SAM size

Sept. 22, 2016

Setup

Normalize path and load libraries etc.

```
library(gsl)
library(dplyr)
library(magrittr)
library(cowplot)
library(viridis)
cbPalette <- c("#999999", "#E69F00", "#56B4E9", "#009E73",
               "#F0E442", "#0072B2", "#D55E00", "#CC79A7")
cols <- viridis(10)

## more packages
library(nlme)
#devtools::install_github("yangjl/quantgen")
library(quantgen)
library(tidyr)
```

Processing Cell Number Data

```
countsize <- read.csv("data/SAM_cellcount.csv", header=T)
head(countsize)
```

```
##   Count_Cells Mean_Area_percell Genotype Plant Growth_Period BAK1 SDA1
## 1         158          791.4494     A554     1            G1  REF  REF
## 2         188          831.1223     A554     1            G2  REF  REF
## 3         199          947.1709     A554     1            G3  REF  REF
## 4         205          812.2780     A554     2            G1  REF  REF
## 5         137          732.2628     A554     2            G2  REF  REF
## 6         228          959.8991     A554     2            G3  REF  REF
##     SAM_V
## 1 2502482
## 2 2502482
## 3 2502482
## 4 2502482
## 5 2502482
## 6 2502482
```

In the above SAM Cell count table, 14 Genotypes were collected for SAM cell counts in 3 growth periods, each period with 3 plants. There are also two factors, BAK1 and SDA1, that associated with some traits. From the data, we learned that `Count_Cells` is significantly correlated with `SAM_V` ($r = 0.75, Pvalue < 0.01$). But `Count_Cells` is not correlated with `Mean_Area_percell` ($r = 0.03, Pvalue = 0.7$).

We estimated BLUE values separately for each **growth period** by fitting a linear mixed model, where `Genotype` as a fixed effect and `Plant` as a random effect.

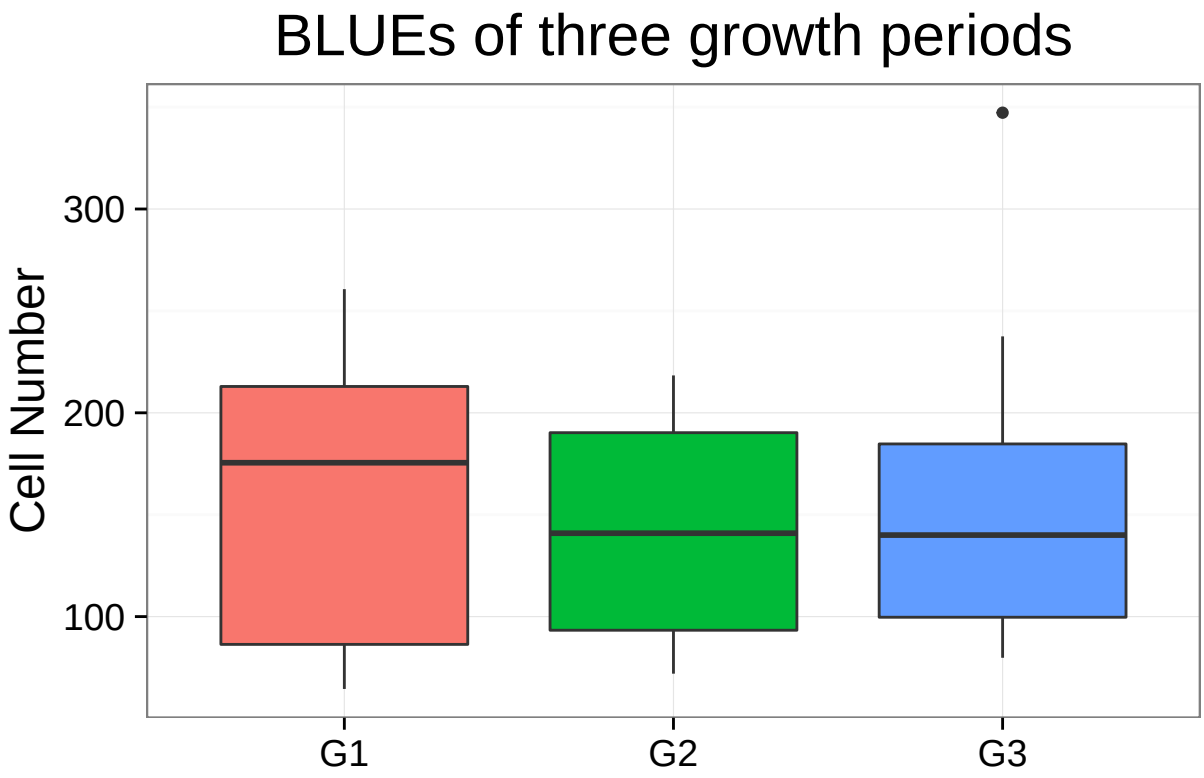
```
csg1 <- mixed_model(data = subset(countsize, Growth_Period %in% "G1"),
                    model = Count_Cells ~ Genotype, random = ~1 | Plant)

csg2 <- mixed_model(data = subset(countsize, Growth_Period %in% "G2"),
                    model = Count_Cells ~ Genotype, random = ~1 | Plant)

csg3 <- mixed_model(data = subset(countsize, Growth_Period %in% "G3"),
                    model = Count_Cells ~ Genotype, random = ~1 | Plant)

cs <- merge(csg1, csg2, by="Genotype")
cs <- merge(cs, csg3, by="Genotype") %>% set_names(c("genotype", "g1", "g2", "g3"))
lcs <- cs %>% gather(key="Growth", value="cellnum", 2:4)

#theme_set(theme_grey(base_size = 18))
p <- ggplot(lcs, aes(x=toupper(Growth), y=cellnum, fill = Growth)) +
    theme_bw(base_size = 18) +
    geom_boxplot() +
    ggtitle("BLUEs of three growth periods") + xlab("") + ylab("Cell Number") +
    guides(fill=FALSE)
p
```



