Bud genetics

Simon Joly

Set working directory

```
setwd("~/Documents/github/budgenetics/analyses")
```

Load packages and functions

```
library(nlme)
require(ape)
require(caper)
require(phytools)
require(picante)
source("duplicate.tips.R")
```

Load data

```
dater <- read.csv("./output/budsummary.csv")
alldater <- read.csv("./output/indforGBS.csv")
mytree <- read.nexus("./input/tree_intra.tre")
mytree0 <- read.nexus("./input/tree_nointra.tre")
```

Prepare data

This part is identical to yours.

```
# Subset the data and duplicate tips
daterch0 <- subset(alldater, chill=="chill0", select=c("site", "sp", "ind",
                                                       "treatcode", "warm", "photo",
                                                       "lday"))
daterch0$indX <- paste(daterch0$ind, rep(1:4, 102), sep="")
rownames(daterch0) <- daterch0$indX

# Duplicate tips
mytree.dup <- duplicate.tips(mytree, 4, sep.char="")
mytree0.dup <- duplicate.tips(mytree0, 4, sep.char="")

# Resolve tree and set min branch length to depth of tree/2000
mytree.dup.di <- multi2di(mytree.dup)
mytree.dup.di$edge.length[mytree.dup.di$edge.length==0] <-
  max(cophenetic(mytree.dup.di))/2000
```

```
mytree0.dup.di <- multi2di(mytree0.dup)
mytree0.dup.di$edge.length[mytree0.dup.di$edge.length==0] <-
  max(cophenetic(mytree0.dup.di))/2000
```

```
# It is a good idea to clean up the tree and dataset to keep
# only the samples present in both
toremove1 <- rownames(daterch0)[!(rownames(daterch0) %in% mytree.dup.di$tip.label)]
toremove2 <- mytree.dup.di$tip.label[!(mytree.dup.di$tip.label %in% rownames(daterch0))]
toremove <- c(toremove1,toremove2)
thetree <- ape::drop.tip(mytree.dup.di,toremove)
thetree0 <- ape::drop.tip(mytree0.dup.di,toremove)

# Clean the data. Note that there are still NA in the table
thedata <- daterch0[thetree$tip.label,]
```

Now, to facilitate the comparisons, convert the predictor to factors (binary variables).

```
thedata$photo <- as.factor(thedata$photo)
thedata$warm <- as.factor(thedata$warm)
compdat <- comparative.data(thetree, thedata, indX)
compdat0 <- comparative.data(thetree0, thedata, indX)
```

run the PGLS

I first run the PGLS with the tree with the intra-specific branch lengths

```
mod <- pgls(lday~warm*photo+site, lambda="ML", data=compdat)
summary(mod)
```

```
##
## Call:
## pgls(formula = lday ~ warm * photo + site, data = compdat, lambda = "ML")
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -265.80  -57.73    1.28   73.78  319.44
##
## Branch length transformations:
##
## kappa  [Fix]  : 1.000
## lambda [ ML]  : 0.900
##    lower bound : 0.000, p = < 2.22e-16
##    upper bound : 1.000, p = < 2.22e-16
##    95.0% CI   : (0.779, 0.949)
## delta  [Fix]  : 1.000
##
## Coefficients:
##                Estimate Std. Error  t value  Pr(>|t|)
## (Intercept)     54.6458    11.1572   4.8978 1.627e-06 ***
## warm20         -20.2265     1.4098 -14.3466 < 2.2e-16 ***
```

```
## photo12        -12.1258     1.3848  -8.7563 2.220e-16 ***
## siteSH           1.9366     1.4896   1.3001  0.194610
## warm20:photo12   6.0685     1.9582   3.0990  0.002136 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 92.83 on 285 degrees of freedom
## Multiple R-squared: 0.5807,  Adjusted R-squared: 0.5748
## F-statistic: 98.67 on 4 and 285 DF,  p-value: < 2.2e-16
```

Let's compare with the analysis with the tree with no intra-specific branch lengths

```
# With no intra-specific branch lengths
mod0 <- pgls(lday~warm*photo+site, lambda="ML", data=compdat0)
summary(mod0)
```

```
##
## Call:
## pgls(formula = lday ~ warm * photo + site, data = compdat0, lambda = "ML")
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -190.922  -36.701    4.768   36.333  134.025
##
## Branch length transformations:
##
## kappa  [Fix]  : 1.000
## lambda [ ML]  : 0.629
##    lower bound : 0.000, p = < 2.22e-16
##    upper bound : 1.000, p = 2.2204e-16
##    95.0% CI   : (0.427, 0.831)
## delta  [Fix]  : 1.000
##
## Coefficients:
##                 Estimate Std. Error  t value  Pr(>|t|)
## (Intercept)     54.97593    5.74453   9.5701 < 2.2e-16 ***
## warm20         -20.12317    1.60520 -12.5363 < 2.2e-16 ***
## photo12        -11.84224    1.57850  -7.5022 8.029e-13 ***
## siteSH           0.87532    1.14406   0.7651  0.444844
## warm20:photo12   5.87109    2.23222   2.6302  0.008998 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 55.66 on 285 degrees of freedom
## Multiple R-squared: 0.5123,  Adjusted R-squared: 0.5054
## F-statistic: 74.84 on 4 and 285 DF,  p-value: < 2.2e-16
```

