Phylogenetic Least Squares

Testing for corellations using phylogenetic least squares

Why?

When species are your unit of inference and you undertake standard statistical analyses, you are implicitly assuming that the evolutionary distance between all species is equal (a.k.a. a phylogenetic "bush" rather than a tree) - we know this is not true!

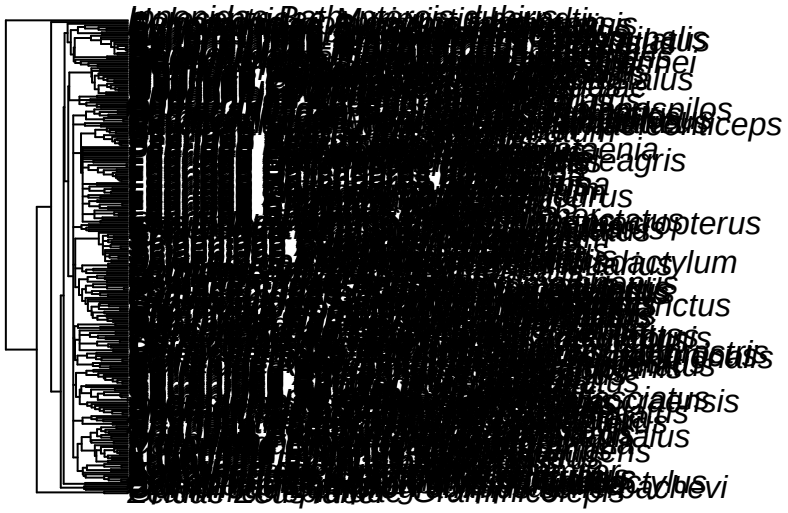
Phylogenetic least squares allows you to explore linear relationships between variables contingent upon covariances extrapolated from the phylogenetic tree.

Example

This is based on a paper that a UQ undergraduate student, Katja Kasimatis, wrote looking at variation in egg sizes for coral reef fishes (Kasimatis & Riginos 2016: Coral Reefs). She went on to get her PhD at the University of Oregon and was awarded the Dobzhansky Prize, one of the highest award for a finishing PhD student in evolutionary biology - way to go Katja!

```
library(ape)
library(geiger)
library(nlme)

fishdata <- read.csv("fishtraits_geography.csv", header=TRUE, row.names=1)  #data file
fishtree<-read.tree("Fish_chronogram.nex")
plot(fishtree)
```



Prepping the data

These example data files have been well prepared. Most likely yours will not. See `BlombergK.html` for some general tips on getting the taxa names in your data file and tree to match. Make sure to check for branch length also!

Analyses

PGLS can be used like any regression. You can model the effect on one predictive variable on one response variable, or you can make a more complex multiple regression model.

```
#reduce the tree to match your data
tree_reduced <-treedata(fishtree, fishdata, sort=TRUE, warnings=TRUE)  #lots of warnings!
```

```
## Warning in treedata(fishtree, fishdata, sort = TRUE, warnings = TRUE): The following tips were not f
##  Acanthuridae_Ctenochaetus_striatus
##  Acanthuridae_Naso_vlamingii
##  Ammodytidae_Ammodytes_hexapterus
##  Ammodytidae_Ammodytes_tobianus
##  Anarhichadidae_Anarhichas_lupus
##  Apogonidae_Apogon_maculatus
##  Apogonidae_Apogon_quadrifasciatus
##  Apogonidae_Phaeoptyx_conklini
##  Aulostomidae_Aulostomus_chinensis
```

