nature portfolio

Peer Review File

Evidence for ecological tuning of anuran biofluorescent signals

Corresponding Author: Dr Courtney Whitcher

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This work is very exciting! The authors show biofluorescent signals in 39 genera (152 species) of frogs, representing 13 families. The authors uncovered a visual signal that is relevant at a broad taxonomic scale, but has been invisible to researchers to date, making this an important finding for (potentially) decade to come. Strengths of this paper include a novel finding on a wide taxonomic scale, which has the likely outcome of propelling the fields of visual ecology, trait selection and speciation forward. I think the paper could be developed by attenuating some of the broad overstatements in parts of the paper – if done correctly, the impact could be even more powerful.

My impression is that the authors place too much emphasis on the expectations of these four criteria, without necessarily having everything in place to test those criteria. It's a nice framework, and I can see the appeal. However, this framework places an unnecessary emphasis on meeting criteria, rather than highlighting that it is interesting that while ~1/3 of the taxa meet the criteria ~2/3 do not. But I would argue that failure to meet these criteria does not necessarily mean that there is not biological function to these signals. (technically, you seem to be looking at species recognition, not ecological function). You could have failed to find 'ecological function' because you didn't' consider the visual pigment of that species, because it is relevant to heterospecifics or because the frog is active during the day (if I understand correctly, they used the common green rod and used twilight as background in all analyses). The table is a nice summary of findings at the taxonomic scale – but could it be organized in a way that helps the reviewer make sense of all of these data (e.g., have a column for ecological function criteria met Y/N, a diurnal/crepuscular column, whether body parts match known visual signaling repertoire).

Criterion 2. In lines 427/428 the authors say that they tested both diurnal and crepuscular/nocturnal frogs. I would assume that the diurnal frog contrast background would be daylight, but that wasn't clear. Perhaps a way to clarify this is to analyze the diurnal frogs separately? I like that both diurnal and crepuscular anurans are included, but their graphics/tables don't make it easy to follow this important distinction.

Criterion 3. The authors test this for 'the most abundant rod in the visual system'. Line 302 is an overstatement. This seems like a big caveat, given that frogs are known to differ in visual pigments (as the authors point out). It seems worth testing whether biofluorescence is detectable if they considered a few of the alternate visual pigments known for frogs, including for diurnal frogs for the diurnal taxa.

There is a missing citation for line 306.

Having two rods and one cone make it possible (but not certain) that the frog has color vision (you need three visual pigments for the possibility of color vision). This is oversimplified in the text at the expense of being incorrect.

Whether the frog has color vision in low light requires visual modeling and behavioral testing. Line 309 says 'most frogs mate in dim light' but that isn't true for diurnal taxa. They mention this as a caveat (line 433) – but instead of a caveat, they could use a visual pigment model from a diurnal taxon.

If I understand correctly, both green and orange peaks provide a contrast in dim light. However, only green peaks are detectable using the green rod sensitivity. There are 316 individuals that have orange peaks, which then, are not detectable. Is it possible that those species have different visual pigments? The authors offer up the alternative hypothesis that orange is

a signal to heterospecifics. Given that more individuals reflect biofluorescence outside the range of the common anuran green rod, it is worth asking if it is variable frog visual pigments that co-vary with this signal.

Criterion 4. Criterion 4 is important for this framework, but I don't think they have the data to test it effectively – or you are not using it to it's potential. That is, what is the alternative, given that you tested every part of the frog body that could be tested? I have two competing thoughts: While it's true, that there are some generalities about frog signaling, visual signals are species specific – and only after behavioral study do we really understand how visual signals are displayed and perceived (as the authors point out in line 397). The authors provide a wonderful example in D. parviceps. The challenge is, therefore, matching the biofluorescence body part with what we know about visual signaling in each species. My second thought is that this study opens the possibility that we don't really understand visual signaling in frogs – and that our understanding of signaling is constrained to the signals that we can perceive (movement, calls). This dataset shows that nearly every part of the frog could contain a signal – even those regions of the body not typically associated with visual signaling (e.g., eyes, vent on most frogs).
Identifying the parts of the body that emit biofluorescence is a great starting place to generate hypothesis to test about visual signaling for any particular species. That is to say, I think the authors miss an opportunity to describe to best use this body part data. The eyes are a great example. Most eye color don't reflect in visible light. It opens doors to think that have identified a biofluorescence signal in the eye -- this makes eye size and shape an interesting character – one that has been overlooked. Is there a more effective way to use these data?

Reviewer #2

(Remarks to the Author)
Overall comments:

I recommend this manuscript for publication after minor revisions. This manuscript examines biofluorescence in anurans. The authors carefully consider biofluorescence through the lens of ecological significance. They provide a sufficiently thorough review of prior studies of this phenomenon in amphibians and other vertebrates. They clearly present their questions, hypotheses, and results. The major conclusions of the study include that 1) biofluorescence is most intense in response to blue excitation light, 2) that the resulting green fluorescence would stand out in chromatic contrast to the light environment during periods of activity for most anurans (dim lighting, twilight), and 3) that the emission spectra generally align with our understanding of photoreceptor cell sensitivities in this group. Their findings will have a substantial impact on our understanding of the evolution and relevance of biofluorescence in amphibians, and more broadly, in tetrapods. With some minor modifications and clarifications to the text, the study is reproducible. The changes below are recommended to improve clarity and to reduce redundancy throughout the manuscript, including in some of the figures. Overall, this was an enjoyable, interesting, and well-composed manuscript.

#####

Line by line comments

Introduction
>>> Overall: Succinct and well written. Important context is given for the study and a thorough review of existing literature is provided. The question, hypotheses, and objectives are clear.
>>> Lines 57 - 60. Recommend revising text to clarify the point. Some of the potential functions mentioned in this text for fluorescence are hypothesized functions, but were not explicitly tested in the papers cited (e.g., camouflage in fishes, Sparks et al.). If there is a way to make that point clearer it would further emphasize the novelty of this study which does propose and test a hypothesis regarding the function of biofluorescence.
>>> Line 75. Recommend changing "at another wavelength" to "at other wavelengths." Some fluorophores seem to be capable of emitting more than one wavelength of light.
>>> Line 83. Recommend revising text to clarify the point. Taboada et al.'s work was perhaps the first publication to focus on biofluorescence in anurans and to document patterns in whole organisms. However, there may be earlier publications in the fields of developmental biology/embryology that document "autofluorescence" in anurans (there are cases of this for caudates).
>>> Lines 106 - 109. Recommend revising to break up into shorter sentences.

Results
>>> Overall: The results are easy to understand, presented in a logical manner, and adequately address the questions and hypotheses posed in the introduction. However, there are some areas where the authors should revise the text to improve clarity. There are also areas where the results and discussion seem redundant (i.e., the same information is presented in nearly identical terms). This section and the discussion should be reviewed and trimmed down to reduce repetition.
>>> Lines 117 - 123. Optional revision if more space is needed elsewhere. These lines detail how many more taxa this study has surveyed for biofluoresence. While important, this section could be shortened or simplified if additional space is needed elsewhere in the manuscript.
>>> Lines 125 - 127. Recommend revising text to make fewer assumptions. The authors state that "novel biofluorescent patterns…were missed in previous studies because the incorrect excitation wavelength was used." Reccomend revising to emphasize that the wavelengths used by other studies may not have excited the most intense fluorescence. If the choice of excitation wavelengths in other studies was driven by a particular prediction or hypothesis, that choice may not necessarily have been "incorrect."

>>> Line 127 - 131. Additional point of clarification may be needed. In some previously published studies, like the one mentioned here (Thompson et al.), authors may not have used a barrier filter to survey for fluorescence. This could mean that reflected excitation light “drowned out” any fluorescence. So in the cases of the species mentioned, was the “incorrect” excitation source used or were fluorescent wavelengths missed for the wavelengths tested?
>>> Line 151. Grammatical error. Capitalize “blue light” towards the end of the line.
>>> Lines 162 - 177. Clarification needed regarding the type of contrast being presented. When talking about contrasting against the background, the authors are focusing on chromatic contrast. Organisms can also exhibit achromatic contrast, meaning that they differ in brightness within a particular chroma or across all chroma. Endler (2006; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1634903/) briefly discusses how the type of contrast may be ecologically and evolutionarily relevant. The authors should revise the manuscript to clarify that they are only considering chromatic contrast.
>>> Line 219 - 227. Recommend revising so that interpretation occurs in the Discussion, and so that it is clearer what is speculation/prediction versus reporting of their results. For example, “could result from secretions from the frog skin” would be more appropriate in the discussion. Additionally, are the authors reporting that ventral biofluoresence was widespread, or that it could be widespread? Phrases like “could be widespread,” and “could be present in distinct regions” should be revised to “was widespread” and “was present in distinct regions.”
>>> Line 203 - 235. Recommend breaking this large chunk of text into more than one paragraph.

Discussion
>>> Overall: The discussion and results need to be streamlined so that there is less repetition between them.
>>> Lines 239 - 298. Recommend revising these sections so that they are less repetitive with respect to the results sections. Also see comments in results regarding the use of some terms (e.g., “incorrect excitation wavelength”). The authors should (and do) reference specific results in their discussion. However, in its current form, large portions of the discussion seem highly repetitive with respect to the presentation and interpretation of information in the results section.
>>> Lines 335 - 337. Recommend revising to only used “additional” once.
>>> Lines 369 - 373. Recommend revising to be less repetitive with the results. This text is very similar to text previously presented.
>>> Line 402 - 404. Recommend revising for clarification. Did the authors examine vocal sac fluorescence when the sac was inflated? It may be that fluorescence changes in intensity when this region of the body is stretched.

