

Dear Karin Pfennig,

We are now re-submitting the paper “Background Noise as a Selective Pressure: Stream-breeding Anurans Call at Higher Frequencies” (JEB-2014-00562). All suggestion made by the referees were fully implemented and we trully believe that this new version achieved the quality standards of Journal of Evolutionary Biology. Additionally, we produced a new supplementary material with a complete discription of all statstical analyses performed, containing also the raw data and the R code for full reproducibility (see supplementary material for details and [Github repository](#) for access to all content). Below we answer each reviewers observations point by point.

---

## Reviewer 1

### 1<sup>o</sup> Comment:

*There’s no need to use the jargon “lotic” and “lentic”. I strongly encourage the authors to replace these throughout with more familiar terms such as running and still.*

**Answer:** Modified accordingly. We changed lotic habitats to either “streams” or “flowing water”

### 2<sup>o</sup> Comment:

*The two environments differ in myriad ways other than sound. It seems likely, for example, that the predators at the two environments will differ in important ways that may impact frog calls. Or water flow may impact tadpole development which in turn affects body size and steroid hormones, both of which would impact calling behavior. Without explicit measures of sound levels in the two habitats, the authors should acknowledge and discuss the multitude of alternatives.*

**Answer:** Modified accordingly. Modified accordingly. See lines: 174-179.

### 3<sup>o</sup> Comment:

*There are many forms of PGLS, and the authors should cite the specific methods that they used here. In addition, each method necessarily assumes an underlying evolutionary model that should be specified and justified here.*

**Answer:** Modified accordingly. See lines: 85-88 and Supplementary Material.

### 4<sup>o</sup> Comment:

*The authors might consider also applying the Hansen (1997) adaptation-Inertia model, which seems a better match to their evolutionary assumptions and would thus offer more specific insights.*

**Answer:** We used a comparative method approach in our analysis specifically to correct for non-independence among species. Our aim was to test the effect of the environment noise on dominant frequency and not to test by which evolutionary model this trait has evolved. Thus, following Kamilar & Cooper (2013) reasoning:

*“Phylogenetic signal is also context (i.e. data and phylogeny) dependent and can be influenced by scale, convergent evolution, taxonomic inflation and cryptic species [10]. Because of this, using  $K$  or lambda (or any other measure of phylogenetic signal) to infer evolutionary processes or rates must be performed with consideration of the traits involved, the hypotheses to be tested and any available external information [35]. Note that this only applies to studies whose aim is to quantify and report measures of phylogenetic signal. For analyses that use  $K$  or lambda to account for phylogenetic non-independence in statistical tests, the evolutionary process underlying the phylogenetic signal is irrelevant.”*

we see no reasons to change our statistical approach regarding this analysis.

Finally, in comparative studies with continuous response variables, the PGLS (sensu Freckleton et al., 2002) is one of the most powerful and used statistical methods available (Garamszegi, 2014, the most recent book on the topic).

## 5<sup>o</sup> Comment:

*Vargas-Salinas & Amezcuita (2014) did a similar analysis but with very different results. This paper is mentioned, but not discussed, in the Discussion. I would like to read in more detail why they think the results disagree.*

**Answer:** Modified accordingly. See lines: 126-130.

---

## Reviewer 2

### 1<sup>o</sup> Comment:

*“Pitch” refers to a perceptual experience, not a measurable, physical feature of a sound. One cannot equate pitch (the percept) with frequency (the physical measure). In addition, we do not know whether frogs experience “pitch” the same way humans do. Therefore, it is technically inappropriate to refer to the “pitch” of frog calls unless one is talking about how a human perceives the call. The appropriate technical term the authors should be using throughout the manuscript is “frequency.” All instances of “pitch” in the manuscript, including the title, should be changed to “frequency.” This may ultimately require more words, but it would make the writing technically correct, which it is not currently. The authors may also want to make crystal clear that they are referring to “acoustic frequency” to avoid any possible confusion with changes in call rate.*

**Answer:** Modified accordingly.

### 2<sup>o</sup> Comment:

*I am somewhat puzzled about why the authors chose the MEAN DF as the response variable, but chose the MAXIMUM SVL as one of their predictor variables. Perhaps they have scientific reasons for doing so, but I can't think of any particularly good ones. I think good arguments could be made for basing their analyses on (i) the MINIMUM DF and the MAXIMUM SVL, or (ii) the MEAN DF and MEAN SVL, or (iii) the MAXIMUM DF and MINIMUM SVL. In fact, I think a much more rigorous and thorough analysis of the data would report results for all three of these analyses. Without reporting these additional analyses, the authors could mistakenly convey the*

*impression that they have gone fishing for statistical significance to find a particular analysis that supports their a priori hypothesis. I would like to see results from all three of these analyses in any revision of this manuscript. Unless something interesting is revealed in these new analyses (e.g., outcomes inconsistent with the authors' favored hypothesis), this should not lengthen the manuscript by more than a few more lines in the tables already presented.*

**Answer:** 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 and mean dominant frequency we:
  1. Re-collected data for maximum, minimum and mean SVL and DF for 30 random species of our previous survey.
  2. Performed a Pearson's correlation analysis between maximum, minimum and mean SVL and DF (see results below).