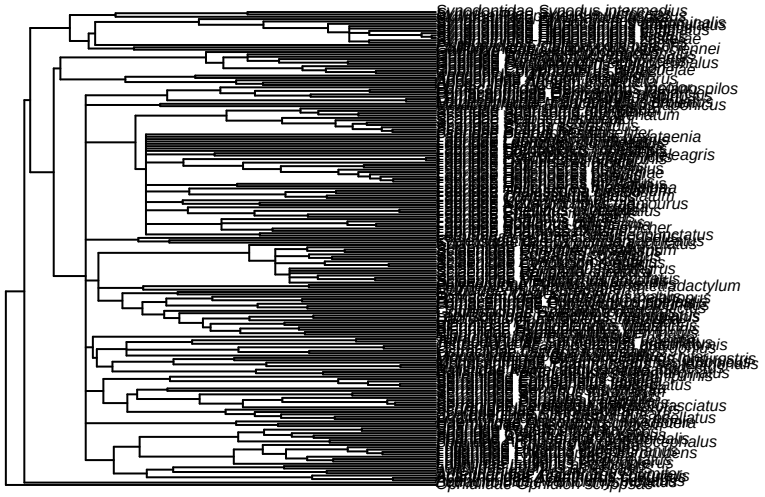
```
##  Balistidae_Balistes_capriscus
##  Batrachoididae_Halobatrachus_didactylus
##  Blenniidae_Parablennius_gattorugine
##  Bothidae_Arnoglossus_imperialis
##  Caesionidae_Pterocaesio_digramma
##  Callionymidae_Callionymus_schaapii
##  Callionymidae_Synchiropus_altivelis
##  Carapidae_Echiodon_cryomargarite
##  Cheilodactylidae_Nemadactylus_macropterus
##  Dactylopteridae_Dactylopterus_volitans
##  Dactyloscopidae_Platygillellus_rubrocinctus
##  Diodontidae_Diodon_hystrix
##  Fistulariidae_Fistularia_petimba
##  Gerreidae_Gerres_cinereus
##  Gobiidae_Entelurus_figaro
##  Gobiidae_Favonigobius_reichei
##  Gobiidae_Periophthalmus_barbarus
##  Gobiidae_Pomatoschistus_minutus
##  Gobiidae_Valenciennea_strigata
##  Grammicolepididae_Grammicolepis
##  Haemulidae_Pomadasys_perotaei
##  Holocentridae_Myripristis_berndti
##  Ipnopidae_Bathypterois_dubius
##  Kyphosidae_Microcanthus_strigatus
##  Labridae_Symphodus_roissali
##  Labrisomidae_Neoclinus_blanchardi
##  Latidae_Lates_calcarifer
##  Lophiidae_Lophius_budegassa
##  Lutjanidae_Lutjanus_sebae
##  Malacanthidae_Lopholatilus_chamaeleonticeps
##  Microdesmidae_Ptereleotris_zebra
##  Molidae_Mola_mola
##  Monacanthidae_Aluterus_monoceros
##  Moronidae_Dicentrarchus_labrax
##  Moronidae_Morone_americana
##  Moronidae_Morone_saxatilis
##  Mugilidae_Liza_sp
##  Mugilidae_Mugil_cephalus
##  Mullidae_Mullus_surmuletus
##  Ophidiidae_Lamprogrammus_scherbachevi
##  Ostraciidae_Lactoria_diaphana
##  Polynemidae_Pentanemus_quinquarius
##  Pomacanthidae_Pomacanthus_maculosus
##  Pomacentridae_Dascyllus_trimaculatus
##  Pomacentridae_Pomacentrus_pavo
##  Priacanthidae_Priacanthus_arenatus
##  Priacanthidae_Priacanthus_tayenus
##  Sciaenidae_Argyrosomus_regius
##  Scorpaenidae_Pontinus_longispinis
##  Sebastidae_Sebastes_rubrivinctus
##  Serranidae_Acanthistius_brasilianus
##  Serranidae_Epinephelus_aeneus
##  Serranidae_Holanthias_chrysostictus
##  Serranidae_Liopropoma_fasciatum
```

```
##  Serranidae_Niphon_spinosus
##  SerranidaeSerranidae_Serranus_accraensis
##  Siganidae_Siganus_vulpinus
##  Soleidae_Solea_solea
##  Sparidae_Spondyliosoma_cantharus
##  Syngnathidae_Nerophis_ophidion
##  Synodontidae_Synodus_foetens
##  Tetraodontidae_Fugu_rubripes
##  Tetraodontidae_Lagocephalus_laevigatus
##  Tripterygiidae_Tripterygion_delaisi
##  Zanclidae_Zanclus_cornutus
##  Zeidae_Zeus_faber
```

```
## Warning in treedata(fishtree, fishdata, sort = TRUE, warnings = TRUE): The following tips were not fo
##  Acanthuridae_Acanthurus_nigroris
##  Apogonidae_Apogon_cyanosoma
##  Apogonidae_Apogon_hungi
##  Apogonidae_Apogon_nitidus
##  Apogonidae_Apogon_ruppellii
##  Apogonidae_Cheilodipterus_lineatus
##  Apogonidae_Holapogon_maximus
##  Apogonidae_Siphamia_permutata
##  Apogonidae_Siphamia_roseigaster
##  Blenniidae_Chasmodes_bosquianus
##  Blenniidae_Chasmodes_saburrae
##  Blenniidae_Hypsoblennius_jenkins
##  Blenniidae_Lipophrys_pholis
##  Blenniidae_Lupinoblennius_nicholsi
##  Blenniidae_Ophiohlennius_steindachneri
##  Blenniidae_Plagiotremus_azalea
##  Callionymidae_Callionymus_enneactis
##  Chaetodontidae_Chaetodon_nippon
##  Ephippidae_Chaetodipterus_zonatus
##  Gobiesocidae_Diplecogaster_bimaculata
##  Gobiesocidae_Lepadogaster_purpurea
##  Gobiidae_Coryphopterus_nicholsii
##  Gobiidae_Elacantinus_figaro
##  Gobiidae_Gobiosoma_robustrum
##  Gobiidae_Lepidogohius_lepidus
##  Gobiidae_Paragobiodon_xanthosoma
##  Labridae_Cheilinus_unifasciatus
##  Labridae_Cirrhilabrus_temminckii
##  Labridae_Coris_variegata
##  Labridae_Halichoeres_poecilopterus
##  Labridae_Labropsis_micronesica
##  Labridae_Labropsis_xanthonota
##  Microdesmidae_Clarkichthys_bilineatus
##  Monacanthidae_Brachaluteres_jacksonianus
##  Monacanthidae_Eubalichthys_bucephalus
##  Monacanthidae_Rudarius_ercodes
##  Ophidiidae_Ophidion_marginatum
##  Pomacanthidae_Centropyge_aurantonotus
##  Pomacanthidae_Centropyge_debelius
##  Pomacanthidae_Genicanthus_semifasciatus
```

