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

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

December 17, 2014

Contents

This doccument follows the principles of reproducible science. All data and code required to repeat the analysis bellow will be available at github

---

1. Packages and Functions used:

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

```
# Function to calculate standard error:
stderr <- function(x) sqrt(var(x,na.rm=TRUE)/length(na.omit(x)))
```

1. Data structure:

A. Species data

The species data contains six variables (see Methods and raw data for full acess and detailed information about 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)

Species data structure

```
##                       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
```

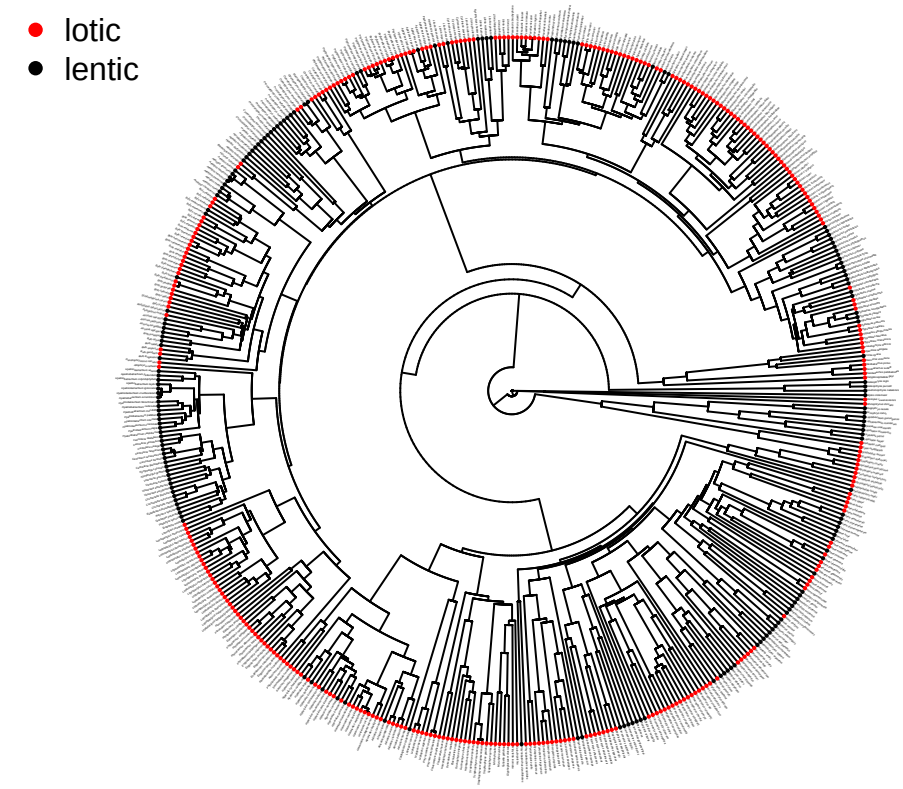
```
## 'data.frame':    509 obs. of  6 variables:
##  $ sp         : Factor w/ 509 levels "Afrixalus_delicatus",..: 16 509 456 458 480 351 508 294 344 174
##  $ environment: Factor w/ 2 levels "running","still": 1 2 2 2 2 2 1 1 1 1 ...
##  $ DF         : int  1491 1500 1650 1500 1300 1075 3456 5588 1071 1500 ...
##  $ SVL        : num  41 44 72 67 49 85 32 29.6 58.4 45 ...
##  $ logDF      : num  3.17 3.18 3.22 3.18 3.11 ...
##  $ logSVL     : num  1.61 1.64 1.86 1.83 1.69 ...
```

