Supporting Information 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;

April, 2015

Contents

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 3.2.0 (2015-04-16) 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
fam	family
sp	species name
environment	Reproductive environment (still or flowing)
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:

	fam	sp	environment	DF	SVL	logDF	logSVL
504	Rhacophoridae	Rhacophorus_schlegelii	still	1900	43	7.549609	3.761200
505	Rhinophrynidae	Rhinophrynus_dorsalis	still	1525	75	7.329750	4.317488
506	Scaphiopodidae	Scaphiopus_couchii	still	1650	72	7.408531	4.276666
507	Scaphiopodidae	Scaphiopus_holbrookii	still	1425	78	7.261927	4.356709
508	Scaphiopodidae	Scaphiopus_hurterii	still	1500	67	7.313220	4.204693
509	Scaphiopodidae	Spea_multiplicata	still	1300	49	7.170120	3.891820

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.
```

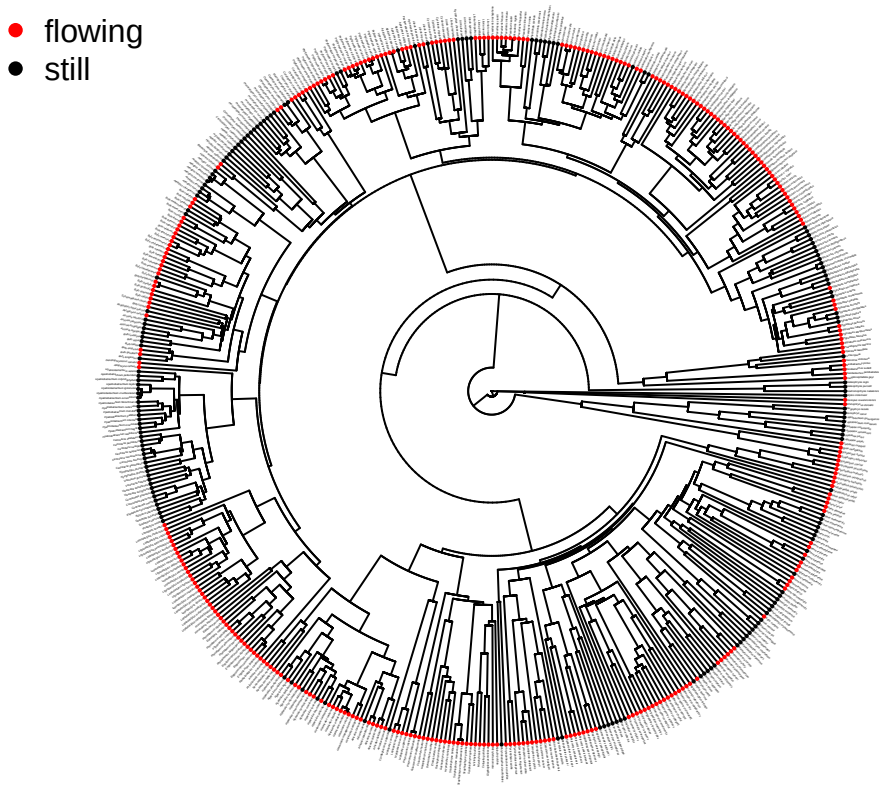

Figure S1: Phylogeny for 509 anuran species sampled in this study extracted from Pyron and Wiens (2011) original tree. Black circles represent pond-breeding species (N = 332) and red circles stream-breeding species (N= 177)

2.3 Summary metrics for Dominant frequency and Sout-vent length

environment	meanDF	sdDF	meanSVL	sdSVL