```
##  Pomacanthidae_Holacanthus_bermudensis
##  Pomacanthidae_Holacanthus_ciliaris
##  Pomacentridae_Acanthochromis_polyacanthus
##  Pomacentridae_Pomacentrus_amboinensis
##  Priacanthidae_Priacanthus_macracanthus
##  Scaridae_Hipposcarus_longiceps
##  Scaridae_Scarus_ovifrons
##  Serranidae_Epinephelus_fulvus
##  Serranidae_Epinephelus_guttaus
##  Serranidae_Mycteroperca_phenax
##  Serranidae_Serranus_psittacinus
##  Siganidae_Siganus_canaliculatus
##  Siganidae_Siganus_randalli
##  Sparidae_Sparidentex_hasta
##  Synanceiidae_Inimicus_japonicus
##  Syngnathidae_Hippocampus_breviceps
##  Syngnathidae_Hippocampus_hippocampus
##  Syngnathidae_Hippocampus_whitei
##  Syngnathidae_Syngnathus_rostellatus
##  Syngnathidae_Syngnathus_typhle
##  Synodontidae_Synodus_ulae
##  Tetrarogidae_Paracentropogon_rubripinnis
```

```
plot(tree_reduced$phy, cex = 0.5)  #simpler tree than uploaded tree - many branches pruned
```



```r
fishdata_reduced<-fishdata[row.names(tree_reduced$data),]  #simplified data file

#now derive a covariance matrix under the assumption of Brownian motion proportional to the tree
bm.tree<-corBrownian(phy=tree_reduced$phy)

#testing with PGLS - simple prediction: body size of fish predicts egg size
univariatePGLS<-gls(log(vol)~sl, correlation=bm.tree, data=fishdata_reduced)
```

```
## Warning in Initialize.corPhyl(X[[i]], ...): No covariate specified, species
## will be taken as ordered in the data frame. To avoid this message, specify a
## covariate containing the species names with the 'form' argument.
```

```r
summary(univariatePGLS)
```

```
## Generalized least squares fit by REML
##   Model: log(vol) ~ sl
##   Data: fishdata_reduced
##        AIC      BIC    logLik
##   572.9257 582.8207 -283.4629
##
## Correlation Structure: corBrownian
##  Formula: ~1
##  Parameter estimate(s):
## numeric(0)
##
## Coefficients:
##                  Value Std.Error   t-value p-value
## (Intercept) -0.6673109 0.4684036 -1.424649  0.1558
## sl           0.0091867 0.0021527  4.267488  0.0000
##
##  Correlation:
##    (Intr)
## sl -0.131
##
## Standardized residuals:
##        Min         Q1        Med         Q3        Max
## -2.4562907 -1.1312705 -0.8097445 -0.3714583  3.8036224
##
## Residual standard error: 1.374412
## Degrees of freedom: 202 total; 200 residual
```

```r
anova(univariatePGLS, test="F")  #this test asks how better is the model than one that is based on an i
```

```
## Denom. DF: 200
##             numDF   F-value p-value
## (Intercept)     1  0.762358  0.3836
## sl              1 18.211457  <.0001
```

```r
## -> body size highly significant!

#testing with PGLS - multiple predictions: egg type and latitude and their interactions
PGLSmodel<-gls(log(vol)~egg.type*latitude, correlation=bm.tree, data=fishdata_reduced)
```

```
## Warning in Initialize.corPhyl(X[[i]], ...): No covariate specified, species
## will be taken as ordered in the data frame. To avoid this message, specify a
## covariate containing the species names with the 'form' argument.
```

```