Merge with other phenotypes

```
#samsize<-read.csv("~/Desktop/samsize.csv",header=T)
#samsize_unsummary<-read.csv("~/Desktop/samsize_unsummarized.csv",header=T)
plantstuff <- read.csv("data/plantstuff.csv",header=T) %>%
```

```
  select(Genotype,PlantH..cm.,EarH..cm.,mean_nodes,mean_dia_ear..mm.,mean_dia_below..mm.,DTA,SAM_V..um.
  set_names(c("genotype","plant_height","ear_height","leaf_nodes","ear_width","stem_width","DTA","SAM_v
plantstuff$genotype <- toupper(plantstuff$genotype)
cs$genotype <- toupper(cs$genotype)


tas <- countsize[!duplicated(countsize$Genotype), ] %>%
  select(Genotype, BAK1, SDA1) %>%
  set_names(c("genotype", "BAK1", "SDA1"))
tas$genotype <- toupper(tas$genotype)
plantcount <- merge(plantstuff, cs, by="genotype") %>%
  merge(tas, by="genotype")

head(plantcount)
```

```
##   genotype plant_height ear_height leaf_nodes ear_width stem_width  DTA
## 1     A554        130.1       42.9   12.66667  15.48333   16.54333 58.7
## 2      B57        135.7       62.8   12.33333  15.23333   16.24000 68.5
## 3      B73        169.3       91.1   14.00000  17.67667   19.05667 67.7
## 4   CML261        207.4      122.5   21.00000  18.21000   18.76667 85.0
## 5   CML322        136.6       83.2   17.00000  16.97667   17.13333 80.0
## 6   CML333        169.8      109.1   17.00000  14.00667   15.23000 77.0
##   SAM_volume       g1       g2        g3 BAK1 SDA1
## 1 1164089.33 174.3333 145.3333 187.46636  REF  REF
## 2 -551935.71  64.5000  74.5000  79.79969  REF  REF
## 3  451176.36 190.6667 182.0000 161.50000  REF  REF
## 4   92414.76 118.3333 144.3333 118.50000  TAS  REF
## 5 -528356.95  67.5000  72.0000  89.13302  REF  REF
## 6 -104991.26  91.5000 113.7500  97.50000  REF  REF
```

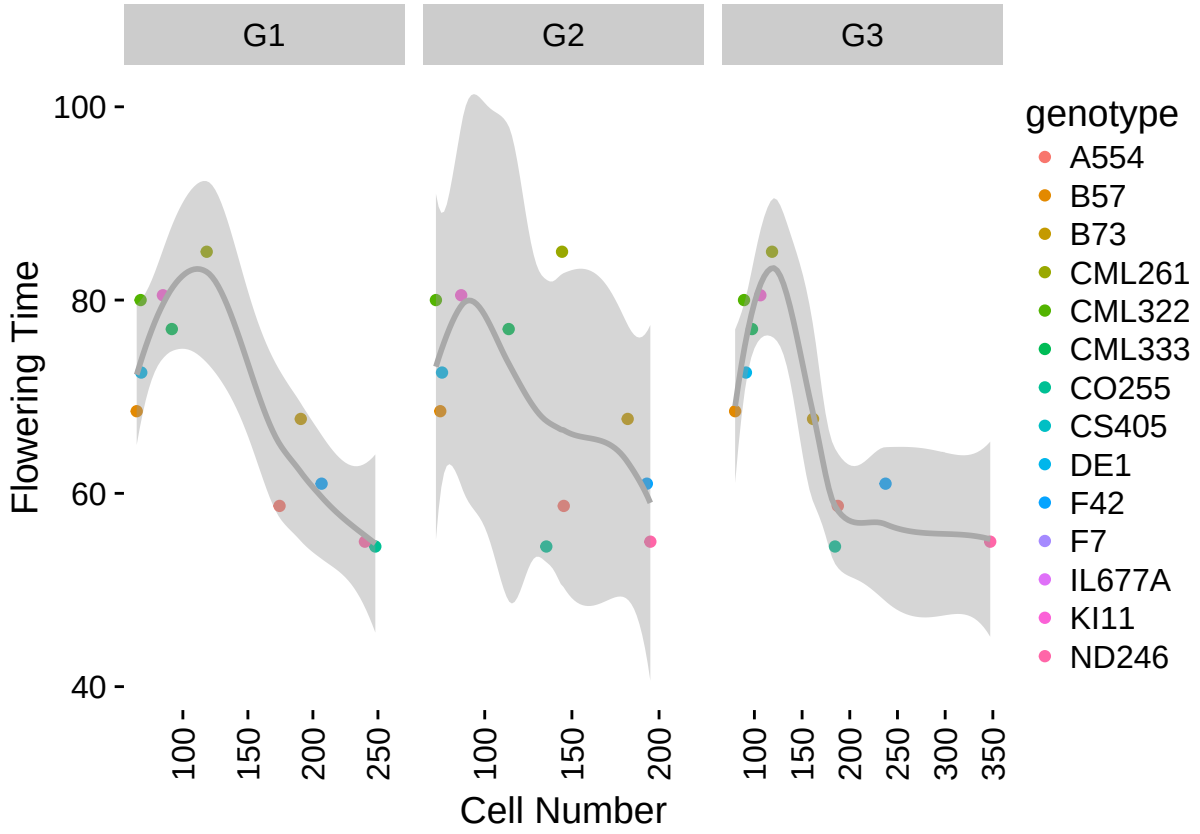
```
#plot(plantcount$DTA, plantcount$g1)
```