Materials and methods
>>> Overall: The materials and methods are relatively clear, and allow the study to be repeated. There are a few points where additional details are needed.
For the three replicate spectra collected, did the authors move or adjust the probe position between readings? Or were spectra collected via a “rapid fire” type approach? Meaning that while the probe and animal were stationary, three measurements were collected.
>>> The authors mention the placement of their probe during spectra acquisition, but what about the position of the light source? They note it was suspended via a gooseneck tripod stand, but not whether that was adjusted over time or with different specimens. I believe that potential movements of their light source may be accounted for in their method of calculating % biofluorescence intensity. However, this is worth pointing out so that others can better understand their approach.
>>> Please clarify the settings used with the Nikon digital camera. Details such as f-stop, exposure time, and ISO settings are relevant to interpreting the figures presented.
>>> Have any other publications approached calculating the maximum percent biofluorescence emission in this way? If so, they should be cited. If I understand the methods correctly, the intensity at peak excitation wavelength and the intensity at peak emission wavelength were calculated from the same spectral reading. I believe this would mean that the intensity at peak excitation wavelength is actually the intensity of the reflected excitation light. I think this value would be lower than the total intensity of the light coming directly from the excitation source due to absorbance, scattering, or transmission. I don’t think that this point of clarification would mean that the results would differ dramatically from how they are presented, but it is an important point to clarify and consider.
>>> Blinding as described in the supplementary files. The authors specify who was handling the amphibian and spectrometer probe, and who was viewing the OceanView application. They specify that the individual handling the amphibian was “blind” to the fluorescence readings so as not to preferentially select a body region for measurement. Was the person handling the amphibian and probe also wearing filtering glasses so that they could target regions that were fluorescing? If not, it seems like the measurements taken could be underestimating fluorescence if the region measured was not in fact the most fluorescent region.

Figures and tables
>>> Overall: The majority of the main figures and tables are necessary and well designed. There are a few that should be split into more than one figure, or otherwise modified to improve readability. None of these are major changes.
>>> Figure 1. See Comment #5 in discussion.
>>> Figure 2. Overall this is a great figure. Some minor modifications will help with clarity and readability.
The depiction of the “shift” from excitation to emission spectra doesn’t need to be included in each of the spectra plots for Criteria 3 and the Observed Data section. It contributes to the Observed data being a bit challenging to interpret at first glance. If the authors want to maintain it in those parts of the figure, then can a different type of line be used to distinguish between the excitation and the emission spectra? Perhaps a dotted line with circles for the former and with squares for the latter?
The caption for Figure 2 is very long and should be shortened so that it is not so repetitive with the results section. Limit the text in the caption to the essentials for understanding the figure.
>>> Figure 3. Some minor modifications will make this figure more readable.

Is there a way to reduce the number of different text sizes used? Some of the sizes used are very small (e.g., for median values and for the axes labels in the inset plot).
I am not sure that the inset is necessary here. It also seems to be covering some of the data about distributions of emission intensities.
>>> Figure 5 - 7. Modifying how the data are presented will improve clarity. I understand that where the open circles fall on the twilight spectrum is meant to indicate what wavelengths of fluorescence were observed. However, it also implies that those irradiance/intensity levels were observed, which I think is what makes the plots confusing. I encourage the authors to consider a different approach where vertical rectangles or bars are used to indicate what wavelengths of fluorescence were observed. This would emphasize that the only information being presented for fluorescence has to do with the wavelengths fluoresced, not with their intensities. The average wavelength for emissions could be presented as a bold line.
>>> Figure 8. Moderate revisions to the figure needed.
I recommend splitting the figure into two figures. One that focuses on the relative proportions of fluorescence by body region and excitation source (pie charts). And a second that focuses on the images of anurans under white light and while fluorescing under royal blue light. This would allow the authors to maximize space and potentially have the white light images be larger. Larger images will facilitate a better understanding of where fluorescence is occurring and what features might be contributing to that fluorescence.
Since the binomial names are given in the figure caption, they can be removed from the figures themselves. This would give the authors more space and make the figure less crowded.

Supplementary materials
>>> Figure S1. Very small text sizes for median and for X axis labels. These can be made larger.

Reviewer #3

(Remarks to the Author)
The authors conducted an impressive survey of biofluorescence in several species of neotropical amphibians. Using their newly generated data, they evaluated the ecological significance of fluorescence in frogs based on a set previously proposed criteria.

I believe the dataset generated by the authors is a significant contribution to the ongoing effort to understand the ecological and evolutionary significance of fluorescent coloration in amphibians. The data were collected in what seems to me as a rigorous manner (I am, admittedly, not an expert in biophotonics), and significantly expand the phylogenetic breadth of biofluorescence reports in amphibians. Unfortunately, there are several issues and shortcomings with the analysis, interpretation, and presentation of the data that preclude me from recommending this paper for publication in Nat. Comm. in its current form. My concerns and some specific comments are detailed below.

>> General comments:

1. The paper is generally well written, but I found getting through the Results and Discussion sections somewhat challenging due to two main reasons:

i. Although the methods section of the paper is detailed well written, the Results section has very little methodological detail for a “results-first” paper, which makes reading the methods and discussion difficult at times. I suggest the authors add basic details/summaries of their methodology to each section of the results to facilitate reading.

ii. Many of the results are repeated in great detail in the discussion section, which makes the text feel repetitive and and confusing at times, and dilutes the points being made in the discussion. I found a few stretches of nearly identical text between both sections (e.g. lines 219-235 vs. lines 369-384), so perhaps this is due to a previous reformatting of the manuscript? In any case, I suggest either succinctly summarizing key results before discussing them, or consolidating the results and discussion into a single section.

2. Most statistical analyses presented were conducted at the level of individual frogs, either across all species or within taxonomic families. This approach essentially assumes that all individual measurements are equally independent from one another, which is not the case in the author’s data, considering they involve measurements on multiple individuals from several different species. This creates structure due to 1. variation within versus between species, and 2. the different levels of relatedness between species (i.e. phylogenetic structure).

The authors should account for structure in the data when doing statistical analyses by 1. Consolidating within-species data into single measurements per species (and perhaps associated variance/uncertainty), and 2. Using phylogenetically-aware statistical approaches to analyze species-level data. There are a few publicly available large-scale phylogenies of amphibians, so this should be relatively straightforward.

Beyond statistical rigor, embracing the inter/intraspecific and phylogenetic structure of the data could help the authors address some of the open questions poised in the discussion. For instance, investigating whether orange and green peaks are polymorphic or fixed within species can speak to whether they serve different functions, and mapping many of the scored traits onto a phylogeny could help gauge if and why they are more labile in some groups than others.

3. From Table 1 it seems like all the species tested in this study displayed biofluorescence. Is this the case, or were only

species that displayed biofluorescence reported? If indeed all ~150 species tested fluoresced, this seems like a major result worth reporting and discussing. Is biofluorescence perhaps ubiquitous in frogs?

If some of the tested species did not display biofluorescence, then they should be reported, and perhaps used in analyses, for example to assess when this trait has been gained or lost in amphibian evolution.

4. I found the way the authors used and interpreted the twilight spectrum problematic for three reasons.

First, although this wasn't specified in the manuscript, looking over the Cronin et al. book it seems like the spectrum used was the one in Fig 3.18 which depicts a "typical habitat" at twilight. Cronin et al. don't specify how or where this spectrum was taken (or I didn't find this information), but considering "Under forest canopy" is one of the alternative habitats depicted, it seems like this spectrum represents twilight in an open habitat. This may pose issues for the interpretation of the paper's data, considering that many (if not all) of the sampled localities are forested, and that tree cover is well-known to influence the irradiance spectrum, so the species sampled may be experiencing different irradiance spectra.

Second, it is worth considering that each of these localities may have different irradiance spectra, so their absorbance/irradiance measurements may not be equally comparable with a single spectrum.

Finally, the presented spectrum does not have irradiance values below 400nm, which complicates the interpretation of results obtained with the UV light source, which is below this wavelength.

The authors should therefore 1. Provide more information on the nature of the used twilight irradiance spectra, and 2. Try to use spectra that represent the light available in the species' habitat as accurately as possible, and that include all of their excitation light sources. I understand that obtaining irradiance spectra specific to each sampling site may be difficult, but if available, I would suggest at least using the irradiance spectra of one or a few tropical forests at twilight. If such spectra are not available the authors should highlight these caveats upfront to better reflect the limitations of their interpretation.

4b. I fully agree that twilight is a very important context to study amphibian biofluorescence. However, considering twilight makes up a small fraction of a nocturnal frog's active period, I believe night time conditions (especially with strong moonlight) should be considered as well. Irradiance spectra at night are markedly different from twilight ones, which is highly relevant given the paper's strong focus on the spectral properties of biofluorescence and their relationship to ambient light.

5. Blue light was found to produce, on average, the highest fluorescent emittance. Based on this, and the fact that this light source is the closest to the short-wavelength peak of the used twilight irradiance spectrum, the authors restricted most of their subsequent analyses to the fluorescence observed under blue light. However, most of the other light sources are also well represented in twilight (and moonlight) spectra, and often produced comparably high emission intensities, so they could produce enough fluorescent light under natural conditions to be perceivable by potential receivers.

With this in mind I think it is warranted that the authors conduct their analyses for measurements gathered under all light sources. This would not only add to the study's biological relevance, but could help the authors evaluate criteria 1-3 better by allowing them to compare the irradiance and perceivability of fluorescence stemming from different excitation wavelengths.

6. It is not entirely clear to me why the authors used the available light (i.e. the irradiance spectrum) as the background against which the fluorescent signal's contrast was evaluated. As I understand it, a visual signal's contrast is perceived against the colors of objects adjacent/around it. For example, a frog's color will be perceived as conspicuous if it is very different from the forest floor (when seen from above), or a spot on that frog will look conspicuous if it is surrounded by patches of a very different color (e.g. yellow spots on a black background). Maybe this is because the signal is emitting its own light? Please explain this choice in more detail.

On a similar note, given the low light conditions, wouldn't it make sense to incorporate the intensity of the fluorescent signal when estimating contrast?

7. The authors argue that criterion 3 is met because the green observed in some individuals matches the sensitivity of ranid and bufonid the green rods. However, the analyses performed provide no evidence for the orange peak passing criterion 3, beyond some speculation about it being perceivable by non-anuran receivers. Considering the orange peak is present in about twice as many individuals as the green peak (316 vs 192), the support for criterion 3 seems ambiguous. Given its high frequency, for these data to support criterion 3 the authors would have to show that the orange peak is perceivable by an ecologically significant receiver.

8. When evaluating whether the emitted fluorescent light would be visible to frogs the authors only use sensitivity data of rods. However the emitted fluorescent light should be bright enough for cones to perceive it in twilight and possibly moonlight conditions. With this in mind, I suggest the authors consider cones too when gauging receiver sensitivity. For example, the red cone of Rana temporaria has a peak sensitivity around 560nm (see doi.org/10.1111/j.1748-1716.1994.tb09790.x), so it should be well able to perceive the orange fluorescence detected in some frogs.

