ELETRONIC SUPPLEMENTARY MATERIAL
Background Noise as a Selective Pressure: Stream-breeding Anurans Call at Higher Frequencies

David Lucas Röhr; Gustavo B. Paterno; Felipe Camurugi; Flora A. Juncá; Adrian A. Garda;

December 17, 2014

Contents

1. Packages versions: **2**

2. Data structure: **3**

2.1 Species data . . . 3

2.2 Phylogentic tree . . . 3

2.3 Summary metrics for Dominant frequency and Sout-vent length . . . 5

3. Phylogenetic signal **6**

3.1 Domiant Frequency and Snout-vent length . . . 6

3.2 Checking the phylogenetic signal of the residuals from stantard OLS regression . . . 6

4. Data analysis **7**

4.1 Data preparation: . . . 7

4.2 Phylogenetic generalized least square model (PGLS) . . . 7

4.3 ANOVA table . . . 8

4.4 Confidence interval for lambda estimation . . . 8

4.5 Model diagnostic: . . . 9

4.5.1 Standard graphic methods for model diagnostics . . . 9

4.5.2 Phylogenetic signal of model residuals . . . 9

4.6 Model comparison: OSL vs PGLS . . . 10

5. References **11**

6. Source code: **11**

This doccument follows the principles of reproducible research Peng, 2011. All Data and code required to repeat the analysis bellow are linked at Github. To dowanload the **source code** used to generate all figures, tables and analysis in the paper, please see: source code. This documment was generated in **R studio** with `kintr` package.

1. Packages versions:

We used R version 3.1.2 (2014-10-31) and the following packages:

```
library(ape);library(caper);library(knitr)
library(dplyr);library(ggplot2);library(picante);library(gridExtra)
```

Please check the Packages versions, for details.

2. Data structure:

2.1 Species data

- To download **raw data**: link.
- See Appendix 1 to download complete table with references used.

The species dataset contains six variables (see Methods for detailed information on data collection).

variable	discription
sp	species name
environment	Reproductive environment (lentic or lotic)
DF	Dominant frequency (hertz)
SVL	snout-vent length (mm)
logDF	log of dominant frequency (DF)
logSVL	log of snout vent length (SVL)

Last six rows of the species dataset:

	sp	environment	DF	SVL	logDF	logSVL
504	Bufo_retiformis	still	3113	47	8.043342	3.850148
505	Bufo_houstonensis	still	2151	77	7.673688	4.343805
506	Pelophryne_misera	still	4000	21	8.294050	3.044522
507	Ansonia_longidigita	running	3500	50	8.160518	3.912023
508	Ansonia_hanitschi	running	5700	32	8.648222	3.465736
509	Ansonia_platysoma	running	8000	25	8.987197	3.218876

2.2 Phylogentic tree

The phylogenetic tree used in this paper was pruned from: Pyron and Wiens (2011) anura super tree.
To dowanload the pruned tree with study species (509): Study Tree.

```
##
## Phylogenetic tree with 509 tips and 508 internal nodes.
##
## Tip labels:
##  Hadromophryne_natalensis, Heleophryne_purcelli, Heleophryne_regis, Calyptocephallela_gayi, Neobatra
## Node labels:
##  209.59, 206.29, 195.63, 151.99, 47.93, 9.02, ...
##
## Rooted; includes branch lengths.
```