There are virtually no differences between the two analyses. However, note that the site effect, although not significant, is twice as important with the analysis with the intraspecific distances.

For the sake of comparison, let's compare with a standard regression.

```r
mod.lm <- lm(lday~warm*photo+site, data=thedata)
summary(mod.lm)
```

```
##
## Call:
## lm(formula = lday ~ warm * photo + site, data = thedata)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -36.122 -13.256   0.593  12.704  35.827
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)      57.122      2.069  27.606  < 2e-16 ***
## warm20          -19.838      2.656  -7.469  9.9e-13 ***
## photo12         -10.949      2.612  -4.192  3.7e-05 ***
## siteSH            2.123      1.846   1.150    0.251
## warm20:photo12    4.468      3.691   1.210    0.227
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 15.69 on 285 degrees of freedom
##   (58 observations deleted due to missingness)
## Multiple R-squared:  0.2824, Adjusted R-squared:  0.2723
## F-statistic: 28.03 on 4 and 285 DF,  p-value: < 2.2e-16
```

Interestingly, the result with the OLS is not very distinct from the PGLS. This suggests that the phylogeny as little to do for this dataset. This might be surprising as the estimated lambda in the PGLS is rather large (0.9). However, note that the interaction is not significant anymore.

Randomize individuals within species

Here, I modified slighlty your script to resample the individuals instead of the replicates. It seems to make more sense to me. Also, not sure why, but it seems that by initially cleaning the data and tree, then the number of observations retained in the comparative data table is the same after resampling as with the original data (not sure I can explain why though!).

```r
# randomize data within species - do this by changing the factor names
randat <- thedata[-c(1:nrow(thedata)),]
randat$randind <- character()
sphere <- unique(thedata$sp)
for (species in c(1:length(sphere))) {
  temp <- subset(thedata, sp==sphere[species])
  ind <- as.factor(as.vector(temp$ind))
  levels(ind) <- sample(levels(ind),length(levels(ind)))
  temp$randind <- ind
  randat <- rbind(randat, temp)
}
randat$randindX <- as.factor(paste(randat$randind, rep(1:4, 87), sep=""))
# PGLS in caper
compdatrand <- comparative.data(thetree, randat, randindX)
```

```
mod.rand <- pgls(lday~warm*photo, lambda="ML", data=compdatrand)
summary(mod.rand)
```

```
##
## Call:
## pgls(formula = lday ~ warm * photo, data = compdatrand, lambda = "ML")
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -319.71  -59.47    2.24   72.01  268.01
##
## Branch length transformations:
##
## kappa  [Fix]  : 1.000
## lambda [ ML]  : 0.904
##    lower bound : 0.000, p = < 2.22e-16
##    upper bound : 1.000, p = < 2.22e-16
##    95.0% CI   : (0.743, 0.953)
## delta  [Fix]  : 1.000
##
## Coefficients:
##                Estimate Std. Error  t value  Pr(>|t|)
## (Intercept)     55.4239    11.4948   4.8217 2.316e-06 ***
## warm20         -20.2290     1.4237 -14.2085 < 2.2e-16 ***
## photo12        -12.0440     1.3983  -8.6132 4.441e-16 ***
## warm20:photo12   5.9994     1.9783   3.0326  0.002647 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 95.62 on 286 degrees of freedom
## Multiple R-squared: 0.5745,  Adjusted R-squared:  0.57
## F-statistic: 128.7 on 3 and 286 DF,  p-value: < 2.2e-16
```