Specific comments:

P. 2 L. 95 (and elsewhere throughout the text): I agree that the availability of light that actually excites a fluorophore to a biologically relevant degree is key here, but it doesn't seem to me like this light has to necessarily be of the most abundant wavelengths in the environment. As long as there is enough light available in the environment for fluorophores to emit a biologically a significant amount of light back into the environment the possibility of a biological function is there. Of course, the fact that fluorophores in many of the tested frogs absorb the very prevalent blue wavelengths is very good evidence for biological meaning, but I suggest the authors rephrase this as evidence in favor but not a requirement of biological function here and elsewhere in the text.

P. 3 L. 127: I suggest replacing the word "incorrect" for something else. There are no incorrect wavelengths to use in experiments, just some that produce higher/lower levels of emittance.

P. 4 L. 151: Capitalize the word "blue"

P. 4 L. 182: It may be worth adding a sentence or two to remind readers about the fact that frogs have a dichromatic rod system.

P. 5 L. 212: It isn't clear why the authors refer to Fig. 2 here. Perhaps they meant Fig. 8? I found the same situation in a few other places after this.

P. 5 L. 223: Please clarify if orange was more frequent in ventral than dorsal.

P. 5 L 229: I suggest providing citations that support this statement.

P. 6 L. 366: It is unclear to me what the authors mean by examining variation "upon which selection can act". From what I understand most if not all of the phenotypes examined in this paper (e.g. absorbance and emittance wavelengths, intensity, location of fluorophores) are heritable (at least to some degree) and variable, and therefore susceptible to natural selection and able to evolve due to it. Please clarify or rephrase.

P. 7 L. 324-327: Fig. 1 in Taboada et al. shows that fluorescence emission peaks at excitation wavelengths between 390–430nm and then starts dropping until ~450nm, which was the highest excitation they tested, sothis particular species may not have a higher emission with blue (440-460nm) excitation wavelengths.

P. 9 L. 409-412:It is not clear to me why fluorescent signals are expected to be less involved in interspecific vs intraspecific communication. Please elaborate.

P. 11 L. 472-488: Consider providing information on the type of habitat at each site, considering different habitats will have different light environments.

P. 12 L. 531-532: Please provide additional details about these qualitative descriptions.

P. 15 L. 676: I suggest adding a short explanation on what the tradeoff spectrum tells us.

Fig. 1: It is difficult to match the boxplots with their corresponding family. I suggest trying to improve the alignment between family labels and plots.

Fig. 2: This comment is more of an editorial nature (so I'll leave it up to the editor), but the legend for Fig. 2 seems excessively long. I'd suggest trying to cut it down substantially.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
I find this submission of high value and recommend it for publication. As I wrote in my previous review, I find this work exciting and important. The authors show biofluorescent signals in 39 genera (152 species) of frogs, representing 13 families, uncovering a visual signal that is relevant at a broad taxonomic scale. The exposure of a signal that has been virtually invisible to researchers make this an incredibly important finding. The authors addressed all concerns of the previous submission, added analyses that help with interpretation and clarified the testing and interpretation of the four criteria of Marshall and Johnson. The new tables help the reader through the immense dataset and will be a valuable resource for future work. This is a beautiful paper. I expect this is a launching point that will propel the fields of visual ecology, behavior, trait selection and speciation.

Reviewer #2

(Remarks to the Author)
I recommend accepting this revised manuscript for publication. The authors have improved its overall clarity and impact. Each of my original comments has been addressed, or appropriately rebutted, and the changes they have made to

incorporate the feedback from other reviewers has improved the manuscript.

I have explored their R code. Most of it is reproducible with minor adjustments, except for the "RandTest_UpdatedFromRevisions.R" script. All script files are well annotated, which is very helpful. See comments associated with feedback on code.

Reviewer #3

(Remarks to the Author)
The authors have addressed several of my and the other reviewers' comments satisfactorily, which has improved the manuscript. That being said, there are still issues with the revised manuscript that need to be addressed, and some of the changes since the previous version have actually added new issues. Specifically, the authors have narrowed down onto the subset of their results that fit the Marshall and Johnsen criteria, obtained from about half of the species and ⅓ of the measured individuals, and have not addressed the rest. In my view this has weakened the manuscript. Below I elaborate on these issues. Without addressing the first three I cannot recommend this paper for publication, the remaining are suggestions that I strongly believe would make the paper better, but are ultimately up to the authors.

1. The statistical analyses are still problematic.

First, although the authors did use species as the unit of analysis in the second randomization test for criterion 2 and 3, other analyses and the text still use individual frogs as the unit of analysis. The issues with this approach were detailed in the previous round of reviews. The authors should use one measurement per species in *all* analyses (except for those comparing excitation light source wavelengths with irradiance curves, which don't involve frogs). Furthermore, when data are reported and interpreted in the text and figures (e.g. Figs. 1 and 6), this should be done in terms of between-species, and hopefully also within-species variation. The interpretation of any pattern will be different if the data come from 5 individuals of the same species versus one individual from each of 5 species.

Second, I disagree with the author's justification to not account for phylogenetic structure in their data. It is true that highly unbalanced sampling can lead to spurious results in some cases, but then again not accounting for phylogeny at all can do so too. In reality the majority of cross-species datasets have missing data and are limited to some degree by opportunistic sampling. This study's dataset spans multiple families across the amphibian tree, and has different levels of sampling across clades, so it is bound to have a high level of phylogenetic non-independence between measurements. The authors should, at the very least, account for phylogenetic structure in their statistical analyses, for instance using phylogenetic generalized least squares for their ANOVA-type analyses. Testing how emission wavelengths compare with irradiance and sensitivity curves in a phylogenetically-informed way will require some more thought. In the current approach, the null hypothesis is that all species (or individuals) are equally likely to reflect any wavelength, and results show that the observed emittances are instead clustered around specific regions of the wavelength spectrum. Whether this clustering is due solely to phylogenetic inertia is not addressed. A way to address could be, for instance, testing how different models of continuous trait evolution fit the data.

Further phylogenetic comparative analyses to address the evolution of biofluorescence would definitely strengthen the paper, but saving these for future publications is fine too.
2. In this version the authors decided to focus mostly on the green fluorescence peak, minimizing the attention given to the orange peak. In doing so, they have restricted their interpretation of the data to 1/3 of the individuals tested, which, in my view, considerably weakens the impact of the paper. I believe orange fluorescence is a central feature of the data, which should be directly tackled when interpreting it, even if it doesn't fit the criteria being tested. This is especially relevant considering that, as far as I could tell, a large number of species are polymorphic in this regard: they contain both green-emitting and orange-emitting individuals. If green fluorescence has an important ecological function, why don't all individuals exhibit it? Intraspecific polymorphism is not predicted by most models of positive selection, and is more often associated with balancing selection. Considering the central axis of this paper is an adaptive explanation for fluorescence (i.e. ecological tuning), intraspecific polymorphism should definitely be addressed in this paper.

3. The authors focus on highlighting the species that do fulfill the four criteria, but don't address the fact that nearly half (44%) of the species don't. If fluorescence appears ubiquitous across the tested species, but is only ecologically tuned in about half of them, how can its existence be explained in the remaining species? Perhaps the criteria are too stringent? These seem like key questions to address when evaluating the evidence for ecological tuning of biofluorescence in amphibians.

4. On a related note, Reviewer 1 pointed out the previous version of the paper had some "broad overstatements" (their words), which could be attenuated to improve the paper's impact. The authors' solution was emphasizing that their statements apply to the majority of tested species. Although 56% is technically the majority, it is also very close to one half. I recommend using more attenuated language, such as "about one half" to more accurately characterize the proportion of species that fulfill these criteria.

5. The authors justify not testing the perception of fluorescence by the red cones under twilight and moonlight conditions, since the light may be too low for cones to work under those conditions. This is a valid point, but I still think this test could be informative, considering that, as the authors point out in the response to reviewers, there is no information on the intensity of ambient (and therefore emitted) light experienced by frogs under twilight/moonlight conditions at the sampling sites, so it is possible that the emitted fluorescence could be above the sensitivity threshold for cones. As per Fig. S5 it seems like the red cone would perceive the orange fluorescence well (if it where bright enough), so this seems like a possibility worth

exploring.

Version 2:

Reviewer comments:

Reviewer #3

(Remarks to the Author)
I appreciate the authors' inclusion of a section discussing the Marshall and Johnsen criteria and the orange peak, as well as the newly added phylogenetic analyses. They have made the manuscript stronger.
My only remaining concern is that, as far as I could tell, some analyses only done at the individual level remain, and not all analyses have included phylogenetic non-independence. Since I have raised this issue multiple times already, I'll leave it to the editor to decide on the degree to which they think it should be addressed.

In addition, a minor comment for this new version is that in some parts of the text there is still language referring to the Marshall and Johnsen criteria as "required" for ecological significance. For example on line 798 "meeting both Criterion 2 and Criterion 3 is necessary to support ecological significance". I suggest combing through the text to avoid language of this sort, given that, as the authors put it (and I strongly agree), strictly enforcing these criteria to support ecological significance can miss biologically relevant scenarios.

Please find our responses to each reviewer comment below. Responses are ***bolded and italicized***. We believe we have thoroughly and effectively addressed each concern proposed by reviewers.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

This work is very exciting! The authors show biofluorescent signals in 39 genera (152 species) of frogs, representing 13 families. The authors uncovered a visual signal that is relevant at a broad taxonomic scale, but has been invisible to researchers to date, making this an important finding for (potentially) decade to come. Strengths of this paper include a novel finding on a wide taxonomic scale, which has the likely outcome of propelling the fields of visual ecology, trait selection and speciation forward.

1). I think the paper could be developed by attenuating some of the broad overstatements in parts of the paper – if done correctly, the impact could be even more powerful.

We have adjusted the broader statements of the paper, instead emphasizing that the majority of the frog species we tested for fluorescence (56.58%) have a maximum fluorescent emission that meets the four criteria for ecological significance proposed by Marshall and Johnsen. We also emphasize after these statements that the only fluorescent emission to meet these criteria are the blue light-induced green fluorescent signals that match the twilight environment and green-sensitive anuran rod (the environment and visual perception abilities unique to frogs) better than expected by chance. See lines 31-33, 159-161, and 396-399.

