Data analysis

Azam et al.

Population structure

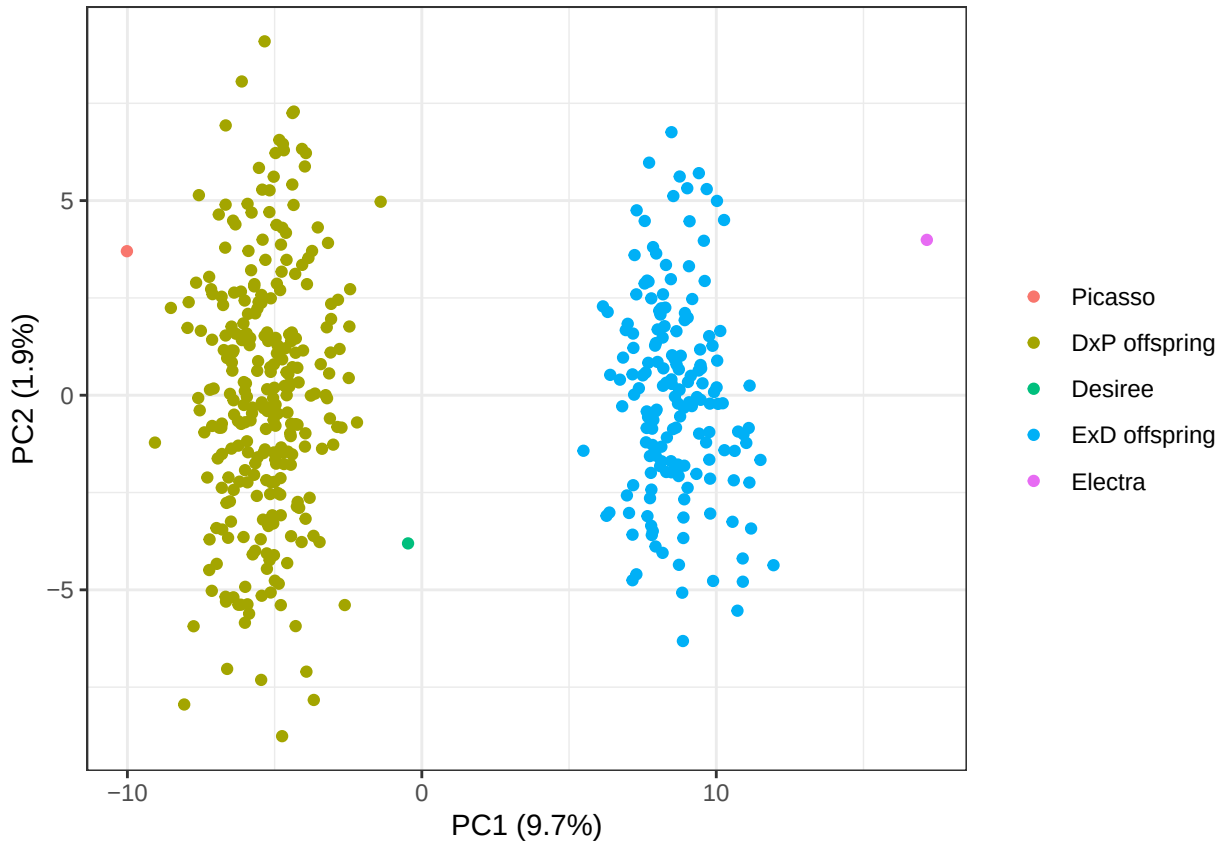
```
library(ggplot2)

data <- read.csv("FileS1.csv",check.names=F)
geno <- as.matrix(data[,-(1:3)])
dim(geno)
```

```
## [1] 1290  439
```

```
rownames(geno) <- data$Marker
id <- colnames(geno)
x <- scale(t(geno), scale=F, center=T)
x[is.na(x)] <- 0 #impute missing data with allele freq
pca <- prcomp(x, scale=F, center=T)
pca$sdev[1:2]^2/sum(pca$sdev^2) #PVE
```

```
## [1] 0.09734759 0.01866406
```

```
pdat <- data.frame(id=id, x=pca$x[id,1], y=pca$x[id,2],
        group = rep(c("Picasso", "Desiree", "Electra","DxP offspring","ExD offspring"),
         times = c(1, 1, 1,270,166)))

pdat$group <- factor(pdat$group, levels = c("Picasso","DxP offspring","Desiree",
                                            "ExD offspring","Electra"))

ggplot(pdat,aes(x=x,y=y,col=group,fill=group)) + geom_point()+
  xlab("PC1 (9.7%)") + ylab("PC2 (1.9%)") + theme_bw()+ theme(legend.title=element_blank())
```



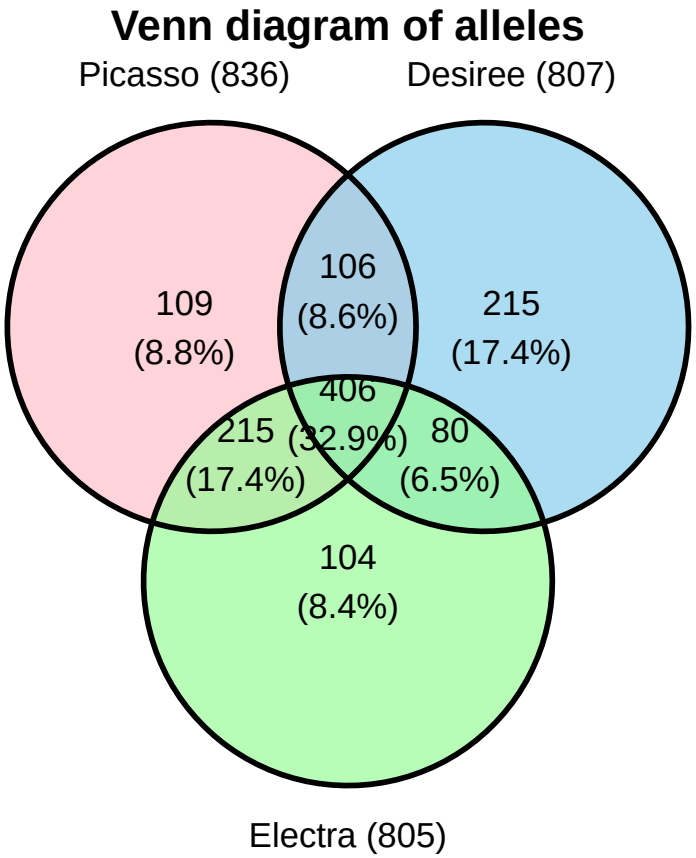

```
ggsave("FigureS6.png",device="png",dpi=600,height=4,width=7)
```

