Supporting Information: Background Noise as a Selective Pressure: Anuran Species from Lotic Environments Call at Higher Pitches

David Lucas R??hr; Gustavo B. Paterno; Felipe Camurugi; Flora A. Junc?? ; Adrian A. Garda

December 17, 2014

Contents

All Data and code required to repeat the analysis bellow are linked at: raw data, source code and the code is to be executed. Thus this reserach reaches the gold standard of reprodibility for full replication (Peng, 2011). This documment was generated in **R studio** with `kintr` package. All files and documentation are available the Github.

1. Packages versions:

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

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

For details about each package version, please check the folder `packrat` folder at the Github.

2. Data structure:

I. Species data

The species dataset contain six variables (see Methods for detailed information on data collection and raw data to download data).

variable	discription
sp	species name
environment	Reproductive environment (lentic or lotic)
DF	Dominant frequency (hertz)
SVL	snout vent length (mm)
logDF	log10 of dominant frequency (DF)
logSVL	log10 of snout vent length (SVL)

Specis data structure

Last six rows of the species dataset:

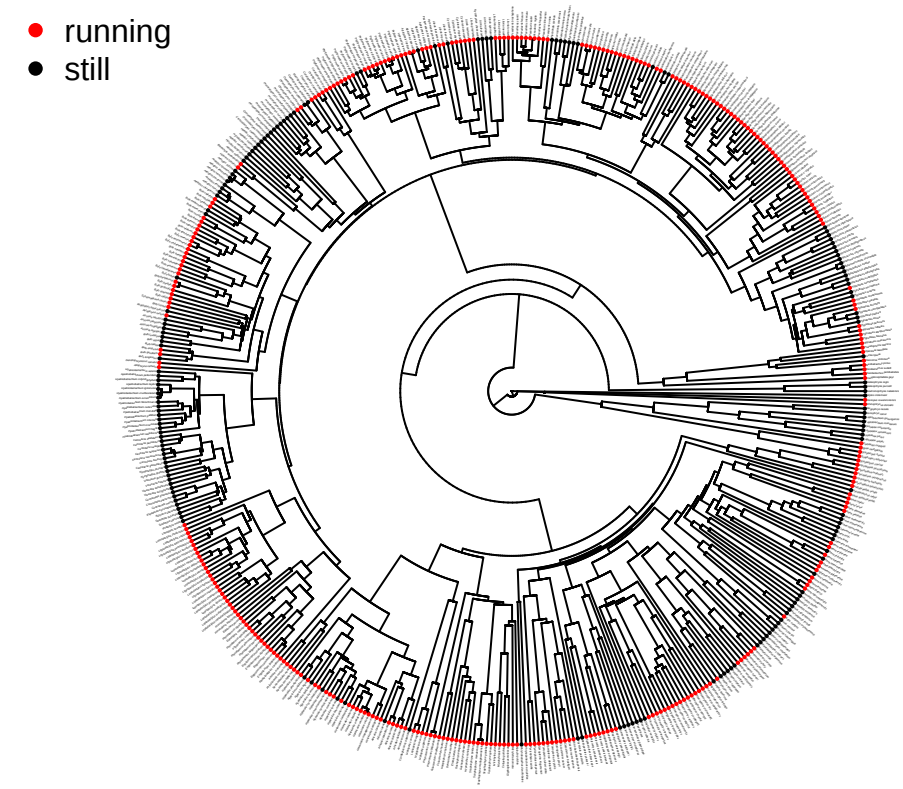
	sp	environment	DF	SVL	logDF	logSVL
504	Bufo_retiformis	still	3113	47	3.493179	1.672098
505	Bufo_houstonensis	still	2151	77	3.332640	1.886491
506	Pelophryne_misera	still	4000	21	3.602060	1.322219
507	Ansonia_longidigita	running	3500	50	3.544068	1.698970
508	Ansonia_hanitschi	running	5700	32	3.755875	1.505150
509	Ansonia_platysoma	running	8000	25	3.903090	1.397940

II. Phylogeny

The phylogenetic tree used in this paper was pruned from: Pyron and Wiens (2014) anura super tree. The pruned tree with study species (509) is available at: study tree.