2). My impression is that the authors place too much emphasis on the expectations of these four criteria, without necessarily having everything in place to test those criteria. It's a nice framework, and I can see the appeal. However, this framework places an unnecessary emphasis on meeting criteria, rather than highlighting that it is interesting that while ~1/3 of the taxa meet the criteria ~2/3 do not. But I would argue that failure to meet these criteria does not necessarily mean that there is not biological function to these signals. (technically, you seem to be looking at species recognition, not ecological function). You could have failed to find 'ecological function' because you didn't' consider the visual pigment of that species, because it is relevant to heterospecifics or because the frog is active during the day (if I understand correctly, they used the common green rod and used twilight as background in all analyses).

As addressed in our response to Comment #1, we have now clarified how many individuals/species meet the criteria and that we are testing this within an intraspecific communication framework. The robustness of our findings was additionally strengthened by the analyses we added (now testing the fluorescent signals found against five different light environments and two different photoreceptor sensitivities). These efforts de-emphasized our previous focus on the criteria only under one environment/vision condition.

Additionally, we added material to the Discussion (see lines 384-392) about how lack of consistency with the four criteria does not mean that the biofluorescence is not ecologically irrelevant. Rather, we propose that it may produce a still visible but not optimal fluorescent signal (line 387-389), or it may be important for a different intended receiver (i.e. a predator or prey species) with different spectral sensitivities (lines 274-282).

3). The table is a nice summary of findings at the taxonomic scale – but could it be organized in a way that helps the reviewer make sense of all of these data (e.g., have a column for ecological function criteria met Y/N, a diurnal/crepuscular column, whether body parts match known visual signaling repertoire).

We have added Tables 2, 3, and S2 which better organize our findings (especially with the addition of extra analyses). Table 2 presents the results of randomization tests for Criteria 1-3 under each light environment and photoreceptor visual sensitivity condition. Table 3 lists the species/individuals meeting all criteria for ecological significance of their fluorescent signal. This list emphasizes in which species the function of fluorescence for intraspecific communication should be explored, with specific focus on testing for the role of visual communication in those species where visual signaling is unknown. Table S2 lists the fluorescent emission specifics and body location of the fluorescent recordings for each individual, again highlighting the lack of literature on visual communication in anurans.

4). Criterion 2. In lines 427/428 the authors say that they tested both diurnal and crepuscular/nocturnal frogs. I would assume that the diurnal frog contrast background would be daylight, but that wasn't clear. Perhaps a way to clarify this is to analyze the diurnal frogs separately? I like that both diurnal and crepuscular anurans are included, but their graphics/tables don't make it easy to follow this important distinction.

We have added this analysis to supplementary methods and materials. We completed randomization tests for Criteria 1-3 (as described in the main text, lines 146-153) under each light environment and photoreceptor visual sensitivity condition for diurnal individuals in daytime conditions. The results of these tests are present in Table S5. We did not find evidence that any of the fluorescent emissions of diurnal

individuals meet any of Marshall and Johnsen's criteria for ecological significance, but we emphasize the extremely limited samples sizes. We caution any strong interpretations from these results and suggest future studies should explore these correlations further with increased sampling.

5). Criterion 3. The authors test this for 'the most abundant rod in the visual system'. Line 302 is an overstatement. This seems like a big caveat, given that frogs are known to differ in visual pigments (as the authors point out). It seems worth testing whether biofluorescence is detectable if they considered a few of the alternate visual pigments known for frogs, including for diurnal frogs for the diurnal taxa.

This was a valid concern, and we made several efforts to clarify and address it. We have added in the citation (39) for the reference for green-sensitive rods being the most abundant rod in the anuran visual system to this statement (now line 237). Denton and Wyllie (1955) found that green-sensitive rods occupy about 60% of the retinal area while blue-sensitive rods occupy only about 8% of the retina in **Rana temporaria** ***(39). Additionally, unpublished data from Rayna Bell (California Academy of Sciences) and colleagues show that the peak emission and abundance of the visual pigments does not change substantially across most anuran species. For example, in the manuscript we state: "Frogs have blue-sensitive (peak absorption ~432nm) and green-sensitive (peak absorption ~500nm) rods." From Rayna Bell and colleagues' unpublished dataset, the range in peak sensitivity of these rods for the species in our study is 425-450nm for blue-sensitive rods and 495-514nm for green-sensitive rods. There is a maximum of a 14nm deviance from the peak sensitivity of the green-sensitive rod to the peak sensitivity of this rod utilized in the analyses, which does not affect the results of our study. Therefore, given the current visual sensitivity data available, we think it is appropriate to utilize the green-sensitive anuran rod to test Criterion 3 of the study.***

Additionally, to address the second concern, we added tests utilizing an alternate frog visual pigment, the red-sensitive cone photoreceptor. We also performed supplemental tests with only diurnal frogs. Please see response to Comment #4 above for specifics on diurnal frog tests.

6). There is a missing citation for line 306.

Added citation (36) here (now line 189). Reference for most frogs mating in dim light conditions.

7). Having two rods and one cone make it possible (but not certain) that the frog has color vision (you need three visual pigments for the possibility of color vision). This is oversimplified in the text at the expense of being incorrect.

I have changed the wording, and added a citation, to better reflect that researchers have found that having two rods allows for color discrimination in dim light conditions (line 260-262).

8). Whether the frog has color vision in low light requires visual modeling and behavioral testing. Line 309 says 'most frogs mate in dim light' but that isn't true for diurnal taxa. They mention this as a caveat (line 433) – but instead of a caveat, they could use a visual pigment model from a diurnal taxon.

Yovanovich and colleagues (2017) found color discrimination in low light via behavioral testing, which we now more clearly discuss (see lines 260-262). To address the second concern, we also added additional tests utilizing a cone visual pigment and analyzed diurnal taxa separately (in Supplemental methods). See response to Comment #4 above regarding testing of diurnal individuals against cone sensitivity.

9). If I understand correctly, both green and orange peaks provide a contrast in dim light. However, only green peaks are detectable using the green rod sensitivity. There are 316 individuals that have orange peaks, which then, are not detectable. Is it possible that those species have different visual pigments? The authors offer up the alternative hypothesis that orange is a signal to heterospecifics. Given that more individuals reflect biofluorescence outside the range of the common anuran green rod, it is worth asking if it is variable frog visual pigments that co-vary with this signal.

As stated in the response to Comment #5, unfortunately, there are no published data for the spectral sensitivity of each visual pigment for each species, though there does not seem to be much variation between species. However, we did add additional tests against other light environments and photoreceptors to help address this comment. Additionally, we have redefined the blue light-induced green and orange peaks to take into account individuals with a bi-modal fluorescence emission spectrum (see Methods lines 579-588).

10a). Criterion 4. Criterion 4 is important for this framework, but I don't think they have the data to test it effectively – or you are not using it to it's potential. That is, what is the alternative, given that you tested every part of the frog body that could be tested?
I have two competing thoughts: While it's true, that there are some generalities about frog signaling, visual signals are species specific – and only after behavioral study do we really understand how visual signals are displayed and perceived (as the authors point out in line 397). The authors provide a wonderful example in D. parviceps. The challenge is, therefore, matching the biofluorescence body part with what we know about visual signaling in each species.

10b) My second thought is that this study opens the possibility that we don't really understand visual signaling in frogs – and that our understanding of signaling is constrained to the signals that we can perceive (movement, calls). This dataset shows that nearly every part of the frog could contain a signal – even those regions of the body not typically associated with visual signaling (e.g., eyes, vent on most frogs).
Identifying the parts of the body that emit biofluorescence is a great starting place to generate hypothesis to test about visual signaling for any particular species. That is to say, I think the authors miss an opportunity to describe to best use this body part data. The eyes are a great example. Most eye color don't reflect in visible light. It opens doors to think that have identified a biofluorescence signal in the eye -- this makes eye size and shape an interesting character – one that has been overlooked. Is there a more effective way to use these data?

Responses to 10a and 10b: We agree with both comments. We have added Tables 3, and S2 which better emphasize the lack of information published on visual signaling in frogs. We also discuss this in line 288-292. Table 3 lists the species/individuals meeting all criteria for ecological significance of their fluorescent signal and emphasizes in which species the function of fluorescence for intraspecific communication should be explored, with specific focus on testing for role of visual communication in those species where this is unknown. Table S2 lists the fluorescent emission specifics and body location of the fluorescent recordings for each individual, again highlighting the lack of literature on visual communication in anurans.

In addition, we highlight proposed biological functions of fluorescence in the observed body locations based of the function of fluorescent signals in those body locations documented in other species (lines 315-355). We hope the combination of the added tables, discussion of lacking literature, and proposed functions emphasize the importance of examining the potential for visual signaling in these frog species as the next step.

Reviewer #2 (Remarks to the Author):

Overall comments:

I recommend this manuscript for publication after minor revisions. This manuscript examines biofluorescence in anurans. The authors carefully consider biofluorescence through the lens of ecological significance. They provide a sufficiently thorough review of prior studies of this phenomenon in amphibians and other vertebrates. They clearly present

their questions, hypotheses, and results. The major conclusions of the study include that 1) biofluorescence is most intense in response to blue excitation light, 2) that the resulting green fluorescence would stand out in chromatic contrast to the light environment during periods of activity for most anurans (dim lighting, twilight), and 3) that the emission spectra generally align with our understanding of photoreceptor cell sensitivities in this group. Their findings will have a substantial impact on our understanding of the evolution and relevance of biofluorescence in amphibians, and more broadly, in tetrapods. With some minor modifications and clarifications to the text, the study is reproducible. The changes below are recommended to improve clarity and to reduce redundancy throughout the manuscript, including in some of the figures. Overall, this was an enjoyable, interesting, and well-composed manuscript.

Line by line comments

Introduction

11). >>> Overall: Succinct and well written. Important context is given for the study and a thorough review of existing literature is provided. The question, hypotheses, and objectives are clear.

Thank you.

12). >>> Lines 57 - 60. Recommend revising text to clarify the point. Some of the potential functions mentioned in this text for fluorescence are hypothesized functions, but were not explicitly tested in the papers cited (e.g., camouflage in fishes, Sparks et al.). If there is a way to make that point clearer it would further emphasize the novelty of this study which does propose and test a hypothesis regarding the function of biofluorescence.

We have now clarified this, splitting the sentence into two to separate experimentally tested and hypothesized functions (lines 57-61).

13). >>> Line 75. Recommend changing "at another wavelength" to "at other wavelengths." Some fluorophores seem to be capable of emitting more than one wavelength of light.