Allele count per parent

```
library(ggvenn)

Picasso <- data[,c(1,4)]
Picasso <- Picasso$Marker[!is.na(Picasso$Picasso) & Picasso$Picasso != 0]

Desiree <- data[,c(1,5)]
Desiree <- Desiree$Marker[!is.na(Desiree$Desiree) & Desiree$Desiree != 0]

Electra <- data[,c(1,6)]
Electra <- Electra$Marker[!is.na(Electra$Electra) & Electra$Electra != 0]

ls <-list("Picasso (836)"=Picasso,"Desiree (807)"=Desiree,"Electra (805)"=Electra)

ggvenn(
  ls, fill_color = c("pink", "skyblue","palegreen"),  stroke_color = "black",
  stroke_size = 1,set_name_color = "black", set_name_size = 4.5,
  text_color = "black",text_size = 4.5,fill_alpha = 0.7) +
  ggtitle("Venn diagram of alleles") +
  theme_void() +
```

```
  theme(plot.title = element_text(size = 16, face = "bold", hjust = 0.5),
    axis.text = element_blank(), axis.ticks = element_blank(),
    axis.title = element_blank()
  )
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## i The deprecated feature was likely used in the ggvenn package.
##   Please report the issue to the authors.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```




```
ggsave("FigureS5.png",device="png",dpi=600,height=5,width=4)
```

Genetic distance estimation

```
library(PolyOriginR)

polyOriginR(genofile = "FileS1.csv",
            pedfile = "FileS2.csv",
            isphysmap = TRUE, refinemap = TRUE, refineorder = FALSE,
```

```
            chrpairing_phase = 22, chrpairing = 44, nworker = 10,
            juliapath="path/to/julia",
            workdir=getwd(),
            outstem = "outstem")
```

```
library(stringr)
library(LinkageMapView)

FileS3 <- read.csv("outstem_parentphased_corrected.csv")
map <- FileS3[,1:3]
map$chromosome <- as.numeric(substring(map$chromosome, 4))
map$locus <- str_extract(FileS3$marker, "^[^_]+_[^_]+")
map <- map[!duplicated(map[c("locus", "chromosome", "position")]), ]
map <- map[,2:4]

titles <- c("chr01", "chr02", "chr03", "chr04", "chr05", "chr06",
            "chr07", "chr08", "chr09", "chr10", "chr11", "chr12")
lmv.linkage.plot(mapthis = map, outfile = "FigureS7.pdf", lg.col = "beige",
                 main = "Integrated chromosomal map", cex.main = 2,
                 cex.lgtitle = 1.5, ruler = TRUE, ylab = "Distance in cM",
                 lgtitles = titles, autoconnadj = FALSE)
```

```
## Required pdf.width = 25.5771666666667

## Required pdf.height = 7.5838

## Using pdf.width = 26

## Using pdf.height = 8
```

```
write.csv(FileS3, "FileS3.csv", row.names = FALSE)
```

Genotype probabilities estimation for interval mapping

```
FileS3 <- read.csv("outstem_parentphased_corrected.csv")
FileS3$chromosome <- as.numeric(substring(FileS3$chromosome, 4))
tmp <- list()
str <- c("2|1|1|1", "1|2|1|1", "1|1|2|1", "1|1|1|2")
sizes <- c(124, 81, 111, 90, 120, 82, 83, 99, 91, 81, 78, 94)
df <- NULL
for (l in 1:12){
  tmp <- expand.grid(chr = l,pos = 1:sizes[l],hom = 1:12)
  df <- rbind(df,tmp)
}
df <- df %>% arrange(chr, pos) %>%
  mutate(id = paste0("chr", chr, ".", pos, ".", hom), .before = chr) %>%
  mutate(
    P1 = ifelse(hom %in% 1:4, str, "1|1|1|1"),
    P2 = ifelse(hom %in% 5:8, str, "1|1|1|1"),
```

```
    P3 = ifelse(hom %in% 9:12, str, "1|1|1|1")
  )
df <- df[,-4]
for (i in 7:442){
  df[,i] <- NA
}
colnames(df) <- colnames(FileS3[,1:442])
combo <- rbind(df, FileS3)
combo <- combo %>% arrange(chromosome, position)
combo[1:12,1:10]
```

