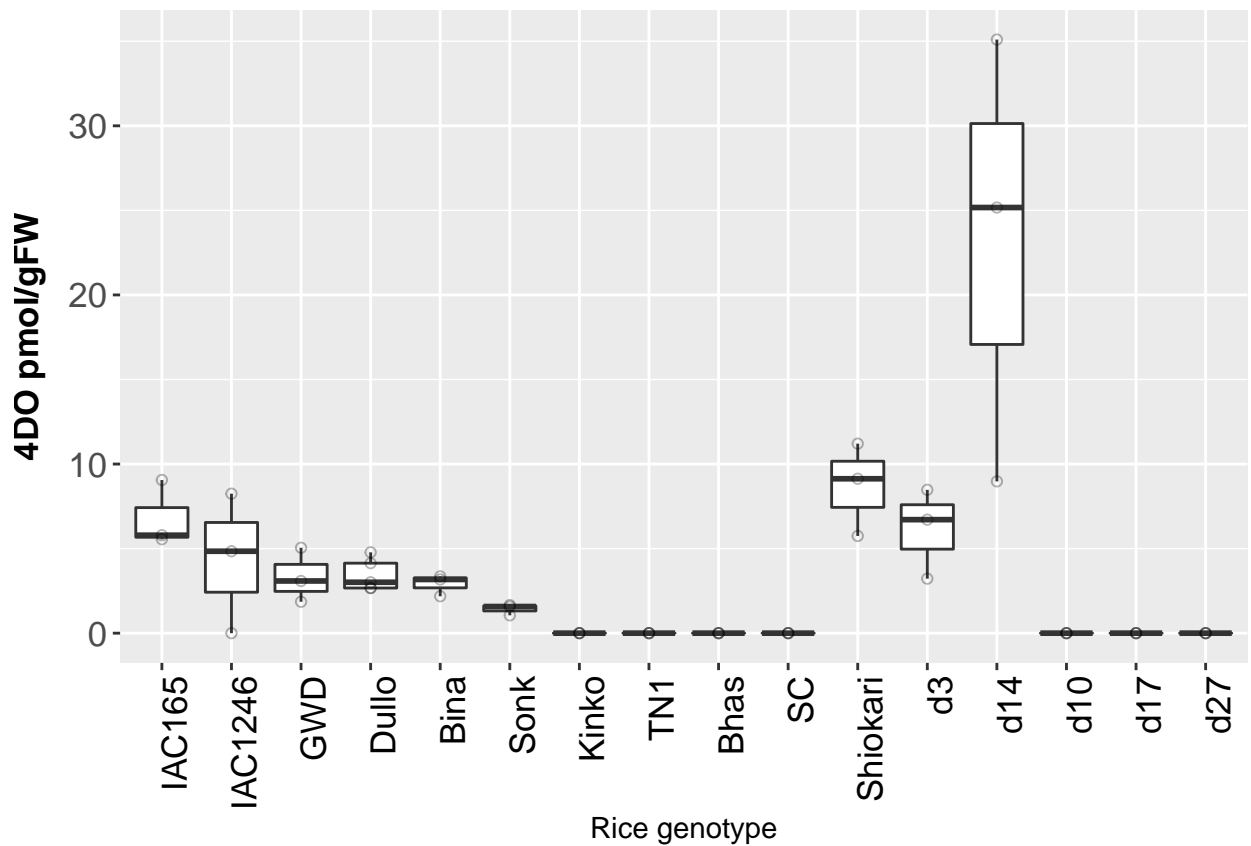
```

foSLs$Genotype<-droplevels(as.factor(foSLs$Genotype))
foSLs$Genotype2<-factor(foSLs$Genotype,
                        c("IAC165", "IAC1246", "GWD", "Dullo","Bina","Sonk",
                          "Kinko", "TN1","Bhas","SC","Shiokari","d3","d14","d10","d17","d27"))

ggplot(foSLs)+
  geom_boxplot(aes(x=Genotype2, y=X4D0_pmol_g), outlier.colour = NA)+labs(x="Rice genotype", y = "4D0 p
  geom_point(aes(x=Genotype2, y=X4D0_pmol_g), alpha = 0.3, shape = 21) +
  theme(axis.text.x = element_text(size = 13, colour = "black",angle=90, hjust=1),

```

```
axis.text.y = element_text(size = 13),
axis.title.y = element_text(face = "bold", size = 13, vjust = 3),
legend.text = element_text(size = 13), legend.title = element_text(size = 13, face = "bold"))
```



version

```
sessionInfo()
```

```
## R version 4.0.3 (2020-10-10)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19042)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.1252
## [2] LC_CTYPE=English_United States.1252
## [3] LC_MONETARY=English_United States.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.1252
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods   base
##
## other attached packages:
```

```

## [1] reshape2_1.4.4    tibble_3.0.4      ggfortify_0.4.11  dplyr_1.0.2
## [5] rcompanion_2.3.27 FSA_0.8.30        ggplot2_3.3.2
##
## loaded via a namespace (and not attached):
## [1] Rcpp_1.0.5          mvtnorm_1.1-1      lattice_0.20-41    tidyr_1.1.2
## [5] class_7.3-17        multcompView_0.1-8 zoo_1.8-8           digest_0.6.25
## [9] lmtest_0.9-38       R6_2.4.1           plyr_1.8.6         EMT_1.1
## [13] stats4_4.0.3        evaluate_0.14      rootSolve_1.8.2.1  e1071_1.7-4
## [17] highr_0.8           pillar_1.4.6       rlang_0.4.10       Exact_2.1
## [21] multcomp_1.4-16     rstudioapi_0.11    Matrix_1.2-18      rmarkdown_2.7
## [25] labeling_0.4.2      splines_4.0.3      stringr_1.4.0      munsell_0.5.0
## [29] compiler_4.0.3      xfun_0.21          pkgconfig_2.0.3    libcoin_1.0-7
## [33] DescTools_0.99.40   htmltools_0.5.1.1  tidyselect_1.1.0   gridExtra_2.3
## [37] lmom_2.8            expm_0.999-6       coin_1.4-0         codetools_0.2-16
## [41] matrixStats_0.57.0 dunn.test_1.3.5    crayon_1.3.4       withr_2.3.0
## [45] MASS_7.3-53         grid_4.0.3         gtable_0.3.0       lifecycle_0.2.0
## [49] magrittr_1.5        scales_1.1.1       gld_2.6.2          stringi_1.5.3
## [53] farver_2.0.3        ellipsis_0.3.1     generics_0.0.2     vctrs_0.3.4
## [57] boot_1.3-25         sandwich_3.0-0     nortest_1.0-4      TH.data_1.0-10
## [61] tools_4.0.3         glue_1.4.2         purrr_0.3.4        parallel_4.0.3
## [65] survival_3.2-7      yaml_2.2.1         colorspace_1.4-1   knitr_1.31
## [69] modeltools_0.2-23

```