Good suggestion, changed.

14). >>> Line 83. Recommend revising text to clarify the point. Taboada et al.'s work was perhaps the first publication to focus on biofluorescence in anurans and to document patterns in whole organisms. However, there may be earlier publications in the fields of developmental biology/embryology that document "autofluorescence" in anurans (there are cases of this for caudates).

Clarified we are discussing <u>macroscopic</u> biofluorescence.

15). >>> Lines 106 - 109. Recommend revising to break up into shorter sentences.

Done. Split into three sentences (now lines 105-108).

Results

16). >>> Overall: The **results** are easy to understand, presented in a logical manner, and adequately address the questions and hypotheses posed in the introduction. However, there are some areas where the authors should revise the text to improve clarity. There are also areas where the results and discussion seem redundant (i.e., the same information is presented in nearly identical terms). This section and the discussion should be reviewed and trimmed down to reduce repetition.

We have checked the entire document for redundancies to improve clarity, rewording/rewriting those areas where the same information was presented in nearly identical terms. Reviewer #3 suggested a combined Results and Discussion section to address this concern. We have taken that suggestion which eliminated the areas where the same information was presented in nearly identical terms.

17). >>> Lines 117 - 123. Optional revision if more space is needed elsewhere. These lines detail how many more taxa this study has surveyed for biofluoresence. While important, this section could be shortened or simplified if additional space is needed elsewhere in the manuscript.

Agreed. Will simplify here if more space needed elsewhere.

18). >>> Lines 125 - 127. Recommend revising text to make fewer assumptions. The authors state that "novel biofluorescent patterns…were missed in previous studies because the incorrect excitation wavelength was used." Recommend revising to emphasize that the wavelengths used by other studies may not have excited the most intense fluorescence. If the choice of excitation wavelengths in other studies was driven by a particular prediction or hypothesis, that choice may not necessarily have been "incorrect."

Reworded (now lines 129-132). Removed the "incorrect" classification. Fixed throughout manuscript.

19). >>> Line 127 - 131. Additional point of clarification may be needed. In some previously published studies, like the one mentioned here (Thompson et al.), authors may not have used a barrier filter to survey for fluorescence. This could mean that reflected excitation light "drowned out" any fluorescence. So in the cases of the species mentioned, was the "incorrect" excitation source used or were fluorescent wavelengths missed for the wavelengths tested?

Added another two sentences to clarify that the lack of utilizing a barrier filter

could have led to misinterpretations of presence/absence of fluorescence as well (lines 132-135).

20). >>> Line 151. Grammatical error. Capitalize "blue light" towards the end of the line.

Fixed (now line 177).

21). >>> Lines 162 - 177. Clarification needed regarding the type of contrast being presented. When talking about contrasting against the background, the authors are focusing on chromatic contrast. Organisms can also exhibit achromatic contrast, meaning that they differ in brightness within a particular chroma or across all chroma. Endler (2006; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1634903/) briefly discusses how the type of contrast may be ecologically and evolutionarily relevant. The authors should revise the manuscript to clarify that they are only considering chromatic contrast.

Clarified we are only considering chromatic contrast (lines 221-222).

22). >>> Line 219 - 227. Recommend revising so that interpretation occurs in the Discussion, and so that it is clearer what is speculation/prediction versus reporting of their results. For example, "could result from secretions from the frog skin" would be more appropriate in the discussion. Additionally, are the authors reporting that ventral biofluoresence was widespread, or that it could be widespread? Phrases like "could be widespread," and "could be present in distinct regions" should be revised to "was widespread" and "was present in distinct regions."

These were all different body locations/patterns of fluorescence we observed. Reworded to clarify that these are results, not interpretations (changed wording from "could" to "was"; lines 304-314).

23). >>> Line 203 - 235. Recommend breaking this large chunk of text into more than one paragraph.

Done. Split into two paragraphs (now lines 293-314).

Discussion

24). >>> Overall: The discussion and results need to be streamlined so that there is less repetition between them.

Done. Removed redundancies in part by combining Results and Discussion sections.

25). >>> Lines 239 - 298. Recommend revising these sections so that they are less repetitive with respect to the results sections. Also see comments in results regarding the use of some terms (e.g., "incorrect excitation wavelength"). The authors should (and do)

reference specific results in their discussion. However, in its current form, large portions of the discussion seem highly repetitive with respect to the presentation and interpretation of information in the results section.

Done. Removed redundancies in part by combining Results and Discussion sections.

26). >>> Lines 335 - 337. Recommend revising to only used "additional" once.

Fixed. Reworded. Now lines 274-282.

27). >>> Lines 369 - 373. Recommend revising to be less repetitive with the results. This text is very similar to text previously presented.

Done. Removed redundancies.

28). >>> Line 402 - 404. Recommend revising for clarification. Did the authors examine vocal sac fluorescence when the sac was inflated? It may be that fluorescence changes in intensity when this region of the body is stretched.

Clarified this in Reporting Summary instead.

Materials and methods

29). >>> Overall: The materials and methods are relatively clear, and allow the study to be repeated. There are a few points where additional details are needed.
For the three replicate spectra collected, did the authors move or adjust the probe position between readings? Or were spectra collected via a "rapid fire" type approach? Meaning that while the probe and animal were stationary, three measurements were collected.

Clarified. Measurements were focused on the same location with minimal movement between recordings (unless frog movement determined otherwise)- line 479.

30). >>> The authors mention the placement of their probe during spectra acquisition, but what about the position of the light source? They note it was suspended via a gooseneck tripod stand, but not whether that was adjusted over time or with different specimens. I believe that potential movements of their light source may be accounted for in their method of calculating % biofluorescence intensity. However, this is worth pointing out so that others can better understand their approach.

Clarified. The gooseneck tripod stand was always at the same height/distance above the frog when possible (line 465). However, there was slight variation in how far away we held the frog beneath the light source, hence why we utilized the method of calculating percent biofluorescent intensity to account for these differences.

31). >>> Please clarify the settings used with the Nikon digital camera. Details such as f-stop, exposure time, and ISO settings are relevant to interpreting the figures presented.

Added camera settings to the Reporting Summary.

32). >>> Have any other publications approached calculating the maximum percent biofluorescence emission in this way? If so, they should be cited. If I understand the methods correctly, the intensity at peak excitation wavelength and the intensity at peak emission wavelength were calculated from the same spectral reading. I believe this would mean that the intensity at peak excitation wavelength is actually the intensity of the reflected excitation light. I think this value would be lower than the total intensity of the light coming directly from the excitation source due to absorbance, scattering, or transmission. I don't think that this point of clarification would mean that the results would differ dramatically from how they are presented, but it is an important point to clarify and consider.

The reviewer is correct. We have clarified that our focal measure is actually one of intensity of fluorescence relative to the amount of reflected excitation light (lines 494-495). To our knowledge, we are the first to calculate the maximum percent biofluorescence emission in this way.

33). >>> Blinding as described in the supplementary files. The authors specify who was handling the amphibian and spectrometer probe, and who was viewing the OceanView application. They specify that the individual handling the amphibian was "blind" to the fluorescence readings so as not to preferentially select a body region for measurement. Was the person handling the amphibian and probe also wearing filtering glasses so that they could target regions that were fluorescing? If not, it seems like the measurements taken could be underestimating fluorescence if the region measured was not in fact the most fluorescent region.

That is correct. Clarified in the Reporting Summary document that the person handling the amphibian was also wearing filter glasses to target fluorescing regions.

Figures and tables

34). >>> Overall: The majority of the main figures and tables are necessary and well designed. There are a few that should be split into more than one figure, or otherwise modified to improve readability. None of these are major changes.

Thank you.

35). >>> Figure 1. See Comment #5 in discussion.

Comment #5 in discussion is:* "Recommend revising for clarification. <u>Did the authors examine vocal sac fluorescence when the sac was inflated</u>? It may be that fluorescence changes in intensity when this region of the body is stretched."** ***We clarified this in the Reporting Summary. Unclear how this is related to Figure 1.

36). >>> Figure 2. Overall this is a great figure. Some minor modifications will help with clarity and readability. The depiction of the "shift" from excitation to emission spectra doesn't need to be included in each of the spectra plots for Criteria 3 and the Observed Data section. It contributes to the Observed data being a bit challenging to interpret at first glance. If the authors want to maintain it in those parts of the figure, then can a different type of line be used to distinguish between the excitation and the emission spectra? Perhaps a dotted line with circles for the former and with squares for the latter?
The caption for Figure 2 is very long and should be shortened so that it is not so repetitive with the results section. Limit the text in the caption to the essentials for understanding the figure.

We believe it is important to keep both the excitation and emission spectra to emphasize that different parts of the fluorescence spectra are evaluated for each criterion. We changed the line type so that excitation and emission spectra now differ in color and line type (blue dotted vs. green dashed). We also changed the format of presentation of the Criterion 4 observed data for clarity. The figure caption has been shortened.

37). >>> Figure 3. Some minor modifications will make this figure more readable.
Is there a way to reduce the number of different text sizes used? Some of the sizes used are very small (e.g., for median values and for the axes labels in the inset plot).
I am not sure that the inset is necessary here. It also seems to be covering some of the data about distributions of emission intensities.

We have increased the text sizes across the figure and decreased the number of text sizes used to improve clarity. We moved the median values outside of the plot to not cover any data as text sizes increased and to improve readability. We also removed the inset.

38). >>> Figure 5 - 7. Modifying how the data are presented will improve clarity. I understand that where the open circles fall on the twilight spectrum is meant to indicate what wavelengths of fluorescence were observed. However, it also implies that those irradiance/intensity levels were observed, which I think is what makes the plots confusing. I encourage the authors to consider a different approach where vertical rectangles or bars are used to indicate what wavelengths of fluorescence were observed. This would emphasize

that the only information being presented for fluorescence has to do with the wavelengths fluoresced, not with their intensities. The average wavelength for emissions could be presented as a bold line.

We liked this suggestion and tried the Reviewer's suggested alternate plot with vertical bars, but that decreased clarity in the relationship between the fluorescent wavelength and the proportion of those wavelengths available in the environment. It also made it difficult to visualize/connect the observed data to the simulated dataset. Instead, we have clarified in the figure caption that the irradiance levels were not observed in this study to help address this concern (lines 959-961).