```
##       marker chromosome position Picasso Desiree Electra P264 P256 P248 P240
## 1     C1_1_1          1        0 1|1|1|2 1|1|1|2 1|2|1|2    2    1    0    1
## 2     C1_1_2          1        0 2|2|2|1 2|2|2|1 2|1|2|1    2    3    4    3
## 3   chr1.1.1          1        1 2|1|1|1 1|1|1|1 1|1|1|1   NA   NA   NA   NA
## 4   chr1.1.2          1        1 1|2|1|1 1|1|1|1 1|1|1|1   NA   NA   NA   NA
## 5   chr1.1.3          1        1 1|1|2|1 1|1|1|1 1|1|1|1   NA   NA   NA   NA
## 6   chr1.1.4          1        1 1|1|1|2 1|1|1|1 1|1|1|1   NA   NA   NA   NA
## 7   chr1.1.5          1        1 1|1|1|1 2|1|1|1 1|1|1|1   NA   NA   NA   NA
## 8   chr1.1.6          1        1 1|1|1|1 1|2|1|1 1|1|1|1   NA   NA   NA   NA
## 9   chr1.1.7          1        1 1|1|1|1 1|1|2|1 1|1|1|1   NA   NA   NA   NA
## 10  chr1.1.8          1        1 1|1|1|1 1|1|1|2 1|1|1|1   NA   NA   NA   NA
## 11  chr1.1.9          1        1 1|1|1|1 1|1|1|1 2|1|1|1   NA   NA   NA   NA
## 12 chr1.1.10          1        1 1|1|1|1 1|1|1|1 1|2|1|1   NA   NA   NA   NA
```

```
write.csv(combo, "FileS4.csv", row.names = FALSE)
```

```
library(PolyOriginR)
polyOriginR(genofile = "FileS4.csv",
            pedfile = "FileS2.csv",
            isphysmap = FALSE, refinemap = FALSE, refineorder = FALSE,
            chrpairing_phase = 22, chrpairing = 44, nworker = 1,
            juliapath="path/to/julia",
            workdir=getwd(),
            outstem = "outstem_IM")
```

Phenotypic data analysis

```
library(StageWise)

effects <- data.frame(name=c("block"),
                      fixed=c(FALSE),
                      factor=c(TRUE))
###Common scab
model_cs <- Stage1(filename="FileS5.csv", traits="Scab_score",
                   spline = c("row","range"), solver = "spats", effects = effects)

### Plant maturity
model_pm <- Stage1(filename="FileS6.csv", traits=c("plant_maturity"),
                   spline = c("row","range"), solver = "spats", effects = effects)
```

```
### Tuber shape
model_ts <- Stage1(filename="FileS6.csv", traits=c("tuber_shape"),
                   spline = c("row","range"), solver = "spats", effects = effects)

model_cs$fit[,c(1,3)] #common scab
```

```
##         env H2.entry
## 2905 GH2023     0.47
## 2925 GH2024     0.71
## 1309 OP2023     0.56
## 1    OP2024     0.74
```

```
model_pm$fit[,c(1,3)] #plant maturity
```

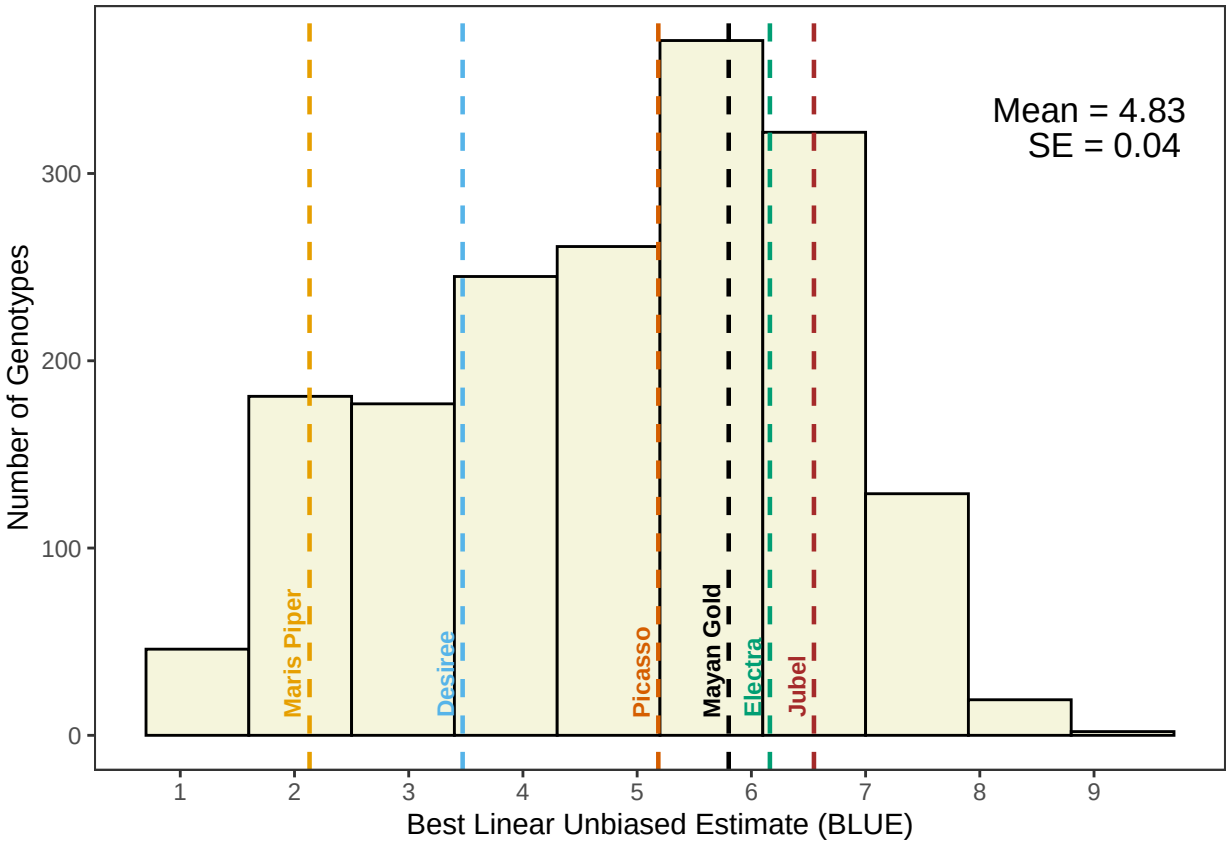
```
##      env H2.entry
## 1 OP2024     0.78
```