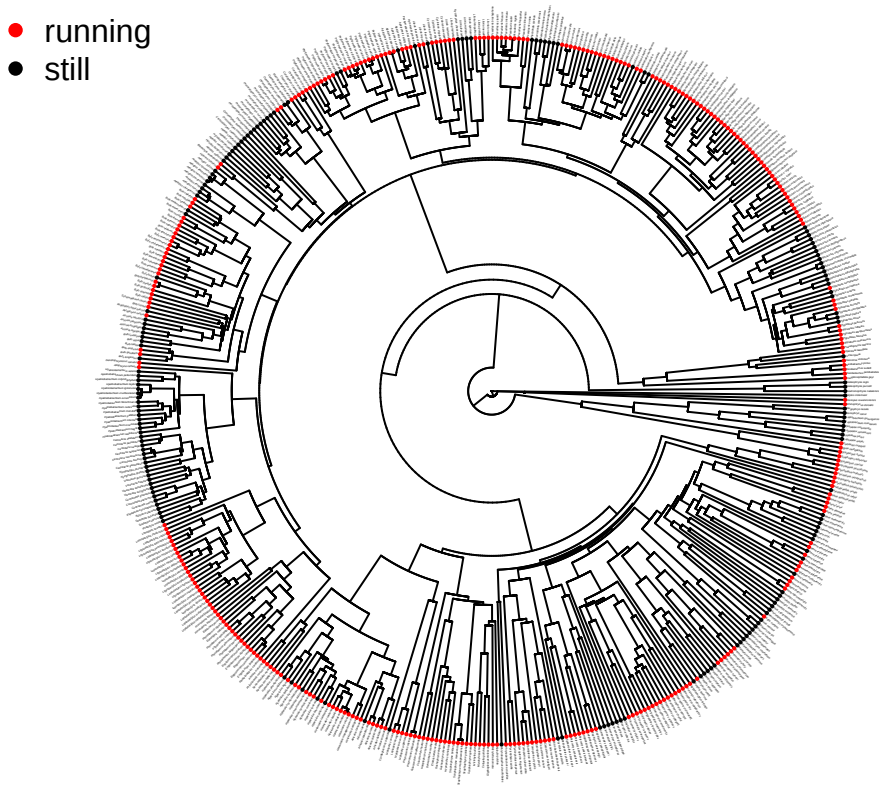


Figure S1: Phylogeny for 509 anuran species sampled in this study extracted from Pyron and Wiens (2011) original tree. Black circles represent pond-breeding species (N = 332) and red circles stream-breeding species (N= 177)

2.3 Summary metrics for Dominant frequency and Sout-vent length

environment	meanDF	seDF	meanSVL	seSVL
running	3377.322	2036.938	41.70932	20.03995
still	2180.557	1263.536	51.27440	29.35337

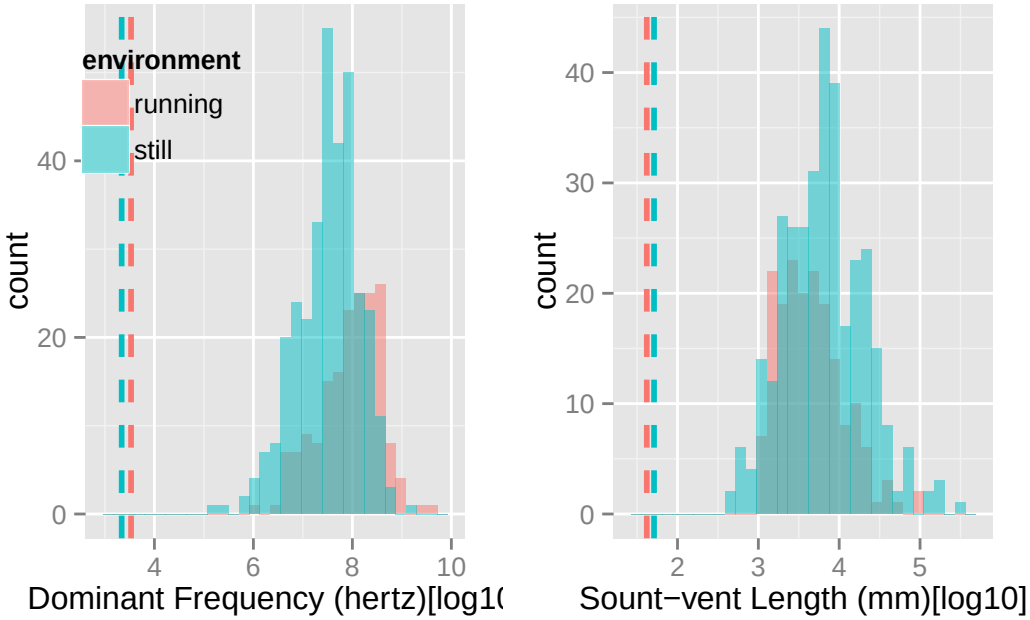


Figure S2: Distribution histograms for Dominant Frequency (log10) and Snout-vent length (log10)