Snout-vent-length:

	SVLmax	SVLmean	SVLmin
SVLmax	1.0000000	0.9836185	0.9077609
SVLmean	0.9836185	1.0000000	0.9656060
SVLmin	0.9077609	0.9656060	1.0000000

Dominant Frequency:

	Dfmax	Dfmean	Dfmin
Dfmax	1.0000000	0.9870885	0.9189477
Dfmean	0.9870885	1.0000000	0.9513402
Dfmin	0.9189477	0.9513402	1.0000000

For both parameters (SVL and DF), the correlation coefficient between maximum, minimum and mean values was higher than 0.9, thus we truly believe that our choice of maximum SVL and mean DF by no means affected any aspect of our main results or final conclusions.

### 3<sup>o</sup> Comment:

*For those instances in which there were “multiple hits” for a species in the literature, I think basing the decision about which to include in the dataset on “recency” (line 77) is a bad idea. The choice should be based on the quality of the study, sample sizes, etc. Not all studies are created equal. A recent study is not necessarily the best or most accurate. I think decisions about which data to include in all analyses should be based on including either the “most” data or the “best” data, but certainly not the “most recent.” This will require some re-analyses of the data.*

**Answer:** To construct a large data set, we used 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:  mat_t_test[, 1] and mat_t_test[, 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.

### 4<sup>o</sup> Comment:

*The authors need to report estimates of effect size (e.g., something like a partial eta-squared or R-squared). They will also need to make a discussion of the relative effects of SVL and habitat on DF a more prominent component of the paper. From looking at the results of the ANOVA table, it would seem that SVL has a MUCH larger effect than habitat type, but this is not something the authors appear to clearly acknowledge in the manuscript. What we might find is that, yes, there is some effect of habitat type, but it is a relatively small effect that turns up as statistically significant in a very large study like this one. This could also explain why some studies find it, and others do not. This would be important information that would allow the community of frog bioacousticians to determine whether we could (finally) agree to lay this issue to rest once and for all, and whether or not it is really all that interesting to keep devoting effort to study such a small effect. This paper has some potential to make the following very important point: “Yes, there’s an effect of noise in lotic environments; it is a small effect; now let’s all move on to something else more interesting.”*

**Answer:** Modified accordingly: See lines: 131-143

## 5<sup>o</sup> Comment:

*I think the authors have provided a somewhat superficial treatment of the literature in places. They also appear to place more emphasis on citing recent papers documenting a particular phenomenon, instead of the original or best papers doing so. One place this problem is particularly relevant is on lines 132 to 139. The authors here make three points: (1) Body size (and hence DF) may be an important predictor of the outcomes of aggressive interactions. They cite a recent study by Meuche et al., 2012, but this idea goes back more than three decades to the work of Davies and Halliday (1978). (2) Male frogs can lower the frequency of their calls during aggressive interactions. While the cited 2002 paper by Bee & Bowling is one of the most recent documenting the effect, that paper is a short communication in a herpetology journal; Bill Wagner's work on this topic, going back to 1989, is the best original work showing male frogs can do this (though there are some even earlier hints of this in the literature). Both Wagner and Bee published earlier and much better papers documenting the effect than the one cited by the authors. (3) Female frogs can exert selection on DF through mate choice preferences. The authors cite a paper by Gerhardt (1991); certain authors in the field who shall remain nameless have hammered this supposed preference into accepted frog dogma (= frogma?), but a much more careful and thorough analysis of all the data, reviewed by Schwartz and Gerhardt in (2001), would seem to suggest that any such preferences for frequency are relatively weak, and there are now examples of female frogs absolutely rejecting males having the lowest-frequency calls, and even female frogs preferring relatively higher-frequency calls. The take-home point is that the studies cited to support the generalizations being made in this particular paragraph are probably not the best, and leave the impression that the authors may possess a somewhat limited familiarity with the frog acoustic communication literature. Similar arguments could be made for other places in the manuscript, where the scholarship just doesn't seem to be quite up to par with what one familiar with more of the literature might expect. I realize that there may be space limitations associated with Short Notes that ultimately limit the number of citations. If so, I would encourage the authors to cite the best papers or the original papers on a topic, and not simply the most recent.*

**Answer:** Modified accordingly. See Lines: 152-158.

## 6<sup>o</sup> Specific comments:

All minor suggestions were implemented in the new version of the manuscript.

Sincerely,

David Lucas Röhr