```
model_ts$fit[,c(1,3)] #tuber shape
```

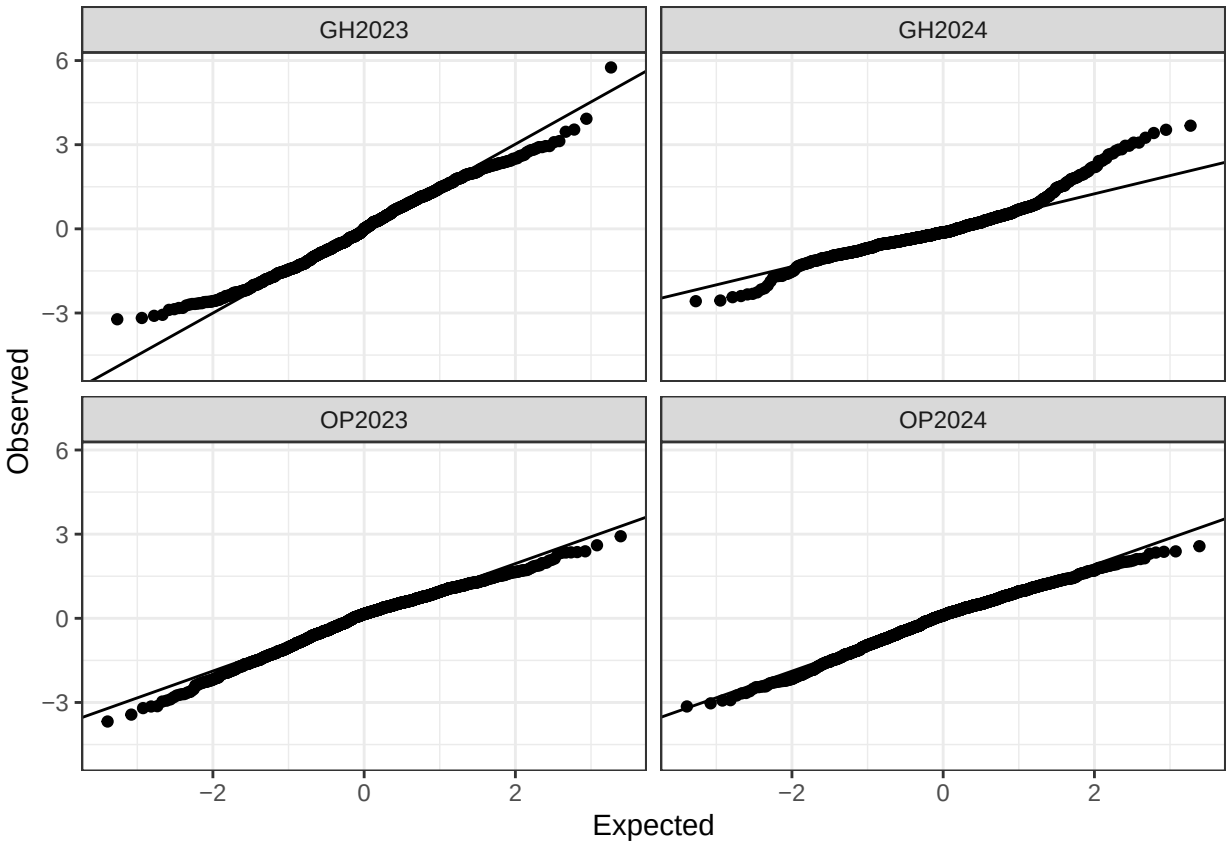
```
##      env H2.entry
## 1 OP2024     0.82
```

```
blues_cs <- model_cs$blues
colnames(blues_cs)[3] <- "Common.Scab"

write.csv(blues_cs[, c(2, 3, 1)], "FileS7.csv", row.names = FALSE)

mean <- mean(blues_cs$Common.Scab)
se <- sd(blues_cs$Common.Scab) / sqrt(length((blues_cs$Common.Scab)))

highlight <- data.frame(Controls = c("Maris Piper", "Desiree", "Electra","Jubel","Picasso",
             "Mayan Gold"),BLUE = c(mean(blues_cs$Common.Scab[blues_cs$id =="MarisPiper"]),
              mean(blues_cs$Common.Scab[blues_cs$id == "Desiree"]),
              mean(blues_cs$Common.Scab[blues_cs$id == "Electra"]),
              mean(blues_cs$Common.Scab[blues_cs$id == "Jubel"]),
              mean(blues_cs$Common.Scab[blues_cs$id == "Picasso"]),
              mean(blues_cs$Common.Scab[blues_cs$id == "MayanGold"])))

p <- ggplot(blues_cs, aes(x = Common.Scab)) +
  geom_histogram(binwidth = 0.9, fill = "beige", color = "black", boundary = 2.5) +
  scale_x_continuous(breaks = 0:9) +
  labs(x = "Best Linear Unbiased Estimate (BLUE)", y = "Number of Genotypes") +
  theme_bw()+
  theme(legend.position = "none",panel.grid =  element_blank())
p + geom_vline(data = highlight, aes(xintercept = BLUE, color = Controls),
               linetype = "dashed",size = 0.8) + geom_text(data = highlight,
               aes(x = BLUE, y = 10, label = Controls,color = Controls), angle = 90,
               vjust = -0.5, hjust = 0.01,fontface = "bold", size = 3.1) +
               scale_color_manual(values = c("Maris Piper" = "#E69F00",
               "Desiree" = "#56B4E9", "Electra" = "#009E73","Jubel"="brown",
               "Picasso"="#D55E00", "Mayan Gold"="black"))+ annotate("text",
```