3. Phylogenetic signal

3.1 Domiant Frequency and Snout-vent length

We used K statistics to test the phylogentic signal for Dominant Frequency (*logDF*) and Snout-vent length(*logSVL*)(for details about the method, see Blomberg et al (2003)).

```
k.signal <- multiPhylosignal(select(comp.data$data,logDF,logSVL),comp.data$phy,reps=999)
kable(k.signal)
```

	K	PIC.variance.obs	PIC.variance.rnd.mean	PIC.variance.P	PIC.variance.Z
logDF	0.3660099	0.0075415	0.0292342	0.001	-4.365948
logSVL	0.4387480	0.0032133	0.0149648	0.001	-4.661169

Dominant Frequency and Snout-Vent Length show significant phylogenetic signal, however, K values are low.

3.2 Checking the phylogenetic signal of the residuals from stantard OLS regression

In order to check the need to include the phylogeny in our analysis, first it is important to check if there is phylogenetic signal in the residuals of an Ordinary Least Square regression (OSL) (Kamilar & Cooper, 2013; Freckleton, 2009).

```
mod.osl <- lm(logDF ~ environment*logSVL,anura.data)
# Extracting residuals from the model:
comp.data$data$lm.res <- residuals(mod.osl)
osl.resi.sig <- phylosignal(comp.data$data$lm.res,reps=999,comp.data$phy)
kable(osl.resi.sig)
```

K	PIC.variance.obs	PIC.variance.rnd.mean	PIC.variance.P	PIC.variance.Z
0.1472979	0.0096954	0.0146178	0.003	-1.777386

Because the residuals from OSL regression show phylogenetic signal k = 0.15, it is necessary to correct for phylogenetic non-independence in data.

4. Data analysis

We used a phylogenetic generalized least square model (PGLS) with dominant frequency as the response variable and reproduction habitat (lentic/lotic) and SVL as the explanatory variables to test if dominant frequency was affected by reproduction environment. Dominant frequencies and body sizes were log transformed (log10) before the analysis. To optimize branch length transformation, the lambda value was set by maximum likelihood (see Freckleton et al., 2002; Orme et al., 2013 for details). PGLS analysis were performed with the function `pgls` from the package `caper`.

4.1 Data preparation:

Using the function `comparative.data` we combined our phylogenie with the species dataset

```
## Comparative dataset of 509 taxa:
## Phylogeny: study.tree
##    509 tips, 508 internal nodes
##    chr [1:509] "Hadromophryne_natalensis" "Heleophryne_purcelli" ...
## VCV matrix present:
##    VCV.array [1:509, 1:509, 1:25] 47.9 3.3 3.3 3.3 3.3 ...
## Data: anura.data
##    $ environment: Factor w/ 2 levels "running","still": 1 1 1 2 2 2 2 2 1 2 ...
##    $ DF         : int [1:509] 1500 2250 1800 866 816 1509 1300 2676 2735 2260 ...
##    $ SVL        : num [1:509] 45 47 43 120 39.6 40 55 27 25.2 35 ...
##    $ logDF      : num [1:509] 7.31 7.72 7.5 6.76 6.7 ...
##    $ logSVL     : num [1:509] 3.81 3.85 3.76 4.79 3.68 ...
```

4.2 Phylogenetic generalized least square model (PGLS)

Fitting pgls model with with lambda adjusted by maximum likelihood:

```
mod.pgls <- pgls(logDF ~ environment*logSVL, data=comp.data,lambda="ML")
summary(mod.pgls)
```

```
##
## Call:
## pgls(formula = logDF ~ environment * logSVL, data = comp.data,
##     lambda = "ML")
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.073524 -0.014980 -0.001757  0.010458  0.060338
##
## Branch length transformations:
##
## kappa  [Fix]  : 1.000
## lambda [ ML]  : 0.889
##    lower bound : 0.000, p = < 2.22e-16
##    upper bound : 1.000, p = < 2.22e-16
##    95.0% CI   : (0.823, 0.933)
## delta  [Fix]  : 1.000
##
```