B. 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 github

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.
```

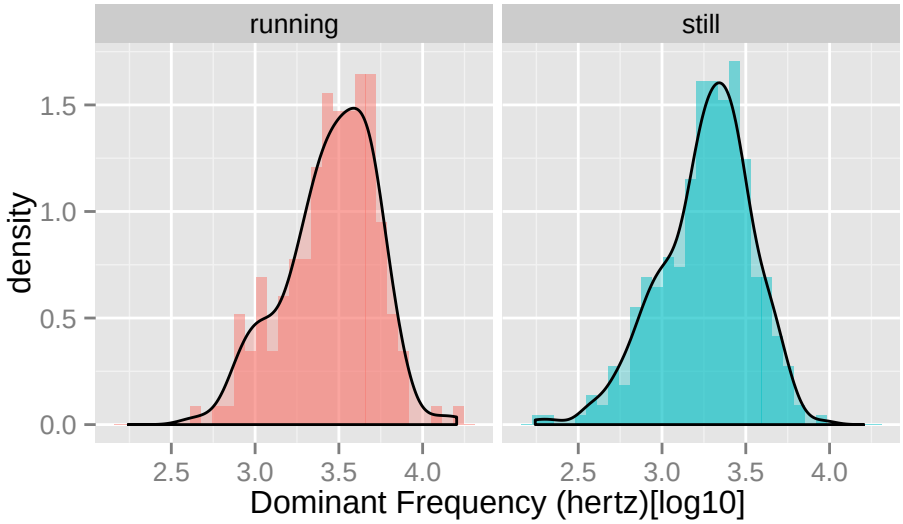


2. Data summary:

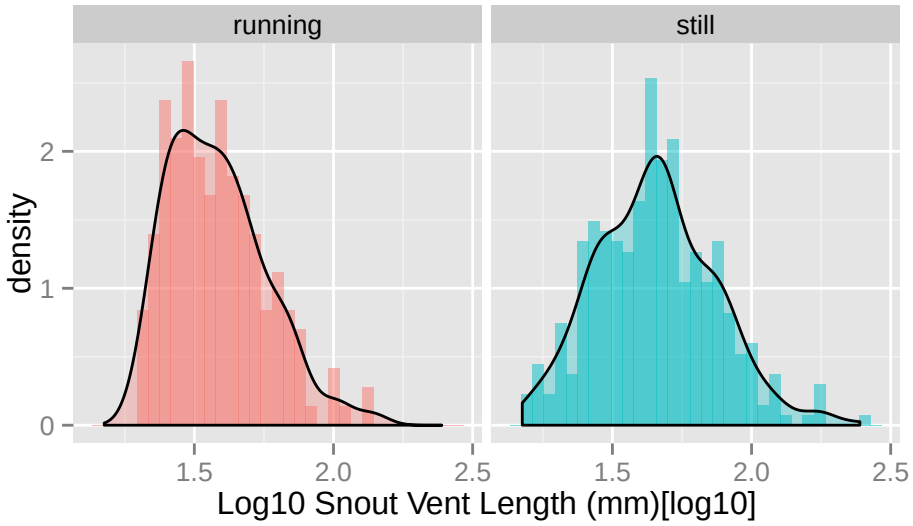
A. Mean and standard deviation for Dominant frequency and Sout-vent-length

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 distribuion



C. Snout Vent Length distribuion



3. 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	0.0055149	0.001	-4.449867
logSVL	0.4387480	0.0006061	0.0028197	0.001	-4.767759

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

4. 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: filo
##    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: mat
##    $ 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 ...
```

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
```

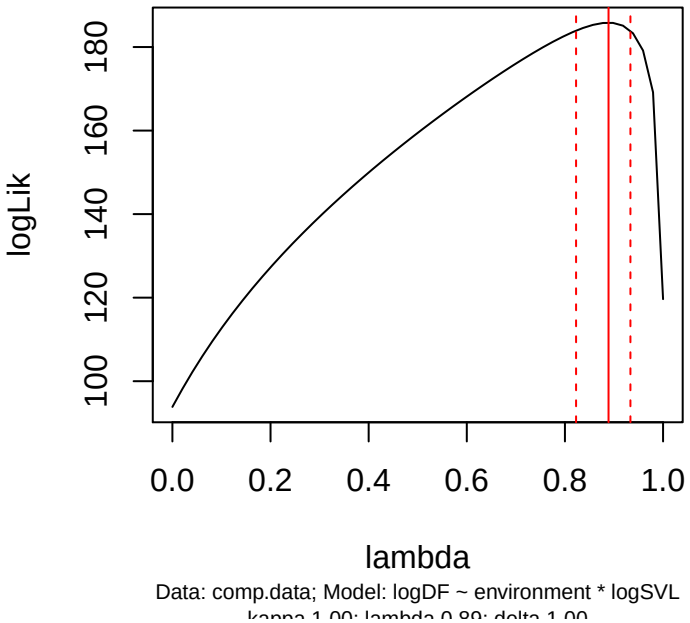
```
##    95.0% CI   : (0.823, 0.933)
## delta  [Fix]  : 1.000
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)              4.874025   0.190792 25.5462   <2e-16 ***
## environmentstill        -0.182756   0.173309 -1.0545   0.2922
## logSVL                  -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