Phylogeny structure

```
##
## 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.
```

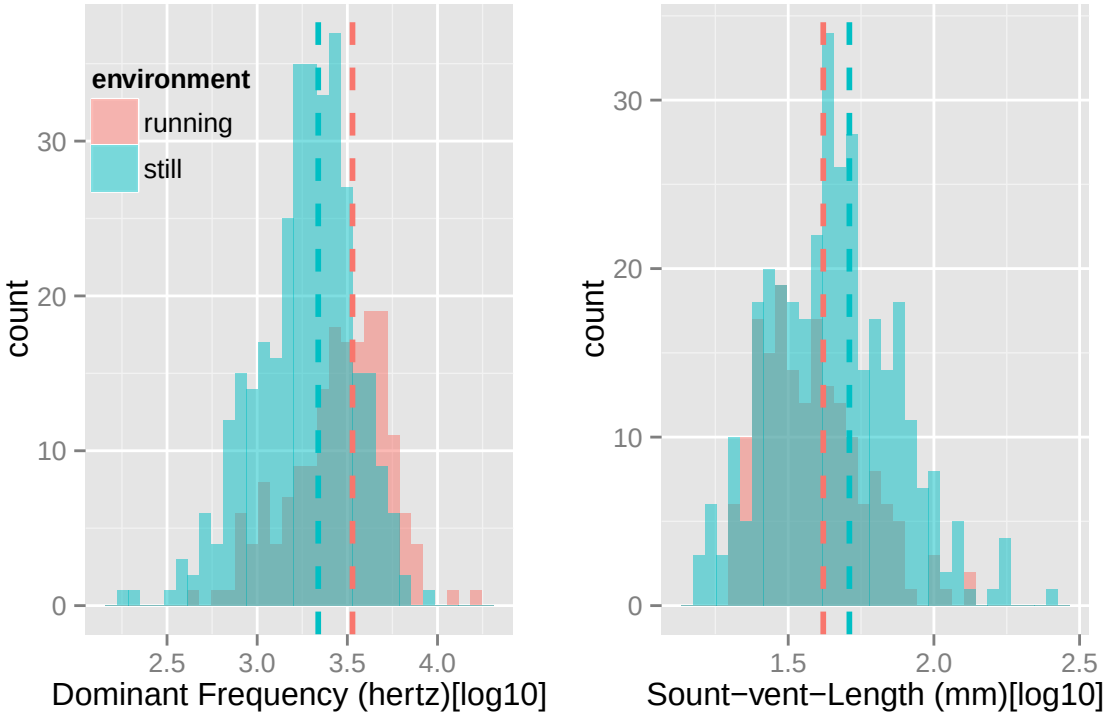

running
still

3. Data summary:

A. Mean and standard deviation 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

B. Dominant Frequency and Snout-Vent-Length distribution



4. Phylogenetic signal

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	0.0014224	0.0054636	0.001	-4.352642
logSVL	0.4387480	0.0006061	0.0028160	0.001	-4.719586

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

5. Phylogenetic generalized least square

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`.