Correlation plot between cell number and flowering time in three growth period

```
lcs <- cs %>% gather(key="Growth", value="cellnum", 2:4)

p1 <- plantcount[, c("genotype", "g1", "g2", "g3")] %>%
  gather(key="Growth", value="cellnum", 2:4) %>%
  merge(plantcount[, 1:8], by="genotype")
p1$Growth <- toupper(p1$Growth)

ggplot(p1, aes(y=DTA, x=cellnum))+
  geom_point(aes(color=genotype))+
  geom_smooth(color="dark grey")+
  facet_wrap(~Growth, scales="free_x")+
  theme(axis.text.x=element_text(angle = 90, vjust = 0.5)) +
  xlab("Cell Number")+
  ylab("Flowering Time")
```




```
#summary(lm(data=p1, DTA ~ cellnum + Growth))
```

Fit a Linear Mixed Model with relatedness matrix as random

The standardized relatedness matrix was estimated with `GEMMA` using GBS data.

```
fam0 <- read.table("cache/GBSv2.7_id14_flt.fam", header=F)

idcurated <- read.csv("cache/cellnum_GBS_sampleid_curated.csv")
fam <- merge(fam0, idcurated[, c("FullName", "DNASample")], by.x="V1", by.y="FullName", sort=FALSE)
fam$DNASample <- toupper(fam$DNASample)

#### Relatedness estimated from GEMMA
mx <- read.table("cache/mx.sXX.txt")

row.names(mx) <- fam$DNASample
names(mx) <- fam$DNASample
#library(d3heatmap)
#d3heatmap(mx, scale = "column", dendrogram = "none", color = "Blues")
mx[mx < 0] <- 0
mx[mx > 1] <- 1
mx <- as.matrix(mx)
```

After fitting the related matrix as random, growth period, BAK1 and SDA1 as fixed effects, DTA significantly associated with G1 (effect=-0.11, P value=0.0076) and G3 (effect=-0.08, P value=1.3e-09), but not G2 (effect=-0.08, P value=0.170).

```
library(bdsmatrix)
library(coxme)
row.names(plantcount) <- plantcount$genotype

gfit1 <- lmekin(DTA ~ g1 + BAK1 + SDA1 + (1|genotype), data=plantcount, varlist= mx, method="REML")
gfit2 <- lmekin(DTA ~ g2 + BAK1 + SDA1 + (1|genotype), data=plantcount, varlist= mx, method="REML")
gfit3 <- lmekin(DTA ~ g3 + BAK1 + SDA1 + (1|genotype), data=plantcount, varlist= mx, method="REML")

gfit1
gfit2
gfit3
```

I did not test for other traits, but should be straight forward.

```
#cell count, SAM volume
ggplot(plantcount,aes(y=SAM_volume,x=cell_number))+
  geom_point(color=cols[1])+
  geom_smooth(method="loess",color="dark grey")+
  facet_wrap(~growth_period,scales="free_x")+
  ylab("SAM volume")+
  xlab("Cell Number")

#cell size, SAM volume
ggplot(plantcount,aes(y=SAM_volume,x=cell_size))+
  geom_point(color=cols[2])+
  geom_smooth(method="loess",color="dark grey")+
  facet_wrap(~growth_period,scales="free_x")+
  ylab("SAM volume")+
  xlab("Cell Size")

#cell size, cell number
ggplot(plantcount,aes(x=cell_number,y=cell_size))+
  geom_point(color=cols[3])+
  geom_smooth(method="loess",color="dark grey")+
  facet_wrap(~growth_period,scales="free_x")+
  xlab("Cell Number")+
  ylab("Cell Size")

#more cells -> no diff plant height. smaller ear height, weakly smaller ear width & stem_width, shorter

summary(lm(data=plantcount,DTA~cell_number+SDA1+BAK1))
```