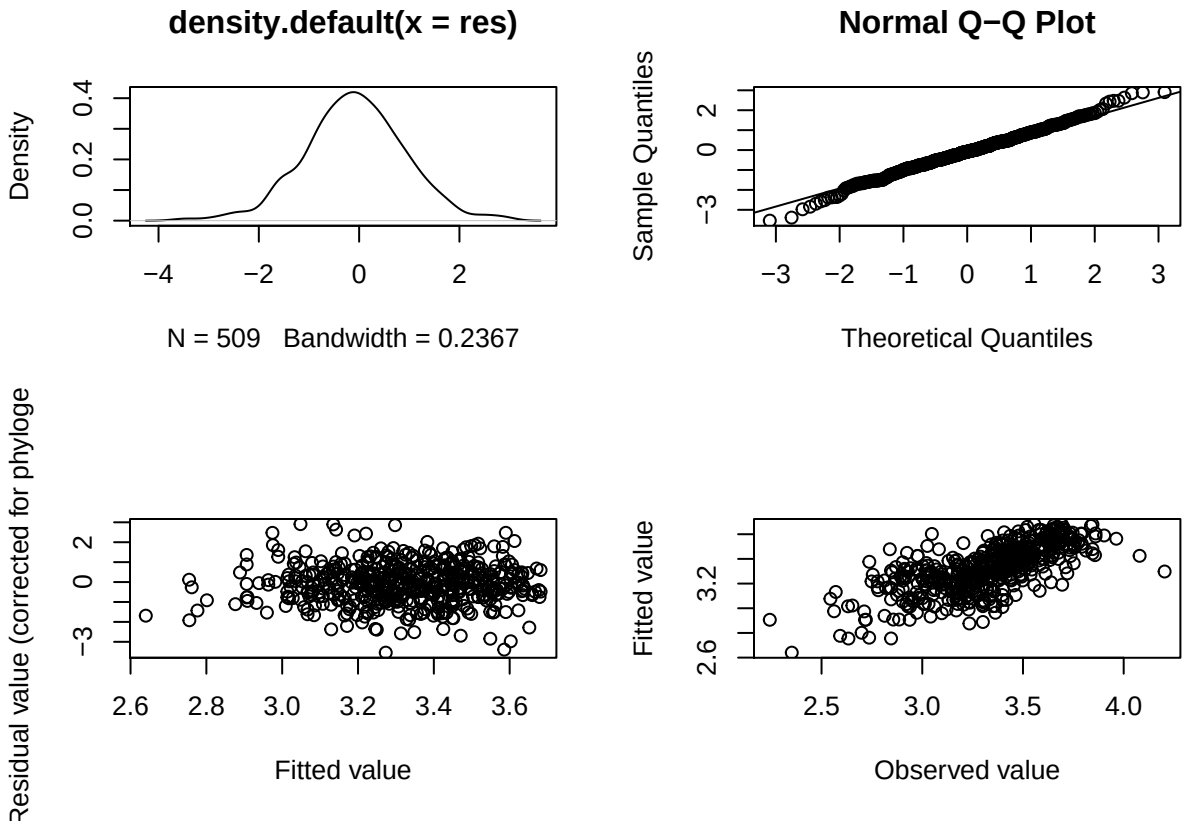
Confidence interval for lambda estimation:

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



C. Model diagnostic:

1. Standard graphic methods for model diagnostics



Residulas do not show any tendency.

2. 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	2.93e-05	0.116	-1.011824

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

5. 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. ## Links summary:

```
* [raw data](https://gist.github.com/paternogbc/d73612e22c36e538ee54)
* [source code](https://github.com/paternogbc/2014_Rohr_et_al_JEB)
* [species list](https://github.com/paternogbc/2014_Rohr_et_al_JEB/blob/master/species_list.csv)
* [Study phylogeny](https://github.com/paternogbc/2014_Rohr_et_al_JEB)
* [Sensitive analysis code]()
```