A. 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] 3.18 3.35 3.26 2.94 2.91 ...
##    $ logSVL     : num [1:509] 1.65 1.67 1.63 2.08 1.6 ...
```

A. 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	0.0018287	0.0027995	0.001	-1.877866

Since the residuals from OSL regression show phylogenetic signal k = 0.15, there is strong need to correct for phylogenetic nonindependence in data.

B. Data analysis:

Fitting pgls model with with lambda adjusted by maximum likelihood:

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

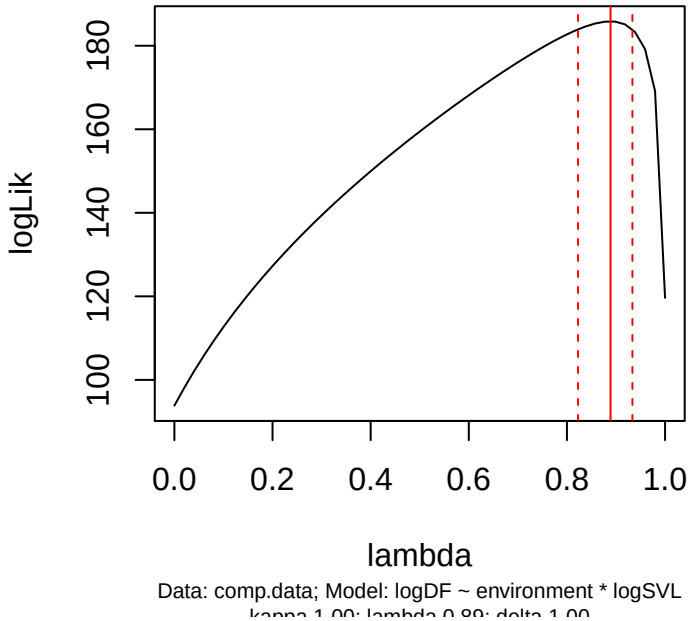
```
##
## Call:
## pgls(formula = logDF ~ environment * logSVL, data = comp.data,
##     lambda = "ML")
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.073524 -0.014980 -0.001757  0.010458  0.060338
##
## Branch length transformations:
##
## kappa  [Fix]  : 1.000
## lambda [ 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
```

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

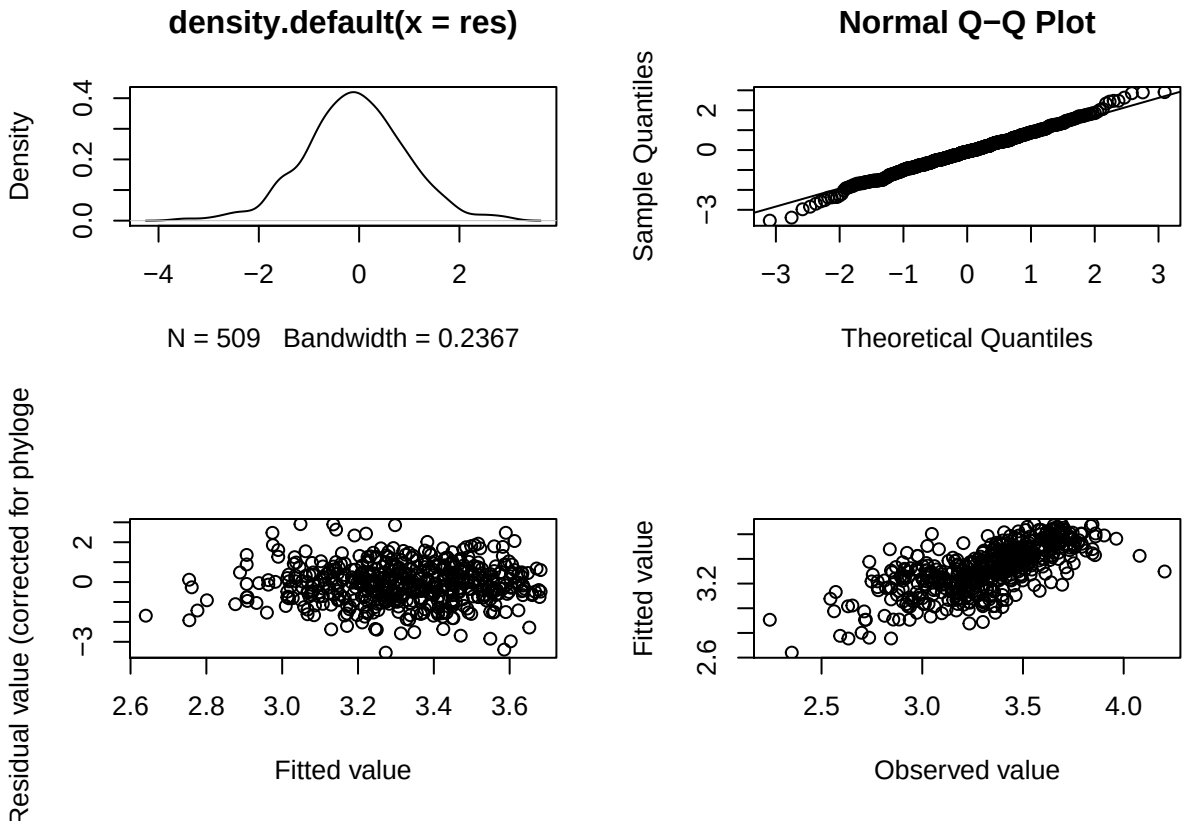
Confidence interval for lambda estimation:

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



C. Model diagnostic:

I. Standard graphic methods for model diagnostics



Residulas do not show any tendency.

II. Phylogenetic signal of model residuals
After performing PGLS analysis it is important to check the phylogenetic signal of model residuals (reference)

```
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.

D. Model comparison: OSL vs PGLS

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

	df	AIC
mod.osl	5	-177.6885
mod.pgls	4	-363.6674

AIC comparison shows that PGLS model has much lower AIC value (-363.6674412) then OSL model (-177.6884567). Thus, PGLS model is a better fit for the data.

6. References

1. Pyron, R.A. & 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, 543-583.

2. Blomberg, S. P., T. Garland, Jr., and A. R. Ives. 2003. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. Evolution 57:717-745.

3. David Orme, Rob Freckleton, Gavin Thomas, Thomas Petzoldt, Susanne Fritz, Nick Isaac and Will Pearse (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 JM, Cooper N. 2013 Phylogenetic signal in primate behaviour, ecology and life history. Phil Trans R Soc B 368: 20120341. http://dx.doi.org/10.1098/rstb.2012.0341
5. Freckleton, R.P. (2009) The seven deadly sins of comparative analysis. J Evol Biol, 22, 1367-1375.

7. Links summary:

1. raw data
2. source code
3. species list
4. Study phylogeny