39). >>> Figure 8. Moderate revisions to the figure needed.
I recommend splitting the figure into two figures. One that focuses on the relative proportions of fluorescence by body region and excitation source (pie charts). And a second that focuses on the images of anurans under white light and while fluorescing under royal blue light. This would allow the authors to maximize space and potentially have the white light images be larger. Larger images will facilitate a better understanding of where fluorescence is occurring and what features might be contributing to that fluorescence. Since the binomial names are given in the figure caption, they can be removed from the figures themselves. This would give the authors more space and make the figure less crowded.

We liked the suggestion of emphasizing the images of fluorescence. Unfortunately, we are currently at our maximum number of figures/tables for the main manuscript. So, instead we have readjusted the format and layout of this Figure (now Figure 6) to increase the size of the white light photos and the fluorescent photos. We also removed the binomial names from the figure since they are given in the figure caption.

Supplementary materials
40). >>> Figure S1. Very small text sizes for median and for X axis labels. These can be made larger.

Increased text sizes.

Reviewer #3 (Remarks to the Author):

The authors conducted an impressive survey of biofluorescence in several species of neotropical amphibians. Using their newly generated data, they evaluated the ecological

significance of fluorescence in frogs based on a set previously proposed criteria.

I believe the dataset generated by the authors is a significant contribution to the ongoing effort to understand the ecological and evolutionary significance of fluorescent coloration in amphibians. The data were collected in what seems to me as a rigorous manner (I am, admittedly, not an expert in biophotonics), and significantly expand the phylogenetic breadth of biofluorescence reports in amphibians. Unfortunately, there are several issues and shortcomings with the analysis, interpretation, and presentation of the data that preclude me from recommending this paper for publication in Nat. Comm. in its current form. My concerns and some specific comments are detailed below.

>> General comments:

1. The paper is generally well written, but I found getting through the Results and Discussion sections somewhat challenging due to two main reasons:

41). i. Although the methods section of the paper is detailed well written, the Results section has very little methodological detail for a "results-first" paper, which makes reading the methods and discussion difficult at times. I suggest the authors add basic details/summaries of their methodology to each section of the results to facilitate reading.

Done. Added brief overview of methods to improve clarity before each section of the results (lines 116, 141, 146-152).

42). ii. Many of the results are repeated in great detail in the discussion section, which makes the text feel repetitive and and confusing at times, and dilutes the points being made in the discussion. I found a few stretches of nearly identical text between both sections (e.g. lines 219-235 vs. lines 369-384), so perhaps this is due to a previous reformatting of the manuscript? In any case, I suggest either succinctly summarizing key results before discussing them, or consolidating the results and discussion into a single section.

Done. To help remove redundancies and improve clarity, we have taken this reviewer's suggestion to consolidate the results and discussion into a single section.

43).

2. Most statistical analyses presented were conducted at the level of individual frogs, either across all species or within taxonomic families. This approach essentially assumes that all individual measurements are equally independent from one another, which is not the case in the author's data, considering they involve measurements on multiple individuals from several different species. This creates structure due to 1. variation within versus between

species, and 2. the different levels of relatedness between species (i.e. phylogenetic structure).
The authors should account for structure in the data when doing statistical analyses by 1. Consolidating within-species data into single measurements per species (and perhaps associated variance/uncertainty), and 2. Using phylogenetically-aware statistical approaches to analyze species-level data. There are a few publicly available large-scale phylogenies of amphibians, so this should be relatively straightforward.
Beyond statistical rigor, embracing the inter/intraspecific and phylogenetic structure of the data could help the authors address some of the open questions poised in the discussion. For instance, investigating whether orange and green peaks are polymorphic or fixed within species can speak to whether they serve different functions, and mapping many of the scored traits onto a phylogeny could help gauge if and why they are more labile in some groups than others.

1. We consolidated within-species data into a single measurement per species as directed. We then performed our randomization tests by producing 10,000 subsamples of the individuals. For each subsample, we chose one individual from each species at random and repeated the randomization test, calculating test statistics and null values as for the original analysis by species. All of our results remained consistent across individual and "species" datasets.

2. We completely agree with the reviewer that a phylogenetically-aware approach should be used for our analyses. We are limited, however, by opportunistic sampling that resulted in missing data in our matrix. As a consequence, a phylogenetic approach could only possibly be applied to limited, small groups or within a handful of species on the amphibian tree. Due to the likelihood of obtaining a spurious result, we chose to not apply these methods here but rather wait until we have a more complete data set. We agree, however, that in future studies it will be very interesting to examine the lability of orange and green peak mono- and polymorphism across the amphibian tree.

44). 3. From Table 1 it seems like all the species tested in this study displayed biofluorescence. Is this the case, or were only species that displayed biofluorescence reported? If indeed all ~150 species tested fluoresced, this seems like a major result worth reporting and discussing. Is biofluorescence perhaps ubiquitous in frogs?
If some of the tested species did not display biofluorescence, then they should be reported, and perhaps used in analyses, for example to assess when this trait has been gained or lost in amphibian evolution.

That is the case, we found <u>some</u> level of biofluorescence is ubiquitous in frogs. We have highlighted this ubiquitous result in the abstract (line 28) and our results section (line 128).

45). 4. I found the way the authors used and interpreted the twilight spectrum problematic for three reasons.
First, although this wasn't specified in the manuscript, looking over the Cronin et al. book it seems like the spectrum used was the one in Fig 3.18 which depicts a "typical habitat" at twilight. Cronin et al. don't specify how or where this spectrum was taken (or I didn't find this information), but considering "Under forest canopy" is one of the alternative habitats depicted, it seems like this spectrum represents twilight in an open habitat. This may pose issues for the interpretation of the paper's data, considering that many (if not all) of the sampled localities are forested, and that tree cover is well-known to influence the irradiance spectrum, so the species sampled may be experiencing different irradiance spectra.
Second, it is worth considering that each of these localities may have different irradiance spectra, so their absorbance/irradiance measurements may not be equally comparable with a single spectrum.
Finally, the presented spectrum does not have irradiance values below 400nm, which complicates the interpretation of results obtained with the UV light source, which is below this wavelength.
The authors should therefore:

1. Provide more information on the nature of the used twilight irradiance spectra, and

2. Try to use spectra that represent the light available in the species' habitat as accurately as possible, and that include all of their excitation light sources. I understand that obtaining irradiance spectra specific to each sampling site may be difficult, but if available, I would suggest at least using the irradiance spectra of one or a few tropical forests at twilight. If such spectra are not available the authors should highlight these caveats upfront to better reflect the limitations of their interpretation.

We have provided more information on the nature of the twilight irradiance spectra throughout the manuscript (lines 189-196). In addition, we added additional analyses under four other light environments (Daylight, Under Forest Canopy, Full Moon, and Starlight). These were the available irradiance spectra (Cronin et al. Visual Ecology) relevant to frog behavior/ecology. Unfortunately, we still have a limitation of no below forest canopy during night hours spectrum under which to test our data. However, we more clearly define these limitations in the manuscript now (lines 589-601).

46). 4b. I fully agree that twilight is a very important context to study amphibian

biofluorescence. However, considering twilight makes up a small fraction of a nocturnal frog's active period, I believe night time conditions (especially with strong moonlight) should be considered as well. Irradiance spectra at night are markedly different from twilight ones, which is highly relevant given the paper's strong focus on the spectral properties of biofluorescence and their relationship to ambient light.

We followed the reviewer's advice (also see *response to comment #45 above)* *and re-ran the analyses with all five relevant environment spectra provided in Cronin et al. Visual Ecology book to address these concerns. Additionally, we added the following information about twilight to the manuscript: In the manuscript we state: "Twilight is defined as the light environment during the time between sunset and full night when the sun is between 0° and 18° below the horizon (27). Illuminance values during twilight range from 103 lux to nearly 10-4 lux, decreasing approximately a millionfold from when the sun is 10° below the horizon to when it is 15° below the horizon (27). These illuminance values are often the light levels which we consider "nocturnal" activity for frogs (48). Hence, twilight conditions are the most relevant environmental spectra in which to test the hypothesis of a function in intraspecific communication." (lines 189-196). Therefore, moonlight conditions are likely not the most relevant environmental spectra in which to test the hypothesis of a function in intraspecific communication.*

47). 5. Blue light was found to produce, on average, the highest fluorescent emittance. Based on this, and the fact that this light source is the closest to the short-wavelength peak of the used twilight irradiance spectrum, the authors restricted most of their subsequent analyses to the fluorescence observed under blue light. However, most of the other light sources are also well represented in twilight (and moonlight) spectra, and often produced comparably high emission intensities, so they could produce enough fluorescent light under natural conditions to be perceivable by potential receivers.
With this in mind I think it is warranted that the authors conduct their analyses for measurements gathered under all light sources. This would not only add to the study's biological relevance, but could help the authors evaluate criteria 1-3 better by allowing them to compare the irradiance and perceivability of fluorescence stemming from different excitation wavelengths.

We re-analyzed our emission data for all excitation light sources as recommended by the reviewer with two exceptions. Due to the limitations of the available irradiance spectra only having recordings for 400-700 nm, the fluorescent excitation and emission could only be evaluated for those recordings within this range. So, UV excitation could not be statistically evaluated in respect to Criterion 1, though the relative amount of ultraviolet light at ground level is estimated to be only 4%, in comparison to 44% visible light due to the blocking of UV wavelengths by the

atmosphere (46), so we estimated that these UV wavelengths would not be dominant in the daytime environments (Table 2).

Additionally, the emission of cyan and green-light-induced fluorescence extended beyond 700 nm and often into the infrared range (Fig. S1). While likely not relevant to anuran sensory biology, we suggest this unique fluorescent pattern be explored further with respect to potential predators with infrared detection capabilities (such as snakes). We could not analyze these emission wavelengths with our currently available irradiance spectra. We did, however, analyze all excitation and emission data possible with the currently published (and comparable) irradiance spectra.

We also clearly state these limitations in the manuscript (lines 589-601).

48). 6. It is not entirely clear to me why the authors used the available light (i.e. the irradiance spectrum) as the background against which the fluorescent signal's contrast was evaluated. As I understand it, a visual signal's contrast is perceived against the colors of objects adjacent/around it. For example, a frog's color will be perceived as conspicuous if it is very different from the forest floor (when seen from above), or a spot on that frog will look conspicuous if it is surrounded by patches of a very different color (e.g. yellow spots on a black background). Maybe this is because the signal is emitting its own light? Please explain this choice in more detail.