```
 x = min(blues_cs$Common.Scab) + 5,
 y = max(ggplot_build(p)$data[[1]]$count) * 0.9, label = paste("Mean =",
 round(mean, 2)), hjust = -1.4, size = 4.5, color = "black")+ annotate("text",
x = min(blues_cs$Common.Scab) + 5,
y = max(ggplot_build(p)$data[[1]]$count) * 0.85,
label = paste("SE =", round(se, 2)), hjust = -2, size = 4.5, color = "black")
```



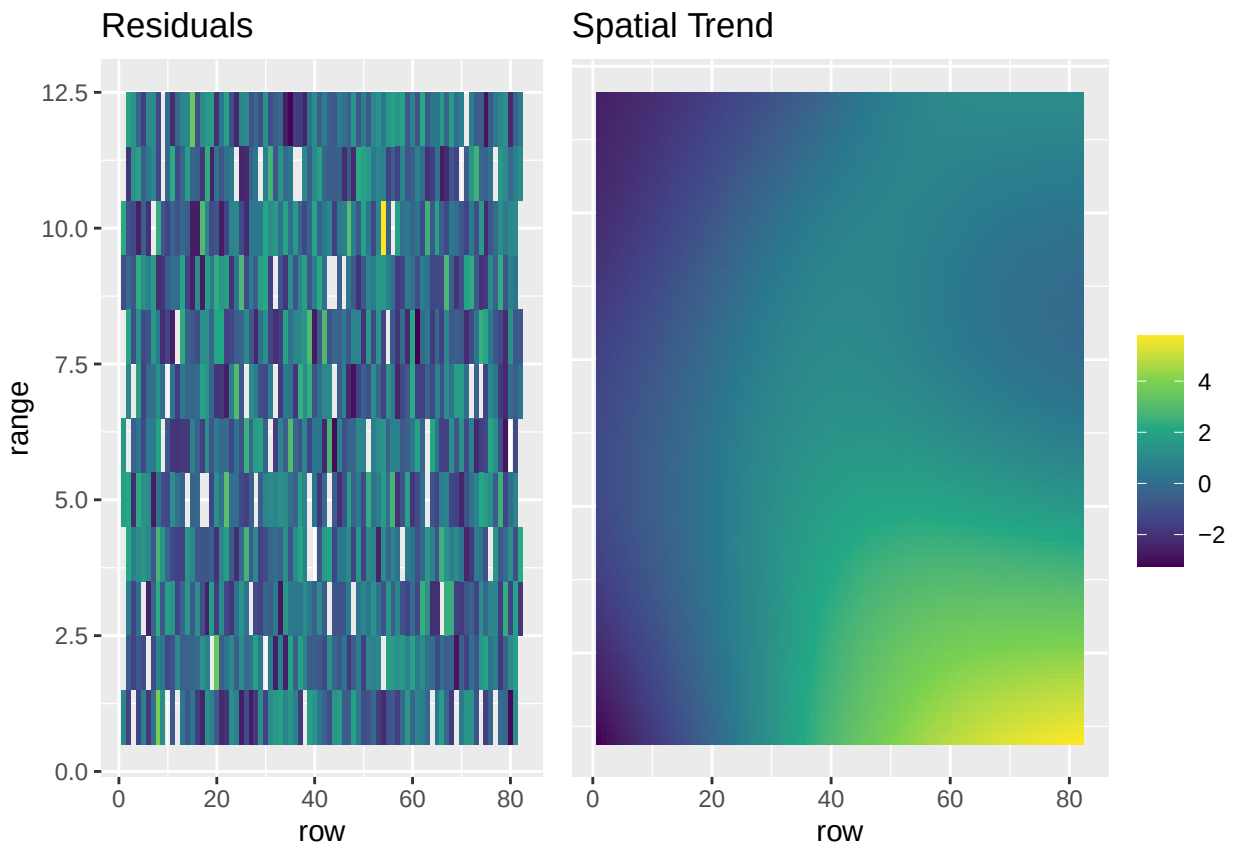

```
ggsave("Figure2.png",device="png",dpi=600,height=5,width=7)