flowing	3377.322	2036.938	41.70932	20.03995
still	2180.557	1263.536	51.27440	29.35337

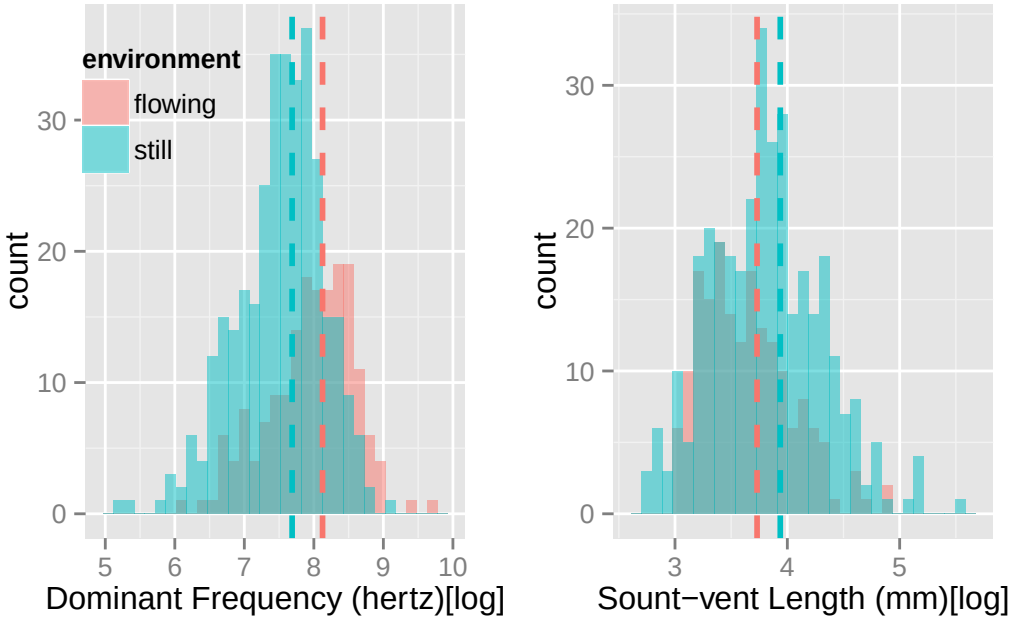

Figure S2: Distribution histograms for Dominant Frequency (log) and Snout-vent length (log)

2.4 Testing for bias in Dominant Frequency between old and new publications

To construct a large data set, we difened some practical, pre-established criteria of data inclusion that enabled us to use more than 500 species. We agree that the most recent publication is not always the best, however the recent improvement of recording equipment and sound analysis software is unquestionable. Furthermore, judging the merit of multiple papers takes a large amount of time because factors such as the number of individuals recorded, geographical location of individuals included and recording equipment used must be considered. We believe that in multiple occasions these decisions can be subjective, leading to uncertainty about possible bias in the resulting dataset. Our pre-established criterion explicitly avoids this, leaving no space to speculations about such biases. To be sure, we collected data on dominant frequency from different papers for 30 randomly-selected species from our dataset and found no significant differences between the most recent compared to the older data:

```
##
##  Paired t-test
##
## data:  DFpaired.data[, 1] and DFpaired.data[, 2]
## t = 1.0339, df = 31, p-value = 0.3092
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -101.573  310.448
## sample estimates:
## mean of the differences
##                104.4375
```

This provides strong evidence that our criterion does not affect our results and conclusions.

2.5 Testing correlation between mean and maximum SVL

In large comparative studies it is always hard to choose the best approach to represent species, however, we do not believe that the choice of using maximum body size affected the results presented in the manuscript.

1. The decision to use maximum SVL was a pre-established criterion applied to all species, thus we find very unlikely that this choice could cause any non-random tendency in our analysis.
2. We decided to use maximum SVL because data for most species was available in multiple publications and it would therefore be hard to calculate a mean value from many sources. It would be possible to calculate the weighted mean among publications, however this approach would be unpractical considering the number of species included (509). Furthermore, we could also use the mean value provided from one specific publication, however, this could include subjectivity to the analysis, diminishing its reproducibility. Thus, choosing the maximum value among papers excludes bias and represents the adult potential size of species considering multiple publications.
3. Finally, in order to consider the concerns about our choice of maximum SVL:
    1. Re-collected data for maximum and mean SVL for 30 random species of our previous survey.
    2. Performed a Pearson's correlation analysis between maximum and mean SVL (see results below).

SVL_m	ax SVL_me	an
SVL_max	1.0000000	0.9817694
SVL_mean	0.9817694	1.0000000

The correlation coefficient between maximum and mean SVL values was higher than 0.95, thus we truly believe that our choice of maximum SVL by no means affected any aspect of our main results or final conclusions.

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	0.0291375	0.001	-4.434511
logSVL	0.4387480	0.0032133	0.0149021	0.001	-4.734528

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
0.1166795	0.0103429	0.0146352	0.007	-1.521814

Because the residuals from OSL regression show phylogenetic signal k = 0.12, 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 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