summary(PGLSmodel)  #complicated results because there are a lot of categories
```

```
## Generalized least squares fit by REML
##   Model: log(vol) ~ egg.type * latitude
##   Data: fishdata_reduced
##        AIC      BIC    logLik
##   575.0112 607.6381 -277.5056
##
## Correlation Structure: corBrownian
##  Formula: ~1
##  Parameter estimate(s):
## numeric(0)
##
## Coefficients:
##                                            Value Std.Error    t-value p-value
## (Intercept)                                 0.1790729 0.5608619  0.3192815  0.7499
## egg.typeegg.scatterer                      -2.3356158 1.0906442 -2.1415011  0.0335
## egg.typepelagic                            -0.7521775 0.4520378 -1.6639702  0.0977
## latitudetemperate                           0.7654596 0.3812920  2.0075415  0.0461
## latitudetropical                            0.0044751 0.2252996  0.0198627  0.9842
## egg.typeegg.scatterer:latitudetemperate    -0.2133778 1.0698265 -0.1994509  0.8421
## egg.typepelagic:latitudetemperate          -0.8514288 0.5152066 -1.6525970  0.1000
## egg.typeegg.scatterer:latitudetropical      0.3894406 0.9269147  0.4201472  0.6748
## egg.typepelagic:latitudetropical           -0.2773651 0.2672002 -1.0380424  0.3005
##
##  Correlation:
##                                           (Intr) egg.t. egg.ty lttdtm lttdtr
## egg.typeegg.scatterer                     -0.218
## egg.typepelagic                           -0.530  0.346
## latitudetemperate                         -0.230  0.135  0.345
## latitudetropical                          -0.223  0.121  0.288  0.335
## egg.typeegg.scatterer:latitudetemperate    0.082 -0.730 -0.122 -0.356 -0.119
## egg.typepelagic:latitudetemperate          0.171 -0.108 -0.356 -0.749 -0.249
## egg.typeegg.scatterer:latitudetropical     0.054 -0.774 -0.069 -0.081 -0.243
## egg.typepelagic:latitudetropical           0.190 -0.098 -0.319 -0.284 -0.843
##                                           egg.typgg.scttrr:lttdtm
## egg.typeegg.scatterer
## egg.typepelagic
## latitudetemperate
## latitudetropical
## egg.typeegg.scatterer:latitudetemperate
## egg.typepelagic:latitudetemperate          0.266
## egg.typeegg.scatterer:latitudetropical     0.831
## egg.typepelagic:latitudetropical           0.101
##                                           egg.typplgc:lttdtm
## egg.typeegg.scatterer
## egg.typepelagic
## latitudetemperate
## latitudetropical
## egg.typeegg.scatterer:latitudetemperate
```

```
## egg.typepelagic:latitudetemperate
## egg.typeegg.scatterer:latitudetropical   0.060
## egg.typepelagic:latitudetropical         0.262
##                                       egg.typgg.scttrr:lttdtr
## egg.typeegg.scatterer
## egg.typepelagic
## latitudetemperate
## latitudetropical
## egg.typeegg.scatterer:latitudetemperate
## egg.typepelagic:latitudetemperate
## egg.typeegg.scatterer:latitudetropical
## egg.typepelagic:latitudetropical         0.205
##
## Standardized residuals:
##        Min         Q1        Med         Q3        Max
## -2.2925602 -0.9568840 -0.6145851 -0.2759798  3.3784025
##
## Residual standard error: 1.389872
## Degrees of freedom: 202 total; 193 residual
```

```
anova(PGLSmodel, test="F")  # look at p values here - only egg type is significant, so latitude and int
```

```
## Denom. DF: 193
##                   numDF  F-value p-value
## (Intercept)           1 0.745492  0.3890
## egg.type              2 5.993939  0.0030
## latitude              2 2.260274  0.1071
## egg.type:latitude     4 0.968887  0.4257
```