On a similar note, given the low light conditions, wouldn't it make sense to incorporate the intensity of the fluorescent signal when estimating contrast?

We used the available light (irradiance spectrum) as the background against which the fluorescent signal's contrast was evaluated because these are the wavelengths of light that will be reflected by the environment around the frog (assuming no other fluorescence of the surrounding environment, which, admittedly, is likely not the case). We have now defined this assumption in the manuscript (lines 222-225). The next step for evaluation of these signals is to be more specific (e.g., by examining the signal vs. the color reflectance and fluorescence emission of the specific leaf each frog was sitting on, but unfortunately, we do not currently have those data as those measurements did not fall within the scope of this study).

Additionally, we did not incorporate the intensity of the fluorescent signal when estimating contrast because our study is only evaluating the chromatic, not achromatic, contrast (now clarified in the manuscript based on other Reviewer comments, lines 221-222). This choice was made due to a lack of data of intensity of the background emission wavelengths (we have abundance information from the irradiance spectra but not intensity information). However, our results should be consistent across varying intensities because the overall abundance of available

wavelengths in the environment and the fluorescence emission are both correlated in the same direction with intensity (because fluorescence intensity is dependent on how much of the excitation wavelength is available in the environment).

49). 7. The authors argue that criterion 3 is met because the green observed in some individuals matches the sensitivity of ranid and bufonid the green rods. However, the analyses performed provide no evidence for the orange peak passing criterion 3, beyond some speculation about it being perceivable by non-anuran receivers. Considering the orange peak is present in about twice as many individuals as the green peak (316 vs 192), the support for criterion 3 seems ambiguous. Given its high frequency, for these data to support criterion 3 the authors would have to show that the orange peak is perceivable by an ecologically significant receiver.

We redefined the green and orange peaks to include bi-modal emission spectra with two peaks and described these patterns in the manuscript (see Methods lines 579-588). Additionally, we have adjusted the broader statements of the paper, instead emphasizing that the majority of the frog species we tested for fluorescence (56.58%) have a maximum fluorescent emission that meets the four criteria for ecological significance proposed by Marshall and Johnsen. We also emphasize that the only fluorescent emission to meet these criteria are the blue light-induced green fluorescent signals that match the twilight environment and green-sensitive anuran rod (the environment and visual perception abilities unique to frogs) better than expected by chance.

50). 8. When evaluating whether the emitted fluorescent light would be visible to frogs the authors only use sensitivity data of rods. However, the emitted fluorescent light should be bright enough for cones to perceive it in twilight and possibly moonlight conditions. With this in mind, I suggest the authors consider cones too when gauging receiver sensitivity. For example, the red cone of Rana temporaria has a peak sensitivity around 560nm (see doi.org/10.1111/j.1748-1716.1994.tb09790.x), so it should be well able to perceive the orange fluorescence detected in some frogs.

We added analysis against the red-sensitive anuran cone for environments in which it is relevant. However, we want to clarify that it is not reasonable to assume that the emitted fluorescent light should be bright enough for cones to perceive it in twilight and possibly moonlight conditions. As found by Yovanovich and colleagues (2017), only rods are used at these low light levels, and fluorescence always reemits less light than available in the environment (because of energy lost in the process). So, fluorescence will never be brighter than the available amount of light at twilight, and hence only frog rod photoreceptors will be activated, not cones, under twilight conditions.

Specific comments:

51). P. 2 L. 95 (and elsewhere throughout the text): I agree that the availability of light that actually excites a fluorophore to a biologically relevant degree is key here, but it doesn't seem to me like this light has to necessarily be of the most abundant wavelengths in the environment. As long as there is enough light available in the environment for fluorophores to emit a biologically a significant amount of light back into the environment the possibility of a biological function is there. Of course, the fact that fluorophores in many of the tested frogs absorb the very prevalent blue wavelengths is very good evidence for biological meaning, but I suggest the authors rephrase this as evidence in favor but not a requirement of biological function here and elsewhere in the text.

Discussed in lines 383-391, clarifying that even if a wavelength is not dominant in the environment that it can still excite fluorescence. Reworded to clarify Marshall and Johnsen's criteria may suggest biological meaning but are not "requirements" (line 94).

52). P. 3 L. 127: I suggest replacing the word "incorrect" for something else. There are no incorrect wavelengths to use in experiments, just some that produce higher/lower levels of emittance.

Addressed in previous Reviewer's comment (#18). Reworded and removed the "incorrect" classification.

53). P. 4 L. 151: Capitalize the word "blue"

Done.

54). P. 4 L. 182: It may be worth adding a sentence or two to remind readers about the fact that frogs have a dichromatic rod system.

Added. Now lines 235-238.

55). P. 5 L. 212: It isn't clear why the authors refer to Fig. 2 here. Perhaps they meant Fig. 8? I found the same situation in a few other places after this.

Clarified that we are referring to the observed data presented under Criterion 4 within Figure 2. The data is the same as the middle pie chart in Figure 8 (now Fig. 6), but the specific percentages are presented in Figure 2. Additionally, to avoid confusion with body location data from other light sources (in Figure 6) we point the reader to Figure 2 instead.

56). P. 5 L. 223: Please clarify if orange was more frequent in ventral than dorsal.

We have reframed the results to not focus on orange fluorescence, so this comment is no longer relevant to this line. However, we added Table S2 which lists the fluorescent emission specifics and body location of the fluorescent recordings for each individual, documenting what fluorescent signal was found where for each individual.

57). P. 5 L 229: I suggest providing citations that support this statement.

Citations were provided but after the Figure reference in the parentheses, edited to highlight sources (now line 320).

58). P. 6 L. 366: It is unclear to me what the authors mean by examining variation "upon which selection can act". From what I understand most if not all of the phenotypes examined in this paper (e.g. absorbance and emittance wavelengths, intensity, location of fluorophores) are heritable (at least to some degree) and variable, and therefore susceptible to natural selection and able to evolve due to it. Please clarify or rephrase.

We have removed this line to improve clarity.

59). P. 7 L. 324-327: Fig. 1 in Taboada et al. shows that fluorescence emission peaks at excitation wavelengths between 390−430nm and then starts dropping until ~450nm, which was the highest excitation they tested, so this particular species may not have a higher emission with blue (440-460nm) excitation wavelengths.

This reviewer is correct. The most intense fluorescence in **Boana punctata** ***is excited by violet light. We did not test*** **Boana** ***punctata in this study. We found a general pattern across individuals of blue excitation wavelengths producing the most intense fluorescence in our study (Figure 3); however, this pattern is not necessarily consistent across all individuals/species. For example, in our study the most intense fluorescence in*** **Boana atlantica** ***was excited by violet light (Table S2). Our results do not contradict Taboada's, but instead describe general patterns across the species we tested and show that some species (like those*** **Boana** ***species Taboada and colleagues tested) may not follow this general pattern.***

60). P. 9 L. 409-412:It is not clear to me why fluorescent signals are expected to be less involved in interspecific vs intraspecific communication. Please elaborate.

We only evaluated our fluorescent signal data for ecological significance within the framework of intraspecific communication in this study, which is what we found evidence for (a correlation between twilight light environment, frog vision, and their fluorescence). Hence, these lines of text are referring to this finding of the fluorescent signals being tuned to frog behavior/communication and perhaps not interspecific

communication. We do not think interspecific communication is absent but rather it was beyond the scope of this study. We have added text to the manuscript to clarify this point (lines 279-282).

61). P. 11 L. 472-488: Consider providing information on the type of habitat at each site, considering different habitats will have different light environments.

We have identified the specific locations of each site in lines 433-440 so that readers could try to find information to assess the variation in habitat types at each site if desired. Unfortunately, we do not have measures of the specific habitat type or its variation for each collection locality at this time. However, upon request, we have specific GPS coordinates for the collection site of each individual frog that could potentially aid in habitat classification.

62). P. 12 L. 531-532: Please provide additional details about these qualitative descriptions.

Added. Lines 485-487.

63). P. 15 L. 676: I suggest adding a short explanation on what the tradeoff spectrum tells us.

Added. Lines 669-671.

64). Fig. 1: It is difficult to match the boxplots with their corresponding family. I suggest trying to improve the alignment between family labels and plots.

Added x-axes for each graph and right justified family names to improve this.

65). Fig. 2: This comment is more of an editorial nature (so I'll leave it up to the editor), but the legend for Fig. 2 seems excessively long. I'd suggest trying to cut it down substantially.

Figure caption has been shortened.

Please find our responses to each reviewer comment below. Responses are ***bolded and italicized***. We believe we have thoroughly and effectively addressed each concern proposed by reviewers.

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

I find this submission of high value and recommend it for publication. As I wrote in my previous review, I find this work exciting and important. The authors show biofluorescent signals in 39 genera (152 species) of frogs, representing 13 families, uncovering a visual signal that is relevant at a broad taxonomic scale. The exposure of a signal that has been virtually invisible to researchers make this an incredibly important finding. The authors addressed all concerns of the previous submission, added analyses that help with interpretation and clarified the testing and interpretation of the four criteria of Marshall and Johnson. The new tables help the reader through the immense dataset and will be a valuable resource for future work. This is a beautiful paper. I expect this is a launching point that will propel the fields of visual ecology, behavior, trait selection and speciation.

Thank you. We appreciate your feedback. Your comments greatly aided in improving this manuscript.

Reviewer #1 (Remarks on code availability):

The link was not available to view. That being said, this is not my area of expertise.

The Dryad link will become available to the public once the manuscript is published. It will include the script and data files previously submitted (with updates).

Reviewer #2 (Remarks to the Author):

I recommend accepting this revised manuscript for publication. The authors have improved its overall clarity and impact. Each of my original comments has been addressed, or appropriately rebutted, and the changes they have made to incorporate the feedback from other reviewers has improved the manuscript.

I have explored their R code. Most of it is reproducible with minor adjustments, except for the "RandTest_UpdatedFromRevisions.R" script. All script files are well annotated, which is very helpful. See comments associated with feedback on code.

Thank you. We appreciate your time in reviewing this manuscript. Your contributions greatly aided in improving this manuscript. We address the R code adjustments below.

Reviewer #2 (Remarks on code availability):

Overall the code is well written. However, there are some issues and questions the authors should consider. Also, the DOI or URL given in the reviewer portal (10.5061/dryad.djh9w0w5g) doesn't point to the author's code. Perhaps it is just an example DOI?