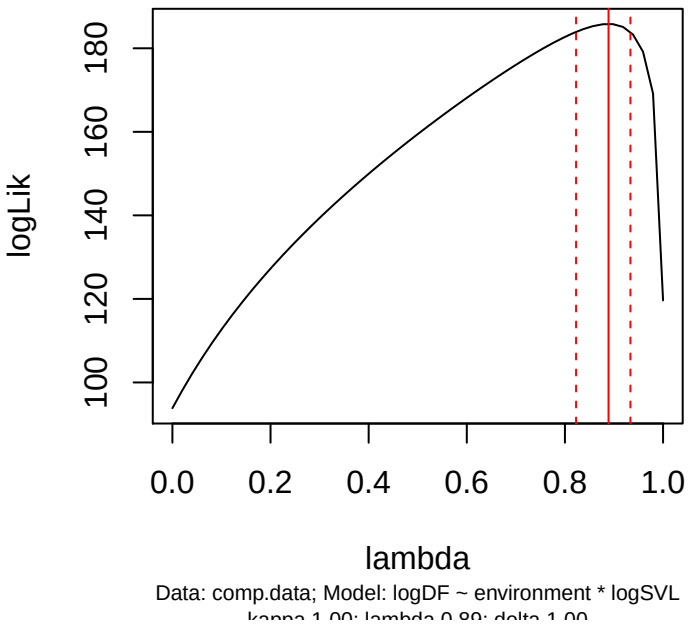
```
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                4.874025   0.190792 25.5462   <2e-16 ***
## environmentstill          -0.182756   0.173309 -1.0545   0.2922
## logSVL                    -0.918595   0.093136 -9.8630   <2e-16 ***
## environmentstill:logSVL    0.060128   0.104687  0.5744   0.5660
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.02088 on 505 degrees of freedom
## Multiple R-squared: 0.3825,  Adjusted R-squared: 0.3788
## F-statistic: 104.3 on 3 and 505 DF,  p-value: < 2.2e-16
```

4.3 ANOVA table

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
environment	1	0.0144038	0.0144038	33.0318903	0.0000000
logSVL	1	0.1218527	0.1218527	279.4422042	0.0000000
environment:logSVL	1	0.0001428	0.0001428	0.3274965	0.5673916
Residuals	505	0.2202086	0.0004361	NA	NA

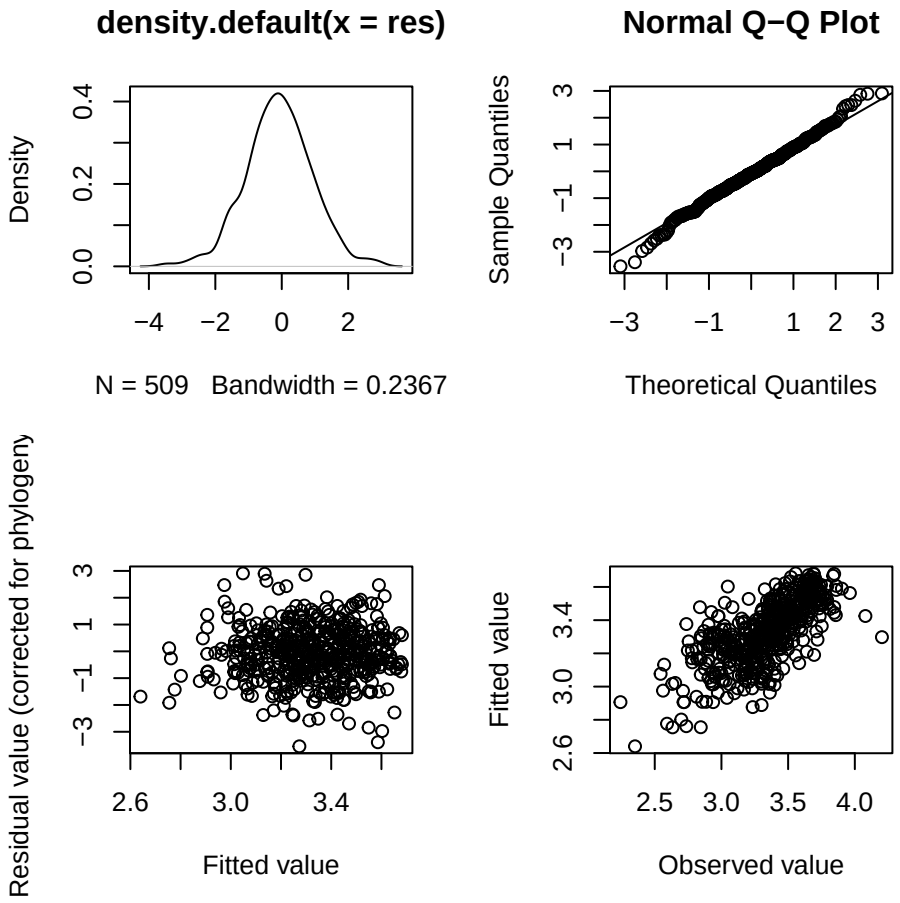
4.4 Confidence interval for lambda estimation

```
profile.lambda <- pgls.profile(mod.pgls)
plot(profile.lambda)
```