model_cs$resid$qqplot
```



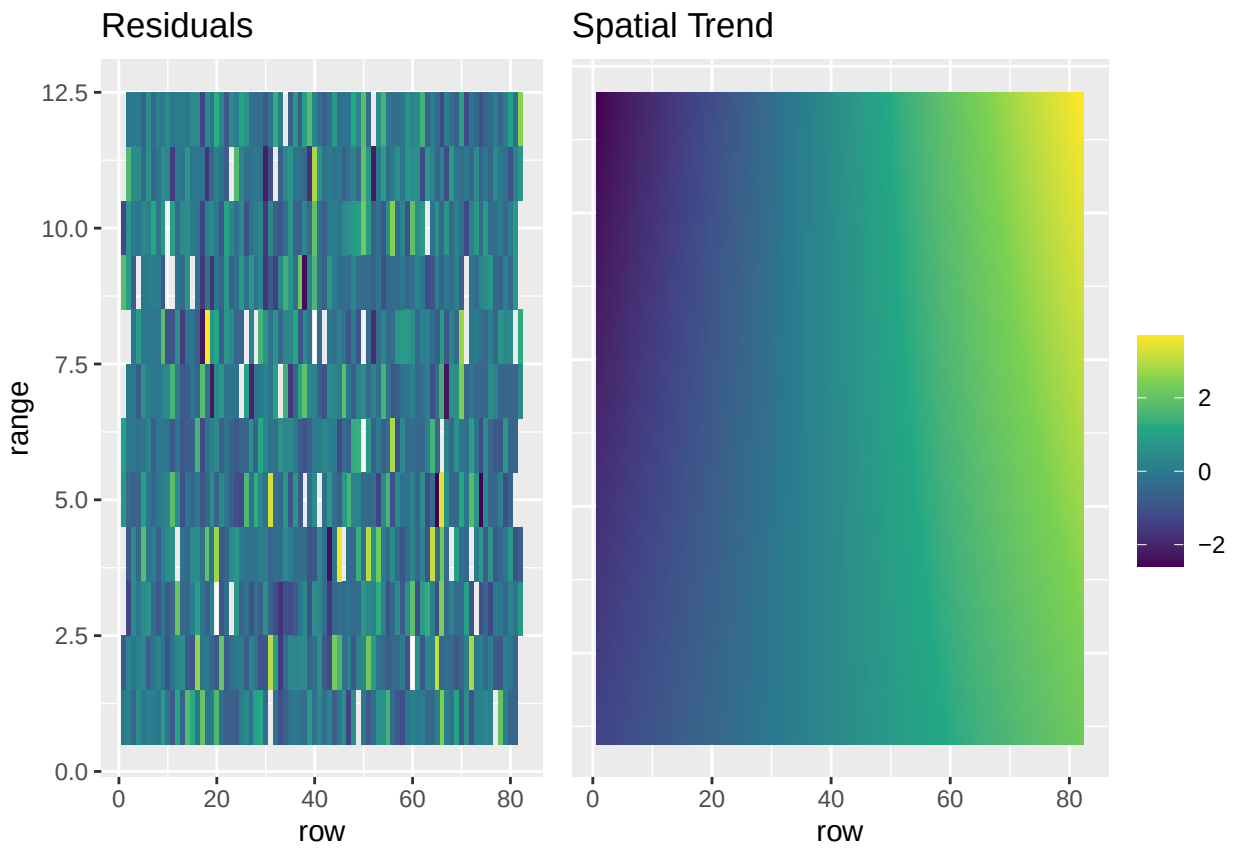

```
ggsave("FigureS1.png",device="png",dpi=600,height=5,width=7)
```

```
model_cs$resid$spatial
```

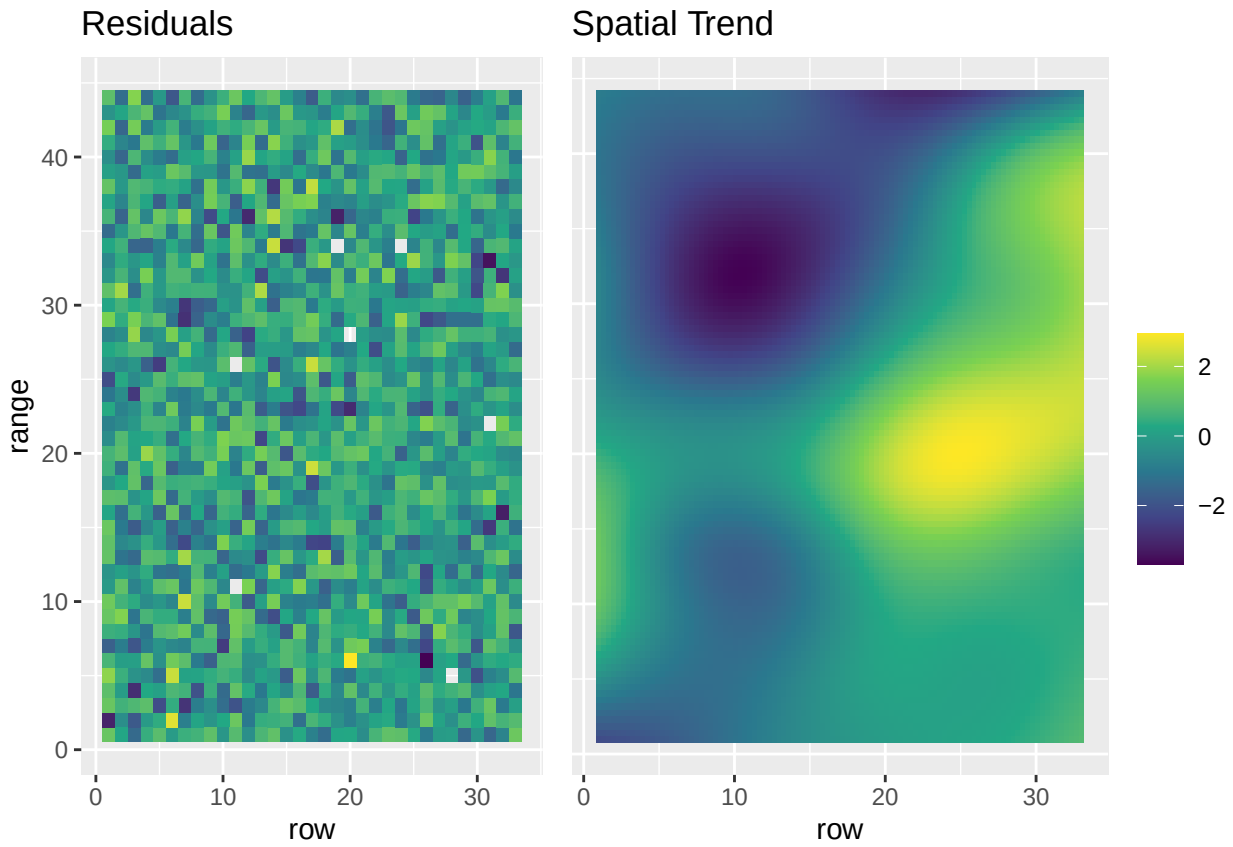
```
## $GH2023
```