The Dryad link will become available to the public once the manuscript is published. It will include the script and data files previously submitted (with updates).

##########
"MainManuscript.R" script
##########

# The authors use the following code to read in datafiles:
dat <-read.table(get_GRfilenames_from_file[[i]], header=FALSE, , skip=14)
# Is the second comma after header necessary?

The second comma is not necessary and has ben removed.

Lines 630 and 637 write data tables to the working directory using the same title for the txt file. This isn't an error, per say, but is a little confusing.

The name for the data file in line 630 has been changed so as to not be a duplicate and to be more descriptive.

Lines 696 and 699 - unnecessary end parentheses? Additionally, another package is needed to run code for this figure (lines 679 - 698). Package Rmpfr required.

Added "library(Rmpfr)" at line 675 and removed unnecessary end parenthesis at line 699, parentheses at line 630 needed.

Lines 732 and 735 - unnecessary end parentheses?

Removed unnecessary end parenthesis at line 735, parentheses at line 735 needed.

Lines 775 and 778 - unnecessary end parentheses?

Removed unnecessary end parenthesis at line 778, parentheses at line 775 needed.

line 884 uses an object named "no" but I can't find the code creating that object in this R script.

Removed line 884, unnecessary.

Lines 916 to 932 - I think access to original data files is required to run this code? Totally fine, but worth mentioning it in the annotations.

True. Added annotation specifying this at line 898.

##########
Script "RandTest_UpdatedFromRevisions.R"
##########

text misspellings in the initial annotations

Misspellings in the initial annotations fixed.

Authors have included "Rod and Twilight Spectra.csv" files in the folder for reviewers, but not files for other spectra or rod sensitives. I'm checking this script file using that csv file.

Changed to "Rod and Spectra.csv." This includes all additional spectra and photoreceptor sensitivities).

I tried checking this text, but I could not. I think perhaps there were objects created that were renamed, or that there are input files needed? (e.g., object testStat?).

Checked. Should work now with corrected file input above.

If this script is going to be published, I recommend restructuring it so that you can use more consistent object names at the start for your input data and variables. That way someone wouldn't have to find and replace object names with different datasets when using data from different emission peaks and light sources.

We left the object names as is to avoid confusion with the multiple vision and environment spectra; however, we have added script annotation to provide clarity in which input data and variables need to be changed to test each criterion (lines 4-11).

The authors should also consider what the goals are for publishing their script. If the goal is for someone else to run through multiple script files in order to reproduce analyses, then I recommend organizing scripts into an RProject. And naming those scripts logically "01_DataOrganization", "02_DataAnalysis", "03_ManuscriptFigures", etc. If the assumption is not that someone would need to move directly from one script to another, then I recommend that the authors include code that will clear the environment and set the seed. Below is what I typically do in my own scripts, but I'm sure there are more ways to do it.

```
# clear the environment & set the seed
# This makes sure that all pertinent files are specified as part of running this script, and that the analysis is repeatable
remove(list = ls())
set.seed(2583722) # helps set a random-number so that what you do later is more reproducible
```

We have added annotation to the RADME.md file specifying that the order of the scripts in the README.md file follows the order of analyses run and their presentation in the manuscript.

Reviewer #3 (Remarks to the Author):

The authors have addressed several of my and the other reviewers' comments satisfactorily, which has improved the manuscript. That being said, there are still issues with the revised manuscript that need to be addressed, and some of the changes since the previous version have actually added new issues. Specifically, the authors have narrowed down onto the subset of their results that fit the Marshall and Johnsen criteria, obtained from about half of the species and ⅓ of the measured individuals, and have not addressed the rest. In my view this has weakened the manuscript. Below I elaborate on these issues. Without addressing the first three I cannot recommend this paper for publication, the remaining are suggestions that I strongly believe would make the paper better, but are ultimately up to the authors.

Thank you for your continued revision. We appreciate your time in helping to improve this manuscript. We have now expanded our focus to include the orange fluorescent emissions and have added additional analyses that take phylogenetic structure of the data into account. We expand on these changes below, with specific attention to the first three issues for which, without addressing, you cannot recommend publication.

1. The statistical analyses are still problematic.

First, although the authors did use species as the unit of analysis in the second randomization test for criterion 2 and 3, other analyses and the text still use individual frogs as the unit of analysis. The issues with this approach were detailed in the previous round of reviews. The authors should use one measurement per species in *all* analyses (except for those comparing excitation light source wavelengths with irradiance curves, which don't involve frogs). Furthermore, when data are reported and interpreted in the text and figures (e.g. Figs. 1 and 6), this should be done in terms of between-species, and hopefully also within-species variation. The interpretation of any pattern will be different if the data come from 5 individuals of the same species versus one individual from each of 5 species.

Second, I disagree with the author's justification to not account for phylogenetic structure in their data. It is true that highly unbalanced sampling can lead to spurious results in some cases, but then again not accounting for phylogeny at all can do so too. In reality the majority of cross-species datasets have missing data and are limited to some degree by opportunistic sampling. This study's dataset spans multiple families across the amphibian tree, and has different levels of sampling across clades, so it is bound to have a high level of phylogenetic non-independence between measurements. The authors should, at the very least, account for phylogenetic structure in their statistical analyses, for instance using phylogenetic generalized least squares for their ANOVA-type analyses. Testing how emission wavelengths compare with irradiance and sensitivity curves in a phylogenetically-informed way will require some more thought. In the current approach, the null hypothesis is that all species (or individuals) are equally likely to reflect any wavelength, and results show that the observed emittances are instead clustered around specific regions of the wavelength spectrum. Whether this clustering is due solely to phylogenetic inertia is not addressed. A way to address could be, for instance, testing how different models of continuous trait evolution fit the data.

Further phylogenetic comparative analyses to address the evolution of biofluorescence would definitely strengthen the paper, but saving these for future publications is fine too.

We have addressed both of these comments combined in the addition of a new set of analyses. We have run a set of analyses with constrained optima under Ornstein-Uhlenbeck evolution, where the constrained optima correlate to the peak wavelength of fluorescence from the tradeoff spectrum of each environmental and vision condition. See the addition of the Assessing Phylogenetic Structure sections (lines 335-419, lines 818-882, and cited additional figures/tables). See lines 843-844

for specific note about within-species variation and sample sizes used in these analyses.

2. In this version the authors decided to focus mostly on the green fluorescence peak, minimizing the attention given to the orange peak. In doing so, they have restricted their interpretation of the data to 1/3 of the individuals tested, which, in my view, considerably weakens the impact of the paper. I believe orange fluorescence is a central feature of the data, which should be directly tackled when interpreting it, even if it doesn't fit the criteria being tested. This is especially relevant considering that, as far as I could tell, a large number of species are polymorphic in this regard: they contain both green-emitting and orange-emitting individuals. If green fluorescence has an important ecological function, why don't all individuals exhibit it? Intraspecific polymorphism is not predicted by most models of positive selection, and is more often associated with balancing selection. Considering the central axis of this paper is an adaptive explanation for fluorescence (i.e. ecological tuning), intraspecific polymorphism should definitely be addressed in this paper.

We agree and have now included further analysis and discussion of the orange fluorescent signal. See newly added section A Proposed Expansion of Marshall and Johnsen's Criteria and Assessment of the Anuran Orange Fluorescent Signal using a New Key for Ecological Significance (lines 421-495, and associated figures/tables). We also added a specific note addressing intraspecific polymorphism (lines 498-500).

3. The authors focus on highlighting the species that do fulfill the four criteria, but don't address the fact that nearly half (44%) of the species don't. If fluorescence appears ubiquitous across the tested species, but is only ecologically tuned in about half of them, how can its existence be explained in the remaining species? Perhaps the criteria are too stringent? These seem like key questions to address when evaluating the evidence for ecological tuning of biofluorescence in amphibians.

We have added this to the discussion and proposed an updated key for ecological significance of biofluorescence. We believe this key better identifies the complexity of the previously defined criteria. See added section A Proposed Expansion of Marshall and Johnsen's Criteria and Assessment of the Anuran Orange Fluorescent Signal using a New Key for Ecological Significance (lines 421-495, and associated figures/tables).

4. On a related note, Reviewer 1 pointed out the previous version of the paper had some "broad overstatements" (their words), which could be attenuated to improve the paper's impact. The authors' solution was emphasizing that their statements apply to the majority of tested species. Although 56% is technically the majority, it is also very close to one half. I recommend using more attenuated language, such as "about one half" to more accurately characterize the proportion of species that fulfill these criteria.

These have now been changed to more attenuated language (see lines 31, 158, and 529).

5. The authors justify not testing the perception of fluorescence by the red cones under twilight and moonlight conditions, since the light may be too low for cones to work under those conditions. This is a valid point, but I still think this test could be informative, considering that, as the authors point out in the response to reviewers, there is no information on the intensity of ambient (and therefore emitted) light experienced by frogs under twilight/moonlight conditions at the sampling sites, so it is possible that the emitted fluorescence could be above the sensitivity threshold for cones. As per Fig. S5 it seems like the red cone would perceive the orange fluorescence well (if it where bright enough), so this seems like a possibility worth exploring.

In our newly added analyses, we include an analysis that indirectly considers the perception of fluorescence by the red cones under dim light conditions. This is done in the first added environmental/vision condition when testing models of evolution for the blue-light induced orange fluorescence. This condition considers the wavelengths that maximize matching both the fluorescent chlorophyll b background environment (which would be excited most in dim Twilight conditions) and the Red-sensitive cone visual sensitivity curve. We then test the evolution of the frog orange fluorescent emission signal when this maximum wavelength is the constrained optimum emission value. We believe that additional specific analyses testing the perception of fluorescence by the red cones under twilight and moonlight conditions specifically are outside of the scope of this study when considering our previous justification that it is unreasonable to assume that the emitted fluorescent light should be bright enough for cones to perceive in twilight and moonlight conditions due to the physical properties of the process of biofluorescence. Fluorescence always reemits less light than is available in the environment (because of energy lost in the process). So, fluorescence will never be brighter than the available amount of light at twilight, and hence only frog rod photoreceptors will be activated, not cones, under twilight conditions.

Reviewer #3 (Remarks on code availability):

The repository cited is not accessible, but I did find the data and code on the zip file. I glanced over it and it seems OK, but I did not look at it in depth.

The Dryad link will become available to the public once the manuscript is published. It will include the script and data files previously submitted (with the updates addressed above).