4.5 Model diagnostic:

4.5.1 Standard graphic methods for model diagnostics



Residulas do not show any tendency.

4.5.2 Phylogenetic signal of model residuals

After performing PGLS analysis it is important to check the phylogenetic signal of model residuals.

```
k.residuals <- phylosignal(mod.pgls$phyres,reps=999,comp.data$phy)
kable(k.residuals)
```

K	PIC.variance.obs	PIC.variance.rnd.mean	PIC.variance.P	PIC.variance.Z
0.1119063	2.39e-05	2.94e-05	0.105	-1.052582

Results above shows that the residuals do not present significant phylogenetic signal.

4.6 Model comparison: OSL vs PGLS

```
kable(AIC(mod.osl,mod.pgls))
```

	df	AIC
mod.osl	5	671.3566
mod.pgls	4	-363.6674

AIC comparison shows that PGLS model has much lower AIC value round(-363.6674412) then OSL model round(671.3565725). Thus, PGLS model is a better fit for the data.

5. References

1. Blomberg, S. P., Garland, T., & Ives, A. R. (2003). Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution, 57(4), 717-745.

2. Freckleton, R. P., Harvey, P. H., & Pagel, M. (2002). Phylogenetic analysis and comparative data: a test and review of evidence. The American Naturalist, 160(6), 712-726.

3. Freckleton, R. P. (2009). The seven deadly sins of comparative analysis. Journal of Evolutionary Biology, 22(7), 1367-1375.

4. Orme, D., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., Isaac, N. and Pearse, W. (2013). caper: Comparative Analyses of Phylogenetics and Evolution in R. R package version 0.5.2. http://CRAN.R-project.org/package=caper

5. Kamilar, J. M., & Cooper, N. (2013). Phylogenetic signal in primate behaviour, ecology and life history. Philosophical Transactions of the Royal Society B: Biological Sciences, 368(1618), 20120341.

6. Pyron, A. R., & Wiens, J. J. (2011). A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. Molecular Phylogenetics and Evolution, 61(2), 543-583.

7. Peng, R. D. (2011). Reproducible research in computational science. Science (New York, Ny), 334(6060), 1226.

When using the **data available** in this paper, please cite the original publication.
Contact davidlucasr@yahoo.com.br for any further information.