Other approach

This is an idea I got to compare the phylogenetic signal between the two trees for the residuals of the standard OLS.

```
residuals <- mod.lm$residuals[thetree$tip.label]
# With intraspecific branch lengths
phylosignal(residuals,thetree)
```

```
## [1] "Dropping taxa from the data because they are not present in the phylogeny:"
## [1] NA
## [1] "Dropping tips from the tree because they are not present in the data:"
##  [1] "VACMYR07_HF2" "VACMYR03_SH2" "VIBLAN07_SH1" "VIBLAN07_SH3"
##  [5] "VIBLAN02_SH2" "VIBLAN09_HF3" "VIBLAN11_HF1" "VIBLAN11_HF4"
##  [9] "VIBLAN11_HF3" "VIBLAN11_HF2" "VIBLAN03_HF1" "VIBLAN03_HF3"
## [13] "VIBLAN03_HF2" "QUERUB21_HF1" "QUERUB05_SH2" "QUERUB07_SH2"
## [17] "QUERUB06_SH2" "FAGGRA06_SH4" "FAGGRA07_SH1" "FAGGRA07_SH2"
## [21] "FAGGRA02_SH4" "FAGGRA02_SH2" "FAGGRA04_SH4" "FAGGRA04_SH2"
```

```
## [25] "FAGGRA05_SH4" "FAGGRA05_SH3" "FAGGRA05_SH2" "FAGGRA10_HF1"
## [29] "FAGGRA10_HF4" "FAGGRA11_HF3" "FAGGRA13_HF4" "FAGGRA03_HF1"
## [33] "FAGGRA03_HF4" "FAGGRA03_HF3" "FAGGRA03_HF2" "ALNINC02_HF4"
## [37] "ALNINC04_SH4" "PRUPEN06_SH2" "PRUPEN04_SH3" "PRUPEN04_SH2"
## [41] "PRUPEN03_SH1" "PRUPEN03_SH4" "PRUPEN03_SH3" "PRUPEN03_SH2"
## [45] "PRUPEN04_HF4" "SPIALB03_SH2" "SPIALB06_SH1" "SPIALB06_SH3"
## [49] "SPIALB08_HF1" "SPIALB08_HF4" "SPIALB08_HF3" "SPIALB08_HF2"
## [53] "POPGRA06_SH3" "POPGRA05_SH4" "ACEPEN02_HF3" "ACEPEN04_HF3"
## [57] "ACEPEN11_HF3" "ACEPEN09_SH1"

##             K PIC.variance.obs PIC.variance.rnd.mean PIC.variance.P
## 1 0.007962177         502907.4               1871974          0.001
##   PIC.variance.Z
## 1      -13.17114
```

```
# Without intraspecific branch lengths
phylosignal(residuals,thetree0)
```

```
## [1] "Dropping taxa from the data because they are not present in the phylogeny:"
## [1] NA
## [1] "Dropping tips from the tree because they are not present in the data:"
##  [1] "VACMYR07_HF2" "VACMYR03_SH2" "VIBLAN02_SH2" "VIBLAN07_SH1"
##  [5] "VIBLAN07_SH3" "VIBLAN03_HF1" "VIBLAN03_HF3" "VIBLAN03_HF2"
##  [9] "VIBLAN09_HF3" "VIBLAN11_HF1" "VIBLAN11_HF4" "VIBLAN11_HF3"
## [13] "VIBLAN11_HF2" "QUERUB07_SH2" "QUERUB06_SH2" "QUERUB05_SH2"
## [17] "QUERUB21_HF1" "FAGGRA03_HF1" "FAGGRA03_HF4" "FAGGRA03_HF3"
## [21] "FAGGRA03_HF2" "FAGGRA10_HF1" "FAGGRA10_HF4" "FAGGRA11_HF3"
## [25] "FAGGRA13_HF4" "FAGGRA02_SH4" "FAGGRA02_SH2" "FAGGRA04_SH4"
## [29] "FAGGRA04_SH2" "FAGGRA05_SH4" "FAGGRA05_SH3" "FAGGRA05_SH2"
## [33] "FAGGRA06_SH4" "FAGGRA07_SH1" "FAGGRA07_SH2" "ALNINC02_HF4"
## [37] "ALNINC04_SH4" "PRUPEN03_SH1" "PRUPEN03_SH4" "PRUPEN03_SH3"
## [41] "PRUPEN03_SH2" "PRUPEN04_SH3" "PRUPEN04_SH2" "PRUPEN06_SH2"
## [45] "PRUPEN04_HF4" "SPIALB08_HF1" "SPIALB08_HF4" "SPIALB08_HF3"
## [49] "SPIALB08_HF2" "SPIALB03_SH2" "SPIALB06_SH1" "SPIALB06_SH3"
## [53] "POPGRA05_SH4" "POPGRA06_SH3" "ACEPEN09_SH1" "ACEPEN11_HF3"
## [57] "ACEPEN04_HF3" "ACEPEN02_HF3"

##             K PIC.variance.obs PIC.variance.rnd.mean PIC.variance.P
## 1 0.004839129         839766.5               2312931          0.001
##   PIC.variance.Z
## 1      -19.42919
```

Interestingly, the phylogenetic signal is slightly higher when using the tree with intraspecific branch lengths.