```
comp.data <- comparative.data(phy=study.tree,data=anura.data,names.col="sp",vcv=T,vcv.dim=3)
```

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
## -0.169581 -0.034834 -0.003471  0.024159  0.162009
##
## 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)             11.222857   0.439315 25.5462   <2e-16 ***
## environmentstill        -0.420810   0.399058 -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: 0.04808 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	0.0763674	0.0763674	33.0318903	0.0000000
logSVL	1	0.6460504	0.6460504	279.4422042	0.0000000
environment:logSVL	1	0.0007571	0.0007571	0.3274965	0.5673916
Residuals	505	1.1675238	0.0023119	NA	NA

4.4 Lambda estimation

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

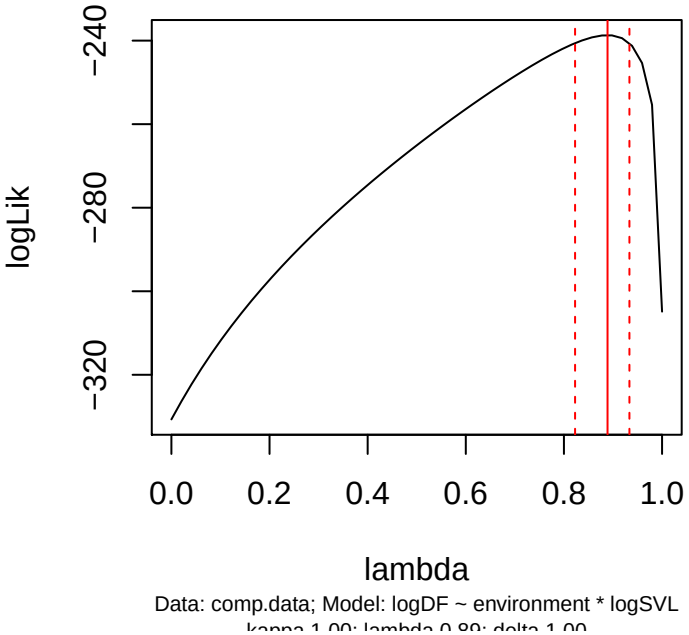

Figure S3: Confidence interval for lambda estimation

4.5 Model diagnostic

4.5.1 Diagnostic graphs

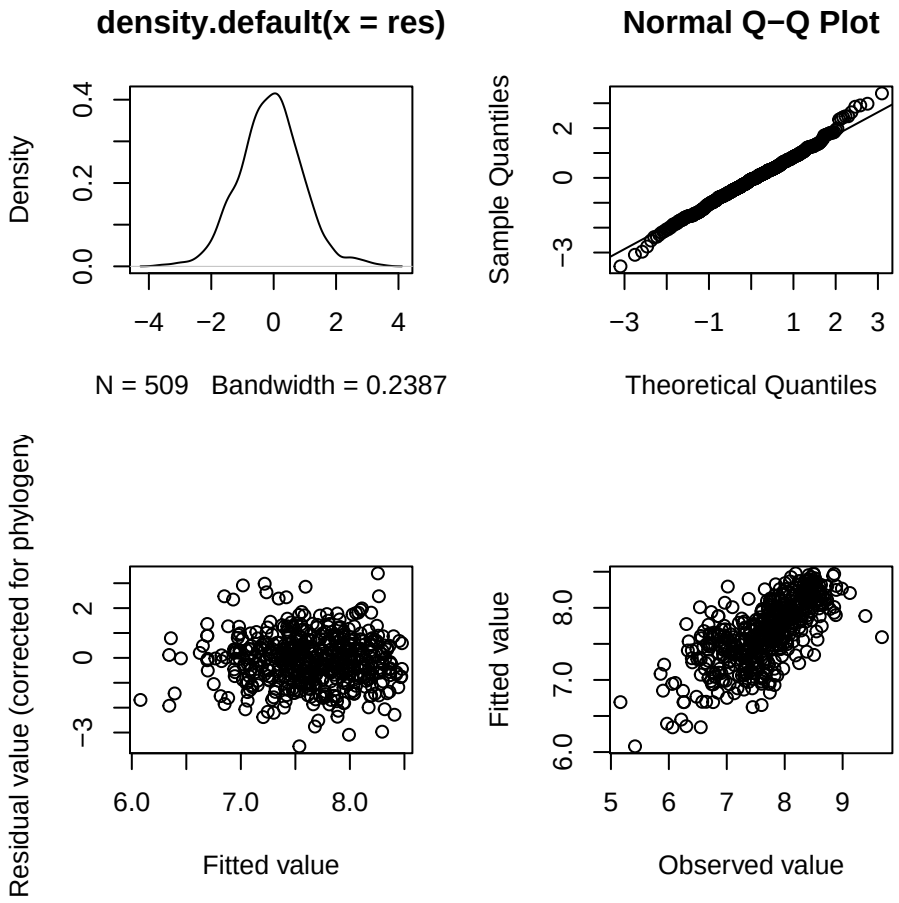

Figure S4: Standard graphic methods for model diagnostics

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
0.1054035	0.0001429	0.0001536	0.402	-0.3904951

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	485.3776

AIC comparison shows that PGLS model has much lower AIC value (485) tham OSL model (671). Thus, PGLS model is a better fit for the data.

4.7 Phylogenetic generalized least square model (PGLS) within families

To test if the environment effect on Dominant Frequency is independent of taxonomic group, we performed pgls models for the three families in this dataset with more tham 30 species (Bufonidae, Ranidae and Hylidae).

4.7.1 Bufonidae:

Fitting pgls model with with lambda adjusted by maximum likelihood:

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

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
environment	1	0.0017044	0.0017044	0.6756231	0.4154320
logSVL	1	0.1610608	0.1610608	63.8456240	0.0000000
environment:logSVL	1	0.0029574	0.0029574	1.1723362	0.2846895
Residuals	45	0.1135197	0.0025227	NA	NA

4.7.2 Ranidae:

Fitting pgls model with with lambda adjusted by maximum likelihood:

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

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
environment	1	0.1111967	0.1111967	30.416038	0.0000037
logSVL	1	0.0044243	0.0044243	1.210182	0.2790242
environment:logSVL	1	0.0094783	0.0094783	2.592628	0.1166087
Residuals	34	0.1242992	0.0036559	NA	NA

4.7.2 Hylidae:

Fitting pgls model with with lambda adjusted by maximum likelihood:

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

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
environment	1	0.0196581	0.0196581	7.828985	0.0057858
logSVL	1	0.1412524	0.1412524	56.254873	0.0000000
environment:logSVL	1	0.0003487	0.0003487	0.138868	0.7099115
Residuals	157	0.3942170	0.0025109	NA	NA

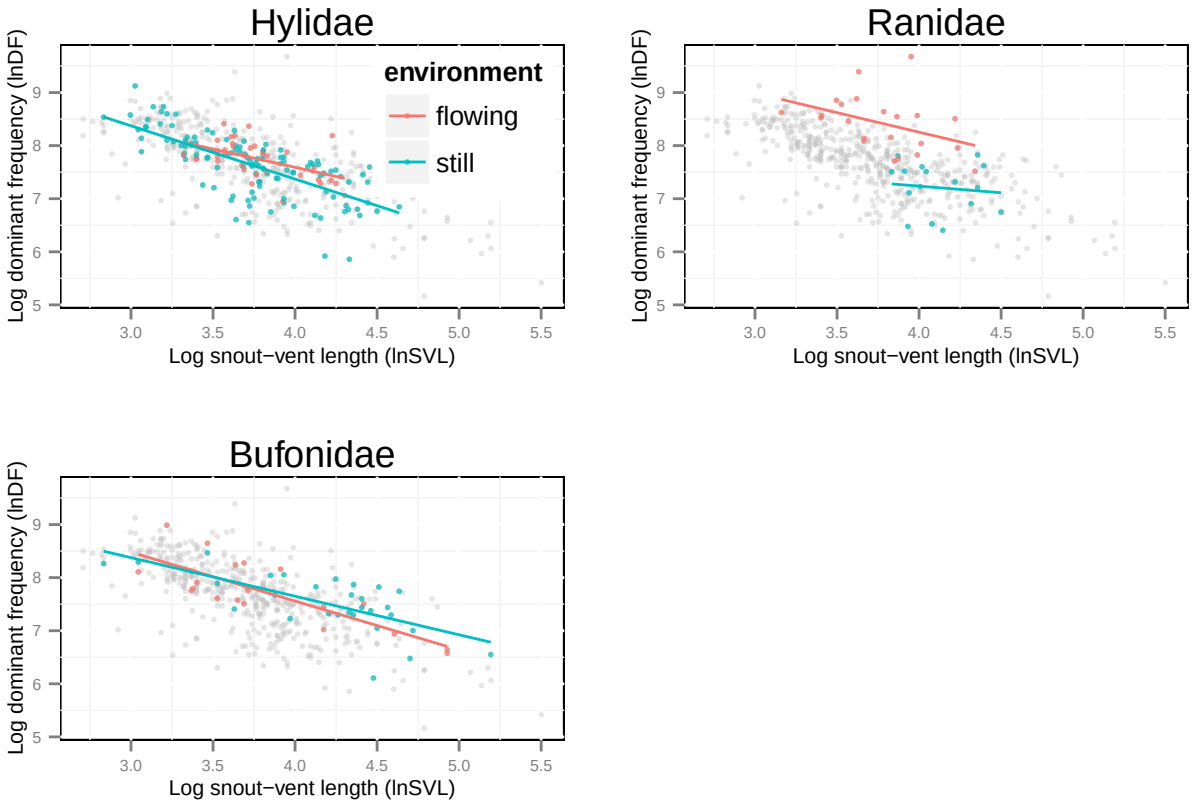

Figure S5: Regression plots for three anura families (Hylidae, Ranidae, Bufonidae).

5. References

1. 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.

2. 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.

3. 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

4. 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.

5. 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.

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

6. Source Code

