SAM size

Sept. 22, 2016

Setup

Normalize path and load libraries etc.

```r
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)
library(quantgen)
library(tidyr)
```

Processing Cell Number Data

```r
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
```
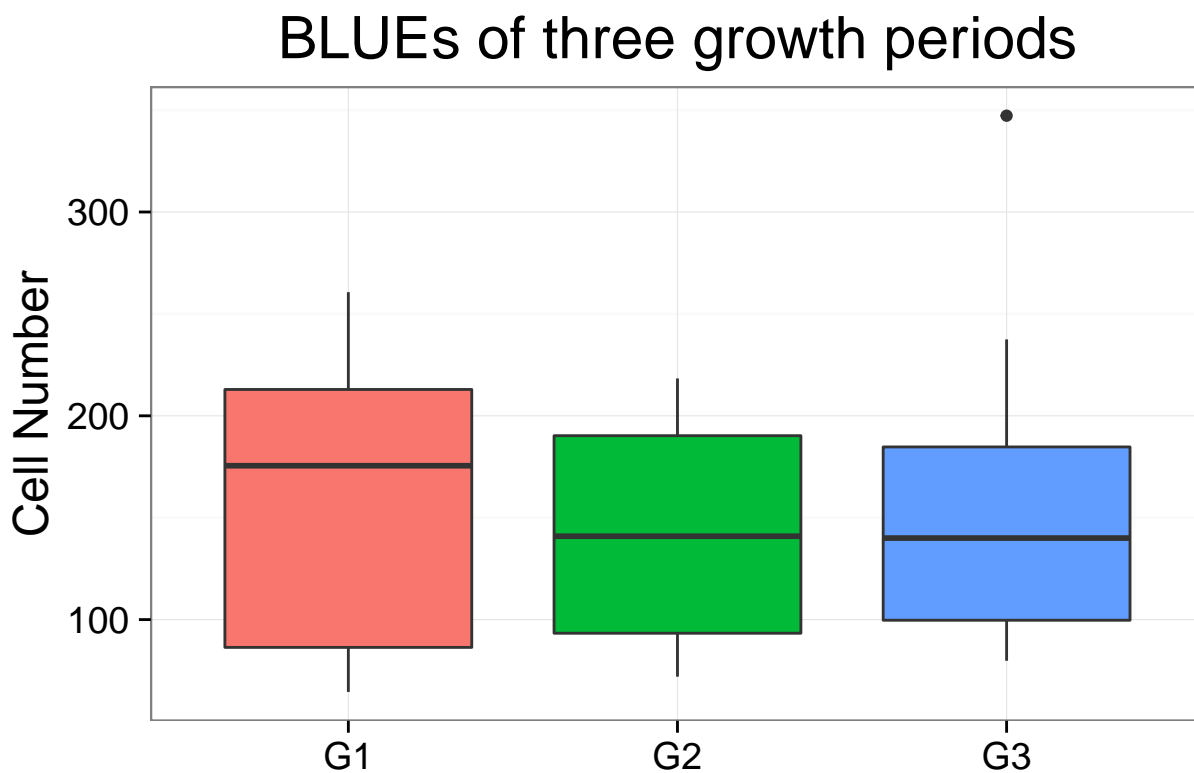

Merge with other phenotypes

```
#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) %>%
```

```r
  select(Genotype,PlantH..cm.,EarH..cm.,mean_nodes,mean_dia_ear..mm.,mean_dia_below..mm.,DTA,SAM_V..um.3
  set_names(c("genotype","plant_height","ear_height","leaf_nodes","ear_width","stem_width","DTA","SAM_vo
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
```