Residuals
Spatial Trend
range
row
row
12.5
10.0
7.5
5.0
2.5
0.0
0
20
40
60
80
4
2
0
−2

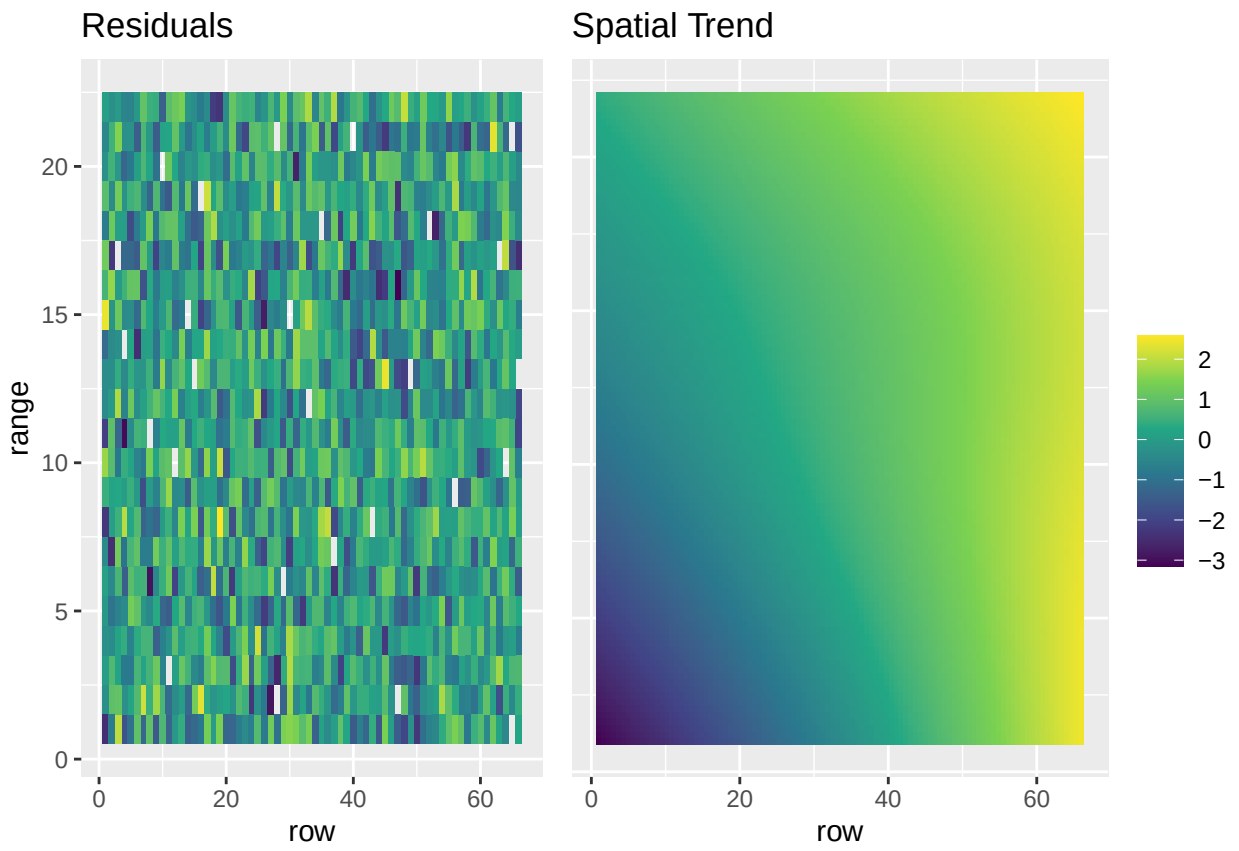
```
##
## $GH2024
```


Residuals
Spatial Trend
range
row
row
12.5
10.0
7.5
5.0
2.5
0.0
0
20
40
60
80
2
0
−2