```
### Source Code for:
##############################################################################################
### Background Noise as a Selective Pressure: Stream-breeding Anurans  Call at
### Higher Frequencies for Phylogenetic Analysis of Pitches in Anura
### Journal of Animal Ecology (2015)
### Authors: David Lucas Röhr; Gustavo B. Paterno; Felipe Camurugi; Flora A. Juncâ;
### Adrian A. Garda;
### Code by: Gustavo Paterno
### last updade: 12.05.15
##############################################################################################

### Packages used (for specific versions see Supp. Mat.):
library(ape);library(caper);library(dplyr);library(gridExtra)
library(picante);library(RCurl);library(foreign);library(ggplot2)

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

# Species data:
url.species <- paste("https://raw.githubusercontent.com",
    "/paternogbc/2015_Rohr_et_al_JAEcol/master/data/data_raw.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_JAEcol/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[,2]))
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=="flowing"),
                                   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=="flowing"))
text(x=5.1,y=10, "Y = -0.87x + C     ",cex=1)
text(x=5.1,y=9.6,"C = 11.05 (flowing)",cex=1)
text(x=5.1,y=9.2,"C = 10.86 (still)  ",cex=1)

### Regression lines:
SVL.ran.running <- with(subset(mat,environment=="flowing"),range(logSVL))
SVL.ran.still <- with(subset(mat,environment=="still"),range(logSVL))
x.running <- seq(SVL.ran.running[1],SVL.ran.running[2],0.01)
x.still<- 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("flowing","still"),pch=c(16,16),col=c("red","black"),
       "topleft",bty="n")

########################## Figure S5 (within family Pgls) ######################
#############################################################################################

### Bufonidae:
mat.buf <- filter(mat,fam=="Bufonidae")

g.fam1 <- ggplot(mat, aes(y=logDF,x=logSVL))+
    geom_point(colour="gray",size=3,alpha=.4)+
    geom_point(data=mat.buf,alpha=.7,
               aes(y=logDF,x=logSVL,colour=environment),size=3)+
    geom_smooth(data=mat.buf,se=F,
                aes(y=logDF,x=logSVL,colour=environment),method="lm")+
    theme(panel.background = element_rect(fill="white",colour="black"),
          axis.text = element_text(size=14),
          axis.title = element_text(size=16),
          legend.position = "none")+
    ggtitle("Bufonidae")+
    ylab("Log dominant frequency (lnDF)")+
    xlab("Log snout-vent length (lnSVL)")+
    annotate("text", x = c(3.4,3), y = c(5.5,6),
             label = c("p = 0.415 (environment)","N = 49"))

### Ranidae:
mat.ran <- filter(mat,fam=="Ranidae")

g.fam2 <- ggplot(mat, aes(y=logDF,x=logSVL))+
    geom_point(colour="gray",size=3,alpha=.4)+
    geom_point(data=mat.ran,alpha=.7,
               aes(y=logDF,x=logSVL,colour=environment),size=3)+
    geom_smooth(data=mat.ran,se=F,
                aes(y=logDF,x=logSVL,colour=environment),method="lm")+
    theme(panel.background = element_rect(fill="white",colour="black"),
          axis.text = element_text(size=14),
          axis.title = element_text(size=16),
          legend.position = "none")+
    ggtitle("Ranidae")+
    ylab("Log dominant frequency (lnDF)")+
    xlab("Log snout-vent length (lnSVL)")+
    annotate("text", x = c(3.4,3), y = c(5.5,6),
             label = c("p < 0.001 (environment)","N = 38"))

### Hylidae:
```

```
mat.hyl <- filter(mat,fam=="Hylidae")

g.fam3 <- ggplot(mat, aes(y=logDF,x=logSVL))+
    geom_point(colour="gray",size=3,alpha=.4)+
    geom_point(data=mat.hyl,alpha=.7,
               aes(y=logDF,x=logSVL,colour=environment),size=3)+
    geom_smooth(data=mat.hyl,se=F,
                aes(y=logDF,x=logSVL,colour=environment),method="lm")+
    theme(panel.background = element_rect(fill="white",colour="black"),
          axis.text = element_text(size=14),
          axis.title = element_text(size=16),
          legend.position = c(0.8,.8))+
    ggtitle("Hylidae")+
    ylab("Log dominant frequency (lnDF)")+
    xlab("Log snout-vent length (lnSVL)")+
    annotate("text", x = c(3.4,3), y = c(5.5,6),
             label = c("p = 0.005 (environment)","N = 161"))

grid.arrange(g.fam3,g.fam2,g.fam1,ncol=2)
```

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