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

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

```r
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")+
  xlab("Cell Number")+
  ylab("Flowering Time")
```
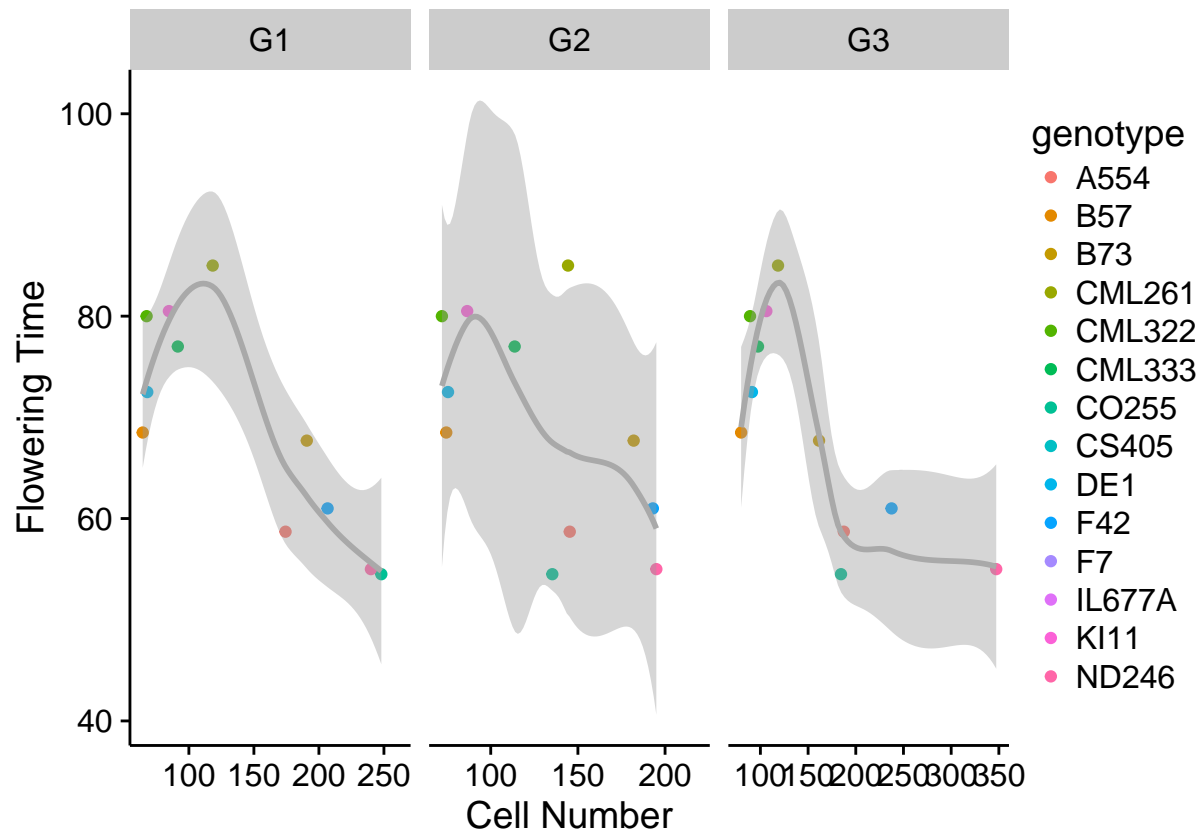
```
#summary(lm(data=p1, DTA ~ cellnum + Growth))
```

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).

4

```r
library(coxme)
```

```
## Loading required package: survival
```

```
## Warning: package 'survival' was built under R version 3.2.5
```

```
## Loading required package: bdsmatrix
```

```
##
## Attaching package: 'bdsmatrix'
```

```
## The following object is masked from 'package:base':
##
##     backsolve
```

```r
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
```

```
## Linear mixed-effects kinship model fit by REML
##   Data: plantcount
##   Log-likelihood = -29.07729
##   n=11 (3 observations deleted due to missingness)
##
## Model:  DTA ~ g1 + BAK1 + SDA1 + (1 | genotype)
## Fixed coefficients
##                  Value  Std Error     z       p
## (Intercept) 84.2343946 5.25581425 16.03 0.0000
## g1          -0.1099299 0.04119906 -2.67 0.0076
## BAK1TAS       9.2883770 5.46303801  1.70 0.0890
## SDA1TAS      -7.3068982 7.54803804 -0.97 0.3300
##
## Random effects
##  Group    Variable Std Dev   Variance
##  genotype Vmat.1    5.396819 29.125653
## Residual error= 3.492829
```

```r
gfit2
```

```
## Linear mixed-effects kinship model fit by REML
##   Data: plantcount
##   Log-likelihood = -30.5373
##   n=11 (3 observations deleted due to missingness)
##
## Model:  DTA ~ g2 + BAK1 + SDA1 + (1 | genotype)
## Fixed coefficients
```

```
##                    Value  Std Error     z      p
## (Intercept)  81.50424800 7.69649100 10.59 0.000
## g2           -0.08355468 0.06018004 -1.39 0.170
## BAK1TAS       8.51496678 6.93880292  1.23 0.220
## SDA1TAS     -17.21332466 7.51639592 -2.29 0.022
##
## Random effects
##  Group    Variable Std Dev   Variance
##  genotype Vmat.1   7.076177 50.072286
## Residual error= 4.080804
```

```
gfit3
```

```
## Linear mixed-effects kinship model fit by REML
##   Data: plantcount
##   Log-likelihood = -30.18341
##   n=11 (3 observations deleted due to missingness)
##
## Model:  DTA ~ g3 + BAK1 + SDA1 + (1 | genotype)
## Fixed coefficients
##                   Value  Std Error     z       p
## (Intercept) 83.14276139 3.40674800 24.41 0.0e+00
## g3          -0.08559534 0.01411914 -6.06 1.3e-09
## BAK1TAS      3.15304052 6.83472396  0.46 6.4e-01
## SDA1TAS     -7.21503651 7.16473607 -1.01 3.1e-01
##
## Random effects
##  Group    Variable Std Dev   Variance
##  genotype Vmat.1   7.974515 63.592885
## Residual error= 0.01373827
```

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))
```