```
##
## $OP2023
```



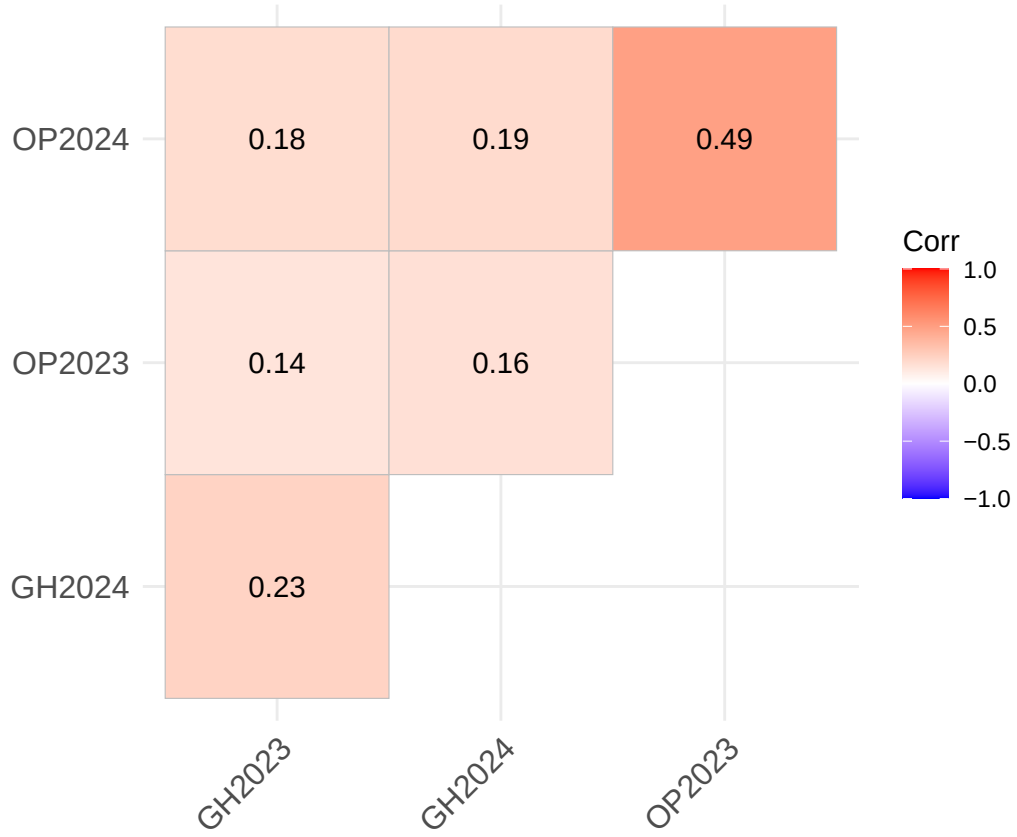

```
##
## $OP2024
```




```
p <- ggarrange(model_cs$resid$spatial$GH2023,model_cs$resid$spatial$GH2024,
               ncol = (2), labels = c("A","B"))
ggsave("FigureS3.1.png",device="png",dpi=600,height=4,width=10)
p <- ggarrange(model_cs$resid$spatial$OP2023,model_cs$resid$spatial$OP2024,
               ncol = (2), labels = c("A","B"))
ggsave("FigureS3.2.png",device="png",dpi=600,height=4,width=10)
```

Common scab correlation across environments

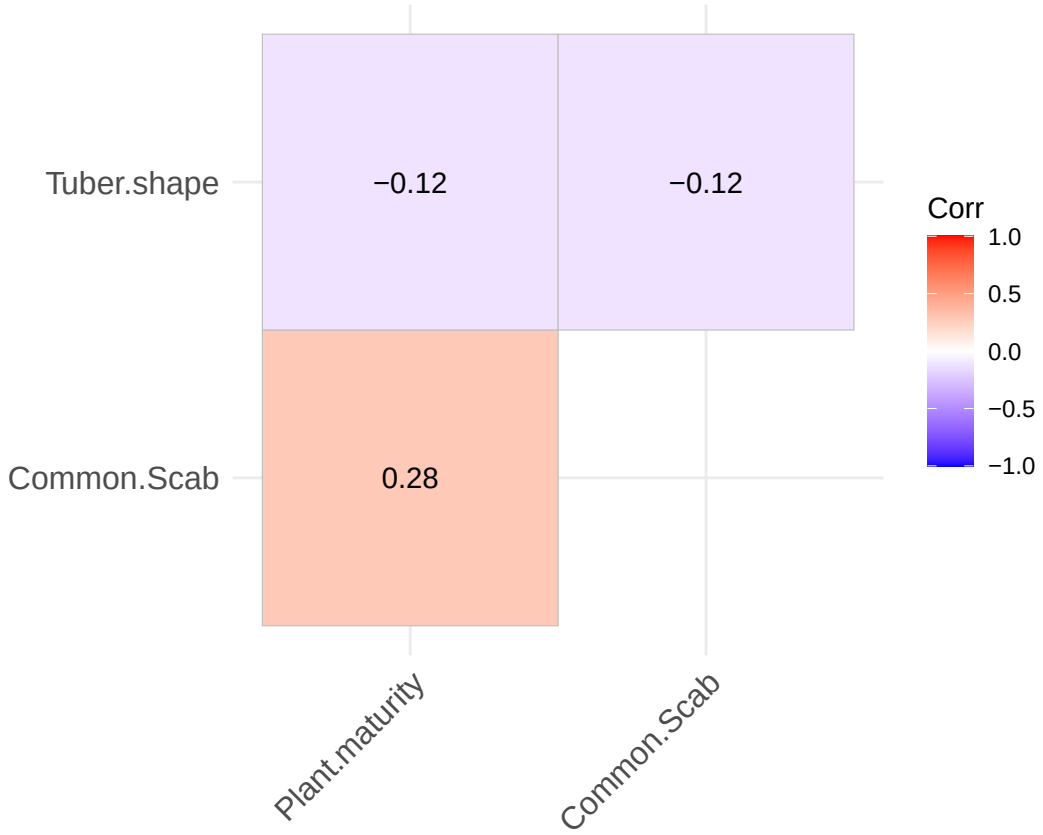
```
library(tidyr)
library(corrplot)
library(ggcorrplot)

blues_cs_byenv <- pivot_wider(blues_cs, names_from = "env", values_from = "Common.Scab")
p <- ggcorrplot(cor(blues_cs_byenv[,2:5], use = "complete.obs"), method = "square",
                type = "upper", lab = TRUE)
p
```




```
ggsave("FigureS2.png",device="png",dpi=600,height=5,width=5)
```

Different trait correlation in the same environment

```
blues_pm <-model_pm$blues
colnames(blues_pm)[3] <- "Plant.maturity"
write.csv(blues_pm[,2:3], "FileS9.csv", row.names = FALSE)

blues_ts <-model_ts$blues
colnames(blues_ts)[3] <- "Tuber.shape"
write.csv(blues_ts[,2:3], "FileS10.csv", row.names = FALSE)

merge <- merge(blues_pm, blues_cs[blues_cs$env == "OP2024", ], by = "id")
merge <- merge(merge, blues_ts, by = "id")
merge <- merge[,c(1,3,5,6,7)]

p <- ggcorrplot(cor(merge[,c(2,3,5)]), method = "square", type = "upper", lab = TRUE)
p
```




```
ggsave("FigureS4.png",device="png",dpi=600,height=5,width=5)
```