6. Source code:

```
### Source Code for:
################################################################################
### Background Noise as a Selective Pressure: Stream-breeding Anurans  Call at
### Higher Frequenciesfor Phylogenetic Analysis of Pitches in Anura
### Journal of Evolutionary Biology (2015)
### Authors: David Lucas Röhr; Gustavo B. Paterno; Felipe Camurugi; Flora A. Juncâ;
### Adrian A. Garda;
### Code by: Gustavo Paterno
### last updade: 23.01.15
################################################################################

### Packages used (for specific versions see Supp. Mat.):
library(ape);library(caper);library(dplyr);
library(picante);library(RCurl);library(foreign)


#################################### DATA ######################################
################################################################################
### Loading Data (raw data from Github repository)
```

```r
# Species data:
url.species <- paste("https://raw.githubusercontent.com",
"/paternogbc/2015_Rohr_et_al_JEB/master/data/raw_data.csv",sep="")
myData <- getURL(url.species,ssl.verifypeer = FALSE)
mat <- read.csv(textConnection(myData))
str(mat)

# Phylogeny:
url.phylogeny <- paste("https://raw.githubusercontent.com",
"/paternogbc/2015_Rohr_et_al_JEB/master/phylogeny/amph_2014.tre",sep="")
myPhy <- getURL(url.phylogeny,ssl.verifypeer = FALSE)
tree <- read.tree(textConnection(myPhy))
str(tree)

# Pruned phylogeny:
tree.drop  <- drop.tip(tree,as.character(mat[,1]))
study.tree <- (drop.tip(tree,tree.drop$tip.label))
study.tree$node.label <- makeLabel(study.tree)$node.label

### Checking for absent species in data:
sum(sort(study.tree$tip.label) != sort(mat$sp))

### Data preparation for pgls:
comp.data <- comparative.data(phy=study.tree,data=mat,names.col="sp",vcv=T,vcv.dim=3)
comp.data.runn <- comparative.data(phy=study.tree,data=subset(mat,environment=="running"),
                                   names.col="sp",vcv=T,vcv.dim=3)
comp.data.still <- comparative.data(phy=study.tree,data=subset(mat,environment=="still"),
                                    names.col="sp",vcv=T,vcv.dim=3)

#################################### Analysis ####################################
##################################################################################

### Table 1 (Phylogenetic signal logSVL and logDF)
k.signal <- multiPhylosignal(select(comp.data$data,logDF,logSVL),comp.data$phy,reps=999)
k.signal

### Table 2 (PGLS with with lambda adjusted by maximum likelihood)
mod.pgls <- pgls(logDF ~ environment*logSVL, data=comp.data,lambda="ML")
anova(mod.pgls)
summary(mod.pgls)

### Model diagnostics:
plot(mod.pgls)

### No interaction environment:logSVL model:
mod.pgls2 <- pgls(logDF ~ logSVL+environment, data=comp.data,lambda="ML")
summary(mod.pgls2)
### Regressions coefficients:
coef.pgls2 <- coef(mod.pgls2)
running.coef <- coef.pgls2[c(1,2)]
still.coef <- c(c(coef.pgls2[1]+coef.pgls2[3]),coef.pgls2[2])
### Mean difference between still and running environments (Hertz)
diff.intercep <- exp(running.coef[1]) - exp(still.coef[1])
```

```
################################## Figure 1 ######################################
##################################################################################

### Figure 1 (dispersion plot with original data (logDF ~ environment + lofSVL))

par(mfrow=c(1,1),las=1,bty="l",oma=c(1,1,1,1))
plot(logDF~ logSVL,col="black",pch=2,data=subset(mat,mat$environment=="still"),
     ylim=c(5,10),xlim=c(2.5,5.5),
     yaxp=c(4,10,6),xaxp=c(2.8,5.8,6),
     ylab="Log dominant frequency (lnDF)",xlab="Log snout-vent length (lnSVL)",las=1,
     yaxp=c(5,10,10),xaxp=c(2.5,5.5,15),
     frame="F",cex=1.2,cex.lab=1.1)
points(logDF~ logSVL,cex=1.2,
       col="red",pch=1,data=subset(mat,mat$environment=="running"))
text(x=5.1,y=10, "Y = -0.87x + C     ",cex=1)
text(x=5.1,y=9.6,"C = 11.05 (running)",cex=1)
text(x=5.1,y=9.2,"C = 10.86 (still)  ",cex=1)

### Regression lines:
SVL.ran.running <- with(subset(mat,environment=="running"),range(logSVL))
SVL.ran.still <- with(subset(mat,environment=="still"),range(logSVL))
x.still <- seq(SVL.ran.running[1],SVL.ran.running[2],0.01)
x.running <- seq(SVL.ran.still[1],SVL.ran.still[2],0.01)

ypred.running <- running.coef[1] + running.coef[2]*x.running
ypred.still <- still.coef[1] + still.coef[2]*x.still

lines(x.running,ypred.running,col="red",lwd=2)
lines(x.still,ypred.still,col="black",lwd=2)

################################## Figure S1 #####################################
##################################################################################

### Figure S1: (Study tree)
comp.data$data$sp <- comp.data$tip.label
plot(comp.data[[1]],"fan",show.tip.label=T,cex=0.08,label.offset=1.2,
     lab4ut="axial")
tiplabels(frame="circle",col=comp.data$data$environment,
          pch=c(16,16),cex=0.3)
legend(legend=c("running","still"),pch=c(16,16),col=c("red","black"),
       "topleft",bty="n")
```