Variances from basic MET model

Given that the file "Boro dataset.txt", which is the exact same file we worked with at the BRRI, is in the working directory, the following code first imports and formats the dataset and then fits a model to the plot data. Thus, this is all the code we need from txt data file to investigating the variance component estimates.

Part 1: Data import & formatting

```
rm(list=ls())
library(lme4)
library(lmerTest)

raw <- read.delim("Boro dataset.txt")

raw$t.k    <- as.numeric(substr(raw$Year, 1, 4))
raw$Env    <- as.factor(paste(raw$Year, raw$Location, sep="-"))
raw$r.i    <- as.numeric(raw$Year.of.release)
raw$G      <- as.factor(raw$Variety)
raw$L      <- as.factor(raw$Location)
raw$Y      <- as.factor(raw$t.k)
raw$Rep    <- as.factor(raw$Rep)
raw$Group  <- as.factor(raw$Group)

plotdata <- raw[ ,c("Y", "t.k", "r.i", "L", "Env", "G", "Rep", "Group", "Yield")]
```

Part 2a: Basic MET model

```
# lme4
lme4 <-lmer(formula = Yield ~
                    (1|G) + (1|L) + (1|Y) +
                    (1|Y:L) + (1|Y:G) + (1|L:G) +
                    (1|Y:L:G) + (1|Rep:Y:L),
            data     = plotdata)

vc.lme4 <- data.frame(VarCorr(lme4))[,c(1,4)]
vc.lme4$vcov <- round(vc.lme4$vcov, 3)
vc.lme4
```

As far as I can see, your model was correct: It included the three main effects (Y, L , G), their three two-way interaction effects and the three-way interaction effect, as well as the Rep:Y:L block effect per environment. (Keep in mind that this one-stage analysis is not the same thing as the two-stage analysis with genotype means per environment we did at the BRRI, since this one-stage analysis assume e.g. a single block variance across all environments)

However, I do not get the same variance component estimates. Note that I did not use the "summary(lme4)" command I showed at the BRRI in order to look at the variance component estimates. Instead, I used the last three lines in the code above. It does not change anything about the results – it simply gives a reduced, rounded output:

```
> vc.lme4
        grp   vcov
1    Y:L:G  0.276
2  Rep:Y:L  0.023
3      Y:G  0.008
4      L:G  0.063
5      Y:L  0.602
6        G  0.090
7        Y  0.086
8        L  0.493
9 Residual  0.202
```
*I get the same estimates when I fit the model with asreml-R

As you can see, the results are different from the ones you sent. I am not sure why that is. Since we seem to have fit the same model, I suspect it's the data. And I am not talking about the txt data file – even that one should be identical– I am talking about the "plotdata" object in R. Could you maybe compare your data preview in the environment window with mine and check whether there are differences? (e.g. number of observations, Factor/num format, number of levels)

```
plotdata              12741 obs. of 9 variables
  Y  : Factor w/ 15 levels "2001","2002",..: 1 1 1 1 1 1 1 2 2 2 ...
  t.k : num 2001 2001 2001 2001 2001 ...
  r.i : num 1970 1970 1970 1970 1970 1970 1970 1970 1970 1970 ...
  L  : Factor w/ 10 levels "Barisal","Bhanga",..: 1 2 3 5 7 8 9 1 2 …
  Env : Factor w/ 137 levels "2001-2002-Barisal",..: 1 2 3 4 5 6 7 …
  G  : Factor w/ 31 levels "BR1","BR12","BR14",..: 1 1 1 1 1 1 1 1 1 1..
  Rep : Factor w/ 3 levels "1","2","3": 1 1 1 1 1 1 1 1 1 1 ...
  Group: Factor w/ 3 levels "Long","Short",..: 2 2 2 2 2 2 2 2 2 2 …
  Yield: num 6.12 5.8 6.87 5.67 6.1 5.17 5.27 4.28 5.56 4.74 ...
```

If that's not it, then I'm afraid that I am quite clueless about the discrepancy.

Analysis per site and phenotypic correlation

Instead of analyzing data across environments as in Part 2a, we can analyze each site (=G:L combination) separately, but across all years, via:

Part 2b: Analyzing sites separately

```r
# within-site analyses #
#######################

# create list to store result per site
number.sites     <- length(levels(plotdata$L))
result.per.site <- matrix(list(), number.sites, 2)
rownames(result.per.site) <- unique(plotdata$L)
colnames(result.per.site) <- c("G.mean", "var.comp")
result.per.site

# loop through sites
for (selected.site in c(1:number.sites)){
  mod <- lmer(formula = Yield ~ G + (1|Y) + (1|Y:G) + (1|Y:Rep),
              data    = subset(plotdata, L==levels(plotdata$L)[selected.site]))

  # adjusted genotype means
  m       <- emmeans(mod, "G")
  means   <- as.data.frame(m)[,c("G", "emmean")]
  names(means) <- c("G", "adj.mean")
  result.per.site[[selected.site, "G.mean"]] <- means

  # variance components
  result.per.site[[selected.site, "var.comp"]] <- data.frame(VarCorr(mod))[,c(1,4)]

  rm(mod, m, means)
}
```
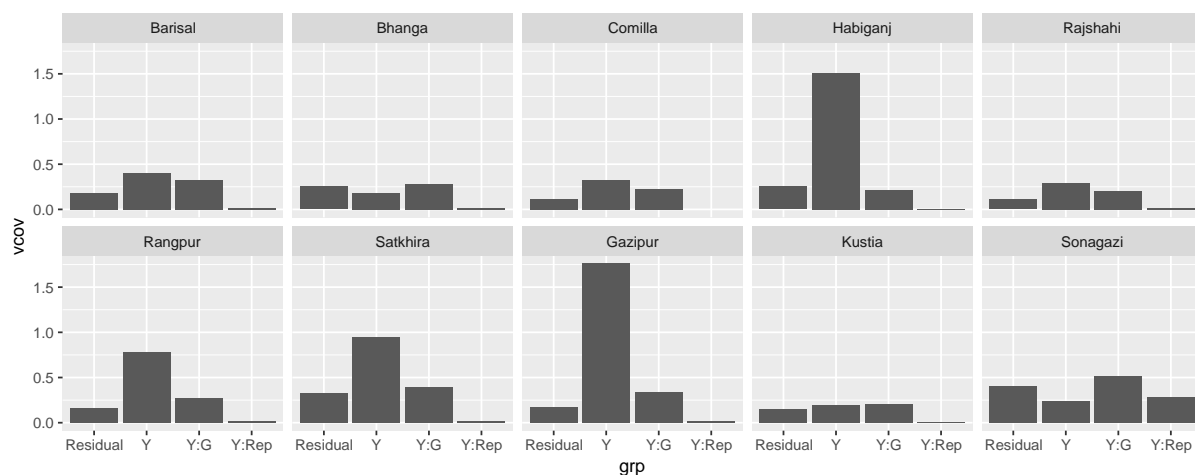
*[and some extra code not shown to format and plot data]

We can then plot the estimated variance components at each site:

Furthermore, we can correlate the adjusted genotype means between sites: