



Peer Community In Paleontology

The origins of Lissamphibia

Robert Asher ? based on peer reviews by [Jennifer Olori](#) and 2 anonymous reviewers

Michel Laurin, Océane Lapauze, David Marjanović (2019) What do ossification sequences tell us about the origin of extant amphibians? Missing preprint_server, ver. Missing article_version, peer-reviewed and recommended by Peer Community in Paleontology. <https://doi.org/10.1101/352609>

Submitted: 22 June 2018, Recommended: 16 October 2019

Cite this recommendation as:

?, R. (2019) The origins of Lissamphibia. *Peer Community in Paleontology*, 100004. [10.24072/pci.paleo.100004](https://doi.org/10.24072/pci.paleo.100004)

Published: 16 October 2019

Copyright: This work is licensed under the Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Among living vertebrates, there is broad consensus that living tetrapods consist of amphibians and amniotes. Crown clade Lissamphibia contains frogs (Anura), salamanders (Urodela) and caecilians (Gymnophiona); Amniota contains Sauropsida (reptiles including birds) and Synapsida (mammals). Within Lissamphibia, most studies place frogs and salamanders in a clade together to the exclusion of caecilians (see Pyron & Wiens 2011). Among fossils, there are a number of amphibian and amphibian-like taxa generally placed in Temnospondyli and Lepospondyli. In contrast to the tree of living tetrapods, affinities of these fossils to some or all of the three extant lissamphibian groups have proven to be much harder to resolve. For example, temnospondyls might be stem tetrapods and lissamphibians a derived group of lepospondyls; alternatively, temnospondyls might be closer to the clade of frogs and salamanders, and lepospondyls to caecilians (compare Laurin et al. 2019: fig. 1d vs. 1f). Here, in order to assess which of these and other mutually exclusive topologies is optimal, Laurin et al. (2019) extract phylogenetic information from developmental sequences, in particular ossification. Several major differences in ossification are known to distinguish vertebrate clades. For example, due to their short intrauterine development and need to climb from the reproductive tract into the pouch, marsupial mammals famously accelerate ossification of their facial skeleton and forelimb; in contrast to placentals, newborn marsupials can climb, smell & suck before they have much in the way of lungs, kidneys, or hindlimbs (Smith 2001). Divergences among living and fossil amphibian groups are likely pre-Triassic (San Mauro 2010; Pyron 2011), much older than a Jurassic split between marsupials and placentals (Tarver et al. 2016), and the quality of the fossil record generally decreases with ever-older divergences. Nonetheless, there are a number of well-preserved examples of "amphibian"-grade tetrapods representing distinct ontogenetic stages (Schoch 2003, 2004; Schoch and Witzmann 2009; Olori 2013; Werneburg 2018; among others), all amenable to analysis of ossification sequences. Putting together a phylogenetic dataset based on ossification sequences is not trivial; sequences are not static features apparent on individual specimens. Rather, one needs multiple specimens representing discrete developmental stages for each taxon to be compared, meaning that sequences are usually available for only a few characters. Laurin et al. (2019) have nonetheless put together the most exhaustive

matrix of tetrapod sequences so far, with taxon coverage ranging from 62 genera for appendicular characters to 107 for one of their cranial datasets, each sampling between 4-8 characters (Laurin et al. 2019: table 1). The small number of characters means that simply applying an optimality criterion (such as parsimony) is unlikely to resolve most nodes; treespace is too flat to be able to offer optimal peaks up which a search algorithm might climb. However, Laurin et al. (2019) were able to test each of the main competing hypotheses, defined a priori as a branching topology, given their ossification sequence dataset and a likelihood optimality criterion. Their most consistent result comes from their cranial ossification sequences and supports their "LH", or lepospondyl hypothesis (Laurin et al. 2019: fig. 1d). That is, relative to extinct, "amphibian"-grade taxa, Lissamphibia is monophyletic and nested within lepospondyls. Compared to mammals and birds (including dinosaurs), crown amphibian branches of the Tree of Life are exceptionally old. Each lissamphibian clade likely had diverged during Permian times (Marjanovic & Laurin 2008) and the crown group itself may even date to the Carboniferous (Pyron 2011). In contrast to mammoths and moas, no ancient DNA or collagen sequences are going to be available from >300 million-year-old fossils like the lepospondyl *Hyloplezion* (Olori 2013), although recently published methods for incorporating genomic signal from extant taxa (Beck & Baillie 2018; Asher et al. 2019) into studies of fossils could also be applied to these ancient divergences among amphibian-grade tetrapods. Ossification sequences represent another important, additional source of data with which to test the conclusion of Laurin et al. (2019) that monophyletic Lissamphibians shared a common ancestor with lepospondyls, among other hypotheses.

References

Asher, R. J., Smith, M. R., Rankin, A., & Emry, R. J. (2019). Congruence, fossils and the evolutionary tree of rodents and lagomorphs. *Royal Society Open Science*, 6(7), 190387. doi: [10.1098/rsos.190387](<https://dx.doi.org/10.1098/rsos.190387>)

Beck, R. M. D., & Baillie, C. (2018). Improvements in the fossil record may largely resolve current conflicts between morphological and molecular estimates of mammal phylogeny. *Proceedings of the Royal Society B: Biological Sciences*, 285(1893), 20181632. doi: [10.1098/rspb.2018.1632](<https://dx.doi.org/10.1098/rspb.2018.1632>)

Laurin, M., Lapauze, O., & Marjanović, D. (2019). What do ossification sequences tell us about the origin of extant amphibians? *BioRxiv*, 352609, ver. 4 peer-reviewed by PCI Paleo. doi: [10.1101/352609](<https://dx.doi.org/10.1101/352609>)

Marjanović, D., & Laurin, M. (2008). Assessing confidence intervals for stratigraphic ranges of higher taxa: the case of Lissamphibia. *Acta Palaeontologica Polonica*, 53(3), 413–432. doi: [10.4202/app.2008.0305](<https://dx.doi.org/10.4202/app.2008.0305>)

Olori, J. C. (2013). Ontogenetic sequence reconstruction and sequence polymorphism in extinct taxa: an example using early tetrapods (Tetrapoda: Lepospondyli). *Paleobiology*, 39(3), 400–428. doi: [10.1666/12031](<https://dx.doi.org/10.1666/12031>)

Pyron, R. A. (2011). Divergence time estimation using fossils as terminal taxa and the origins of Lissamphibia. *Systematic Biology*, 60(4), 466–481. doi: [10.1093/sysbio/syr047](<https://dx.doi.org/10.1093/sysbio/syr047>)

Pyron, R. A., & Wiens, J. J. (2011). A large-scale phylogeny of Amphibia including over 2800 species, and a revised classification of extant frogs, salamanders, and caecilians. *Molecular Phylogenetics and Evolution*, 61(2), 543–583. doi: [10.1016/j.ympev.2011.06.012](<https://dx.doi.org/10.1016/j.ympev.2011.06.012>)

San Mauro, D. (2010). A multilocus timescale for the origin of extant amphibians. *Molecular Phylogenetics and Evolution*, 56(2), 554–561. doi: [10.1016/j.ympev.2010.04.019](<https://dx.doi.org/10.1016/j.ympev.2010.04.019>)

Schoch, R. R. (2003). Early larval ontogeny of the Permo-Carboniferous temnospondyl *Sclerocephalus*. *Palaeontology*, 46(5), 1055–1072. doi: [10.1111/1475-4983.00333](<https://dx.doi.org/10.1111/1475-4983.00333>)

Schoch, R. R. (2004). Skeleton formation in the Branchiosauridae: a case study in comparing ontogenetic trajectories. *Journal of Vertebrate Paleontology*, 24(2), 309–319. doi: [10.1671/1950](<https://dx.doi.org/10.1671/1950>)

Schoch, R. R., & Witzmann, F. (2009). Osteology and relationships of the temnospondyl genus *Sclerocephalus*. *Zoological Journal of the Linnean Society*, 157(1), 135–168. doi: [10.1111/j.1096-3642.2009.00535.x](<https://dx.doi.org/10.1111/j.1096-3642.2009.00535.x>)

Smith, K. K. (2001). Heterochrony revisited: the evolution of developmental sequences. *Biological Journal of the Linnean Society*, 73(2), 169–186. doi: [10.1111/j.1095-8312.2001.tb01355.x](<https://dx.doi.org/10.1111/j.1095-8312.2001.tb01355.x>)

Tarver, J. E., dos Reis, M., Mirarab, S., Moran, R. J., Parker, S., O'Reilly, J. E., & Pisani, D. (2016). The interrelationships of placental mammals and the limits of phylogenetic inference. *Genome*

Reviews

Evaluation round #2

DOI or URL of the preprint: <https://doi.org/10.1101/352609>

Version of the preprint: 1

Authors' reply, 09 October 2019

[Download author's reply](#)

[Download tracked changes file](#)

Decision by Robert Asher ?, posted 30 September 2019

What do ossification sequences tell us about the origin of extant amphibians?

I agree with the reviewer that this manuscript is just about ready for publication. I've made a number of minor comments for the authors to consider, below, so am obliged to tick “revision”. However these are all minor and the authors can incorporate as they see fit.

line 14: I'd slightly edit the first sentence. “Controversial” is a value judgement; I'd delete this term. The authors say as much in the end of this sentence regarding current lack of consensus. You might also add some text to make clear that the lack of consensus is about lissamphibian affinities among fossil groups, not (for example) that they are tetrapods or are the sister taxon to extant amniotes.

line 37: There's only one phylogeny (at least of vertebrate high level taxa) and it's neither “molecular” nor “paleontological”. Rather, we use these kinds of data (among others) to reconstruct what it is. So here & throughout reserve adjectives like “molecular” to describe data, not phylogenies (at least when you're talking about species trees and not gene trees). The authors already use this style nicely in (for example) lines 58, 82, 93, 97.

line 44: Note that Gill (1872 *Smithsonian Misc Collec*, p. xliii) shows a very modern-looking tree with well-corroborated clades like gnathosomtes, cyclostomes, bony fish, actinoptes, sarcoptes including lungfish & coelacanth, and tetrapods with “reptiles” originating out of “batrachians” (see also discussion in Asher & Müller 2012 chap 1, p.2 in *From Clone to Bone CUP*)

lines 111-112: This sentence is a bit long. I'd recommend “...extant amphibians. Recently, Danto et al. (2019) ...”

line 134: We may need some guidance from our PCI colleagues regarding “supplementary material”. At present, this is mentioned in the main text but without a URL or other precise description of exactly where this is (or will be) available. Reference to the URL should eventually be added either as an appendix or directly in the text whenever “supplementary data” or “supplement” is mentioned (e.g., line 270). Relatedly, ensure that the wording for these data is the same throughout, as opposed to writing “supplementary material” on line 134 and “supplement” on line 270.

lines 182-183: The comment about lungfish “seem mostly impossible to homologize” is ambiguous. Ideally the authors might add a bit more justification or background to this statement. I'd recommend adding at least some references to guide readers to previous efforts at recognizing cranial homologies in these groups.

Also please respond to the comment from R1 regarding line 133 in the first version: "133- this is incorrect. Firstly, Schoch 2006 used the actinopt Amia with fairly few homology problems. Secondly, some part of the development of Eusthenopteran were published (Cote, 2002; Schultze 1984), though admittedly little about cranial development. It would provide some data about postcranial though."

line 217: "see below" regarding missing data might also pertain to the text *above*

line 237: the sentence here can end with "...events in ontogeny" without the "would". The following sentence is confusing; it has too many clauses (between commas) and two occurrences of "because". Please rewrite.

line 241: "simple" regarding your formula is a value judgment and should be deleted.

lines 260-261: change to "...only with sequences standardized by position" (or otherwise simplify so that "sequence" isn't repeated). Also "data" as a plural should be modified by "few" rather than "little".

line 292-93: change "maximal" to "maximum". The fact that the analysis took so long is presumably because with few characters the treespace is relatively flat and the algorithm gets bogged down on many local optima. I've found the recent builds of PAUP (including 4.0a165 used here) are faster than 4.0b10 from a few years ago, but (unless you explicitly limit PAUP, e.g., time or iterations per replicate) it'll get stuck among the huge number of possible topologies in a flat-treespace with few characters. Default in TNT is faster; you can limit the tree buffer in advance via "hold" and (in my experience) it will more quickly escape local optima, only filling up the tree buffer upon branch swapping & after finding many local optima. There's no need to redo analyses as far as I can see but for future reference, the authors might consider using TNT: <http://phylo.wikidot.com/tntwiki> <http://phylobotanist.blogspot.com/2015/03/parsimony-analysis-in-tnt-using-command.html> <https://groups.google.com/forum/#!forum/tnt-tree-analysis-using-new-technology>

line 329: It's not quite clear to me why "branch lengths... set to the same length" is an expectation of punctuated equilibria. The latter is (despite occasional Gouldian hyperbole) is just an application of peripatric (= "allopatric" in Eldredge & Gould 1972) speciation to the fossil record & the consequent expectation that small populations will tend not to leave behind a fossil record. "Stasis is data", as the saying goes, & is essentially an indication of population size. Perhaps another sentence or two explaining how punk-eek leads to particular expectations regarding branch lengths would be helpful.

line 350: "mentioned" is more appropriate than "evoked"

line 359: change to "...drawbacks that led us not to use them".

line 361: I'd break this sentence into at least two, e.g., "...can be summarized briefly as follows:..." [new sentence]

line 368: "documented previously" in Germain & Laurin (2009)? Please add ref.

line 386-90: "established consensus" should come with a list of references behind the branching pattern and divergence estimates (ahh I see this is from line 396). Please also add these to the fig. 2 caption. Also "molecular divergence dates" are themselves contingent upon paleontological calibrations, and hopefully you've picked estimates that do not recycle other clock dates as calibrations themselves (see Graur & Martin 2004, Reisz & Muller 2004, both in Trends Genetics).

line 415: as you've assembled a (very impressive) ossification dataset across osteichthyans, data from birds and mammals are relevant. However stating here with "for the birds" is a bit sudden, and you might add a sentence here to remind readers the importance of amniote data for your study of lissamphibian origins. Also I'm not sure about the topologies in Pons et al., Wang et al & Gonzales et al., but the Prum et al. 2015 (very large) dataset shows topological conflict with other, large genomic studies, in particular Jarvis et al. 2014 reflecting what remains a stubborn polytomy at the base of Neoaves (nicely reanalyzed & discussed in Reddy et al. 2017 Syst Bio). How might the competing phylogenies in (say) Prum et al. 2015 vs Jarvis et al. (2014) influence your interpretations of amniote ossification sequences?

line 419: A good summary & rationale for mammalian divergences (and why some estimates may be too old) is Phillips & Fruciano 2018 BMC Ev Biol, also Dos Reis et al. 2012 (Proc Roy Soc B). A good compendium of vertebrate divergence dates in Benton et al. 2015 Paleont Electr.

line 477 (and elsewhere): paragraphs like this that have frequent references to acronyms (DH, LH, PH2, etc.)

are hard to follow. It's fine to shorten the text w/ such acronyms but perhaps you could add parentheticals to remind your readers that "PH2" etc. are shown in your fig. 1a, b, c, etc.

line 516: This sentence would be easier to follow if you broke it up, e.g., rephrase text from "but it is weaker..." as a new sentence. Ditto for the long sentence in lines 518-523.

The Fig. 1 caption is very long. Perhaps move a few qualitative phrases (e.g., "very cautiously Froebisch et al. 2017...") to the main text. The detailed attributions of which authors are associated with which trees are important, but again could probably be moved to the main text while still making reference to those details with a single phrase in the caption, e.g., "See Methods for details on support for these competing topologies".

Fig. 2 caption is too short & makes no mention of the data behind this topology or divergence estimates. Please provide citations to make the caption self-contained and enable your readers to know the data & publications behind this tree. Please state what the horizontal colored lines represent (I guess marine stages?). Also they're garish and make the branches harder to read compared to (for example) grayscale, dotted lines, or similar.

line 1041: Just write "are" rather than "appear to be" (also line 1049). Again this caption has interpretation & detail (e.g., "...there is clearly a phylogenetic signal...") that is more appropriate for Results or Discussion than a figure caption.

line 1044: As noted previously RE: suppdata, write out what "SM 1" means (also line 1050) and add a statement (somewhere) indicating where this can be downloaded.

Appendix 1: without lines to indicate columns & rows, and without a repeating header at the top of each page, this appendix is hard to read. I would recommend moving this to online supplementary data in the form of a spreadsheet. Alternatively, reduce the font size and make this a table or figure with repeating a header and lines to delineate cells.

Reviewed by anonymous reviewer 2, 29 September 2019

The authors have greatly improved this manuscript. The authors approach the results of this analysis with much more caution and acknowledgement of limitations. As such, the data and interpretations appear appropriate.

Thank you to the authors for taking my comments and questions into consideration and clarifying issues found in the first draft of this manuscript.

Evaluation round #1

DOI or URL of the preprint: <https://doi.org/10.1101/352609>

Authors' reply, 25 July 2019

Please see attached detailed reply. Here, I only wish to add that for the clean, complete paper with the figures, please see the posted updated pdf file on BiorXiv. I have attached a MS Word file with tracked changes (lots of changes!), but without figures, which are difficult to integrate into Word files.

[Download author's reply](#)

[Download tracked changes file](#)

Decision by Robert Asher ?, posted 22 August 2018

What do ossification sequences tell us about the origin of extant amphibians?

Recommender comments on Laurin et al. PCI-Paleontology by Robert Asher

22 Aug 2018 I've just got a third review in today and will make that available to the authors. (I don't think it shows up yet on the PCI-Paleo site). My comments below were written after having just the first two reviews;

the third doesn't change my decision to "recommend revision", but does provide further constructive critiques that the authors should consider. My editorial comments are pasted below.

21 Aug 2018

Overall I like this manuscript and am keen to see it as a formally accepted paper in *PCI-Paleontology*. Both reviewers raise a number of issues which need to be addressed. R2 in particular argues that taxon and character sampling is not quite sufficient to reject all hypotheses besides monophyletic origin among lepospondyls, or at least not as strongly as the authors do in this manuscript. I welcome a revision taking these critiques into account, and if possible increasing the scope of taxa and/or characters sampled as per reviewer critiques. Additional, minor comments of my own are pasted below. Please respond to all of these and the reviewer comments in your revision.

In the caption for fig. 1 write out the "LH" abbreviation (and wherever possible minimize acronyms in the text)

line 103: depending on your response to the reviewers, and given that you're looking at cranial sequences, it would be more informative here to note "...extensive database on cranial ossification sequences..."

line 111: state what the software was.

line 122: Characters missing for a given fossil, but present in other taxa, can have an impact on phylogenetic estimation for that fossil by one or both of the following interrelated effects: - changing placement of taxa to which the fossil may be related - changing number of steps (in a parsimony context) on a given tree of other characters that are known for that fossil. Given the above (detailed in Asher et al. 2005 *JVP* 25(4):911-923), are you sure that characters "could not be scored for the temnospondyls *Apateon* and *Sclerocephalus*, so they could not have helped resolve the main question examined in this study"?

line 144: I don't think many readers will remember these acronyms, but I recognize the benefit of not having to write each one out at every occurrence. Perhaps add more frequent references to your fig. 1, for example here, and remind readers that acronyms are defined & figured in your fig. 1.

line 168: It would be straightforward to apply an optimality criterion to these sequence data and actually test if they are indeed "unlikely to provide a well-resolved tree". You wouldn't need to figure anything or write at length, but note simply that—assuming you're correct— method X (e.g., parsimov or others you prefer) "...yields an unresolved tree so instead we tested likelihoods of the competing hypotheses in Fig. 1 ..."

The parenthetical on lines 186-187 sounds a bit too informal & personal and I'd recommend deleting it.

line 203: replace "consensual" with something like "consensus" with relevant citations of the papers/phylogenies behind this consensus.

line 209: I think the term "databases" is more familiar written as one word.

line 234: Fabre et al. 2012 (*BMC Evol Biol*) present a well-sampled rodent phylogeny that includes both *P.melanophrys* and *M.auratus*; ensure that Wilson & Reeder 2005— a more taxonomic than phylogenetic reference— are consistent with their estimate.

line 243: this may seem like a trivial point but it's quite important: unless you're interested in gene trees, phylogenies aren't "molecular" or "morphological" but rather the data used to reconstruct them can entail one or both. So write "molecular phylogenetic analysis" or "... most recent phylogenies based on genomic/molecular data" rather than "molecular phylogenies" (as you've done elsewhere in the manuscript, e.g., lines 15, 50, 82 ...)

line 245: can you better justify your disagreement with Irisarri et al. 2017 (or cite someone to this effect)?

line 261: you might add an "and" before "in case", and/or better explain the connection between the "continuous evolutionary model" and why equal branch lengths are important to minimize bias against a particular hypothesis.

line 266: remind readers why inclusion/exclusion of *Sclerocephalus* & the squamosal is relevant here.

line 291: "previous attempts" sounds pejorative, and agree or not, Anderson's conclusions are not simply "attempts". So delete "attempts" and just write "previous phylogenetic conclusions from ossification sequences..."

line 299: by "untenable" you mean broadly regarded as false, as in amniote or mammalian non-monophyly,

right? Please clarify.

line 356: I'd start a new sentence after "LH" and delete "especially" & start the new sentence with "Similarities ..."

Reviewed by anonymous reviewer 2, 03 September 2018

The conceptual basis of this manuscript is indeed very interesting, especially in light of several studies that concluded ossification sequences don't appear to contain phylogenetic signal. It remained possible that ossification sequences could in fact contain such signal, but the taxonomic level of this signal has yet to be fully explored. I'd first like to congratulate the authors on compiling such an exhaustive list of extant ossification sequence data sources. This appendix alone will be a useful tool for many future research projects.

I have several questions and found areas of ambiguity that in their current state render this manuscript unready for publication.

My main concerns are:

- . 1) Clarity of methods
- . 2) Assumptions of the models and tests being deployed (ie., continuous characters, using branch lengths in a composite reference tree)
- . 3) The strength of conclusions based on largely inference alone

For the first, my recommendation would be to more clearly describe the data and methods. I appreciate the text is concise, but some questions remain and the readership and utilization of the approach would be increased if methods could be explained a little more in depth for non-experts to be able to deploy them in their own work, and to fully understand the present work. Based on the short explanations of the methodologies, I would find myself unable to be able to repeat the work – the key attribute of reproducible science. However, if these issues can be explained and justified in the text this would make an interesting and useful contribution.

Line 111, I am uncertain what 216 characters this original matrix is derived from. All sequence data? Or are these characters from a previous phylogenetic analysis that includes non-sequence data characters? After the missing data criterion was established, 7 characters remained and these are listed as bone names in the text. What are the actual characters? Their position within the sequence relative to one another? Please clarify what exactly these characters are and amend the text to explain this in the applied order.

Line 136, the absence of lepospondyls (and that only 3 fossil taxa in general are included?) is alarming. The obvious question is, **how can a relationship between lissamphibians and lepospondyls be supported by ossification sequence data if no ossification sequence data is available for lepospondyls?** More on this below.

Also, why not try using *Phlegethontia* in a stem tetrapod position? It seems it's position down there is pretty well accepted. Might serve in lieu of a 'fish' basal taxon?

Line 153, just to be clear, these are the position of ossification events in the series of 7 bones, correct? Could an example using the current data be provided?

Line 157, I feel the philosophy of the reasoning as to why skull length was not used to standardize the data is not sound. Just because results are less clear doesn't mean the method isn't working. What seems most likely is that the vast size differences of the organisms at comparable developmental stages would cause problems. Perhaps exploring this justification would make readers feel less like this was being discarded as an option simply because it didn't give a clear answer.

Line 160, These are the seven characters, correct? Perhaps restate that these are the seven characters that can be found in SM2 (with the definitions there also?). It sounds a bit like these are other data from the seven characters mentioned previously, and I am not sure which interpretation is correct.

Line 172, I wonder if these are truly continuous data. The methodology renders the data continuous-like values, but I feel they aren't actually continuous in the real world (they are discrete events). Does this factor violate the assumptions of the models being fitted to the data? Perhaps a little explanation can clarify this so I don't wonder if the tests are all invalidated by this interpretation.

Also in this section, I wonder about the treatment of branch lengths. Since the reference tree is a composite, the original branch lengths are no longer relevant in the composite tree. An analysis would need to be rerun with all the taxa to get those original branch lengths. So I believe any test involving original branch lengths from separate analysis whose trees were stitched together are invalid. Those characters were not given a chance to participate in the branch lengths of parts of the trees not included in the original analysis (e.g. mammal characters do not get to contribute to branch lengths in the amphibian part of the tree). Please explain if this issue is being taken into account somehow.

Line 254, I like this logic, however, it is not caveat free. That is also ok, but a detailed inspection of what is actually being tested, rather than what is the stated goal leaves me very cautiously accepting the conclusions. I will explore this now, and where my interpretations are incorrect, let this guide the author to clarifying the text to justify the conclusions.

It seems that the actual variable being used to determine the correct topology for lissamphibians is the position of *Apateon* (and *Sclerocephalus* in some analyses). This means what is actually being tested is how similar *Apateon*'s ossification sequence is to either salamanders, batrachians, or all lissamphibians with nothing known of variation among fossil taxa (surely there is enough variation among mammals such that sampling only 1 animal could yield drastically different results). In order to test what is described as being tested, a true phylogenetic signal in ossification sequence data needs to be demonstrated **and** *Apateon* needs to be demonstrated as representative of a temnospondyl, or at the very least an amphibamid, condition. Basically, when *Apateon* is better fit on the crown tetrapod stem, I can't help but think this may be due to species specific patterns of ossification or

even neoteny dependent patterns of ossification. Based on the exceedingly limited data from fossils at hand, there is no accommodation of variation. I understand the approach, however, much more caution in the results needs to be expressed in light of the data at hand.

Furthermore, the conclusions of a lepospondyl-lissamphibian link are ultimately entirely based on inference. Simply that if lissamphibians are placed between *Apateon* and amniotes and the models are best fit to that topology is interpreted as meaning that they share a similar ossification sequence to lepospondyls. However, this is entirely not observable. This, in essence, is not testing the LH, since there are no direct data supporting allying them with that clade at all. The results simply show that lissamphibians do not have a sequence more similar to *Apateon* than they do to amniotes. That this is consistent with a LH is an inference alone, and the text should reflect that the results are consistent, nothing more.

I'd lastly just like to verify that the argumentation being presented is not circular. It seems that early on we aren't sure if ossification sequence data carries phylogenetic signal. An analysis was performed that searches for a best fit of the sequence data based on phylogenetic congruence. This inherently means an assumption of is applied to the analyses. Finally, it was concluded that this best fit means ossification sequences are phylogenetically informative. However, the best fit model is the one that was attempting to maximize the phylogenetic signal. Just clarify if this isn't the case.

I find for the main goal and all analyses presented, that some discussion should be made about the actual sequence data analyzed and what about it might be phylogenetically significant. From my experience with development, I find ossification sequences can be strongly influenced by function (e.g., the timing of usage of an element). As such, I don't expect there to be much phylogenetic signal, and as a result I am not surprised that *Apateon* is different from lissamphibians. I do not take that result to mean there is not a close relationship between *Apateon* and lissamphibians. Much of the discussion is spent on topics not related to the study at hand, and while interesting and useful in a broader context, take away from the main findings of the study.

[Download the review](#)

Reviewed by anonymous reviewer 1, 15 July 2018

This paper addresses a fundamental problem of vertebrate phylogeny, one that concerns the relation of a major extant group with extinct ones that were quite diverse and important in the Paleozoic. The authors

examine ontogenetic data in the form of ossification sequences of the skull, and integrate information in a large-scope analysis that is sound in the critical examination of the data and of the limits and power of the method. The alternative hypotheses are clearly laid out and the previous attempts to use the kind of data investigated are properly revised. The study benefits from published data on exceptional fossils and makes the best out of those data - yet sampling is limited to relative few species – that is the nature of the data. The authors took steps to account for biases that could be introduced by stratigraphic (time) provenance of fossils (e.g., lines 261-263). At the end it is tree lengths and few fossils what provide the tests, but given the importance of the subject and the critical and thorough approach used, I find much merit in this paper and its conclusions to advance discussion of temnospondyls / lepospondyls / lissamphibia relationships. I suggest the authors revise a couple of potentially relevant references that contain much data (see below) and clarify some points below. One issue that is left largely ignored is that of intraspecific variation, which can be significant in extant amphibians. The data used is in many cases a studied optimization or rather consensus of that variation, but this needs to be mentioned at least, and its potential effect on the study discussed. The paper concerns skull data, but valuable insights on published, postcranial data on ossification sequences are presented. Abstract, line 17, 'perhaps because the diversity of methods used hampers comparisons'. I would rephrase this, as this paper conducts specific analyses and points into a direction that does not make this clause here fitting – I would write 'integration of data' as opposed to 'comparison'. As formulated, there is a contradiction in the Introduction in terms of the use of molecular data – please rephrase. Line 105 – for mammals, the most comprehensive dataset of cranial ossification is that of Koyabu et al. 2014 Nature Communications (not Weisbecker 2011) Weisbecker and Mitgutsch 2010 presented a comprehensive summary and analysis of anuran cranial ossification patterns. Besides citing it for obvious reasons, please make sure you have considered all the data mentioned there. Weisbecker V, Mitgutsch C (2010) A large-scale survey of heterochrony in anuran cranial ossification patterns. *J Zool Syst Evol Res* 48, 332-347. Lines 245-246 – there is a statement on the disagreement of the authors with Irisarri et al. 2017 calibration dates, but no justification or discussion of why. Some explanation/argument, would be fitting here. This paper is otherwise cited in line 582 – I guess there for the topology but not for the divergence times, correct? Figure big phylogeny: given that the divergence dates of the placental clades is far from being universally accepted, based on empirical work of alternative groups, it would be good to add some statement to this effect.

Reviewed by Jennifer Olori, 14 August 2018

Overview: Major revisions necessary (or a more explicit re-focus regarding what is actually being tested)

This paper is an important review and analysis of the growing number of data sets about skeletal development in early tetrapods and their potential descendants. Although such individual data sets often are published today, comparing them to one another in a comprehensive fashion, and within a phylogenetic framework is cumbersome and rarely done with any breadth. The authors of this paper attempt to do just that while also answering lingering questions in vertebrate paleontology.

In general, the many data sets are brought together with care and thought – something difficult to do given all the different ways that ossification sequences can be put together and interpreted, and the different morphologies across tetrapods. The authors also make a neat statistical assessment of comparisons across taxa, although by using just a single method. That approach is fine, but the paper would be much strengthened by including other methods as well (PGi, other methods of ranking to standardize, etc.), and comparing results across different analyses. Not only would this provide another measure of "confidence" regarding the results, but it would allow the authors' work to be more easily compared to the work of others, who may have used different methods to assess the evolution of skeletal development (hardly anyone seems to use the same methods these days). It may also help future workers select particular methods, if the authors could provide some review and comparisons regarding the strengths and weaknesses of each, and whether results are repeatable across different methods. The authors, in fact, bring up the issue of all these different methods in their abstract, but then make the same mistake they lament, by using just one. An even larger issue, however,

is taxon sampling, discussed in detail below.

Major issues:

The authors compared ossification sequences for cranial elements only. To my knowledge, in most lepospondyls for which we have ossification sequence data, the skulls are already ossified in all preserved material. Occasionally data exists for one or two elements, but not for all seven scored by the authors. Perhaps this type of work would be better focused on postcranial ossification material, so that more lepospondyl taxa may be included? In fact, the lepospondyl taxon for which the most cranial development information exists, is an Aistopod, and that group in the last few years has been supported as a stem tetrapod rather than a lepospondyl (see work by Pardo, Anderson, etc.). That is a major concern for a study that turns up a result of lepospondyl ossification sequences best aligning with those of modern amphibians. Lepospondyl taxa must be included, and to do this, postcranial elements will need to be included. Indeed, it seems very inappropriate to test a topology without including the key taxa upon which it is based. What is really being compared is a situation in which amphibians and amniotes are widely separated from one another by *any extinct tetrapods*, rather than whether amphibians *specifically share a relationship with lepospondyls*, to the exclusion of amniotes (ie what is implied in the LH topology). As discussed below, actually including Lepospondyl taxa with data changes the whole pattern of character tracing, which affects ancestral reconstruction, number of evolutionary steps/events, etc. The answer may be completely different, and a different topology supported. Another more minor issue may be the proportion of extant vs. extinct taxa, wherein the "pull of the recent" may be dictating early tetrapod evolution in terms of pattern of character evolution. Why are we still using living taxa to explain the evolution of their ancestors? It should be the other way around.

Specifics, by line number:

55 – More recent work suggests that salamanders (and maybe caecilians) have lost a tympanic ear that would have been present ancestrally (Anderson et al. 2016). That renders the point here mostly irrelevant, and somewhat more supportive of temnospondyl origins.

84 – Substitute "among" for "between" because this refers to more than two hypotheses being compared.

108-109 – authors need to be more forth coming in the methods about the sources of data and taxa included. Most readers won't access the sup data, and given my reservations above, they need to be honest about which extinct taxa were used (especially among lepospondyls), and the proportion of early tetrapods and outgroups to extant tetrapods. Not including enough extinct taxa will cause a bias of the "pull of the recent", in which the simply more common conditions of the living groups will outweigh, or even mask, the ancestral conditions present in extinct taxa. That would mean very little could actually be said about the evolution of skeletal development, and invalidate the authors' results here.

111- It seems a little unreasonable to choose a method that cannot handle missing data, given that this study focuses on comparisons between fossils and living animals. Most fossil data are incomplete in some way, and this is particularly true for lepospondyls vs. temnospondyls (the latter have a much better fossil record, and more complete ossification sequences).

122- yes another big point in trying to do these comparisons is that some taxa are simply very different. Temnospondyls as a whole, but especially Apateon show early ossification of postcranial material and late ossification of cranial. That is extremely hard to compare with lepospondyls, which generally have a completely ossified skull before the postcranial ossifications. By leaving out either postcranial or cranial elements from the analysis (or, just many other cranial elements, as in this case), the results will be very biased; some taxa that are otherwise wholly different in their total ossification sequence, make look more similar when only a subset is analyzed. This should be done with much more caution, and much more warning to the readers. A lot of information in the methods is left out.

133- this is incorrect. Firstly, Schoch 2006 used the actinopt *Amia* with fairly few homology problems. Secondly, some part of the development of Eusthenopteran were published (Cote, 2002; Schultze 1984), though admittedly little about cranial development. It would provide some data about postcranial though.

138 – The authors themselves bring up one the major concerns noted above, and honestly state that

no lepospondyls were used. How can their results be valid? With no actual lepospondyls, and no non-tetrapod outgroups, it seems fairly impossible to test their hypothesis directly, let alone confidently place living amphibians with a taxon not even present in the study.

157 – size already was shown to not correlate well with developmental stage nor ossification sequence, although my own work suggested that because fossil data are missing so much, using size as an approximation for fossil cases, only, doesn't really change our results too much, given that they are so poorly resolved anyway.

169 – statistical tests are not my strong skill, so an additional reviewer may be helpful to assess the appropriateness of CoMET and AIC for this application. However, I would add that other authors have compared sequence data in a phylogenetic framework (PGI for example, by Harrison and Larsson), so why aren't those methods also used and compared to CoMET's output? It isn't even discussed why more recent methods are not used.

186- perhaps the paper was a bit rushed? Why not wait for the corresponding consultant to reply, before abandoning some of the models? The paper would be strengthened by just waiting a little to see if these can be done, and if they cannot, explaining why more thoroughly.

192- true, but this is primarily character mapping with a more refined and modelled approach. That is different from phylogenetic analysis. In the former case, the authors are mapping characters onto existing hypotheses for check for best fit (more in line with objectives anyway, given that the goal was to test those specific topologies). Doing a phylogenetic analysis would have a different goal: see if the signal from development data agrees or disagrees with topologies based on adult phenotypes. That is a different type of analysis with a different type of goal. It doesn't need to be included here if the explicit focus is testing existing hypotheses of relationships. However, the two approaches should not be conflated in the methods. They are not alternative approaches because they do not accomplish the same thing, as misconstrued earlier in the methods and repeated again here, though implied rather than stated outright.

203 – use a different phrase because "consensual relationship" in English means something of a romantic or sexual nature.

206 – this is a bit puzzling, because molecular divergence estimates often include fossil calibrations anyway. Those gaps cannot be completely avoided. Also what is the purpose of the time tree? It is not explained in the methods. If developmental sequences are being mapped onto existing typologies already, why introduce yet another tree, and do stratigraphic data really add anything to the analysis? This is unclear as presented currently. It seems a time tree is unnecessary, given that so few extinct taxa are included, and as the authors note, there is so much disagreement regarding molecular divergence times anyway. With ossification sequence being so limited, the time tree feels a little redundant/unable to be fully utilized.

238 – no mention is made regarding the horrid state of squamate relationships. Which topology is used, the one based on morphology or the one based on molecular data? Certainly most of the citations favor the molecular tree, but that is not stated, and the disagreement/issues are not mentioned. The disparity would probably affect divergence estimates for squamates.

245 – no reasons are provided for "disagreeing" with Irisarri's dates. Please elaborate so that the reader is informed and the choice may be assessed.

254-255 – this is not really true. Software will test any hypotheses given to it, with any data set of scored characters. However, the lack of lepospondyl taxa in the analysis means that the program is filling in missing data for the taxon, or if the taxon is just left off completely, the character evolution may not be correct, even if the remaining topology can computationally be assessed. In other words, adding in those missing taxa could change which pattern of character evolution is the best match, and thus which topology best explains the data.

263- it is unclear why branch lengths would all be made equal in the end, after all the methodology regarding the different evolutionary models that the authors implemented earlier in the methods section. Were those other models used and tested? Perhaps this just needs to be explained better.

277 – the LH topology minus the actual lepospondyls might be best supported when lepospondyls also are not included in the other topologies, but what happens where their ossification data are included?? As noted

above, that changes the whole pattern of character tracing, ancestral reconstruction, number of evolutionary steps/events, etc. The answer may be completely different. It seems very inappropriate to test a topology without including the key taxa upon which it is based. What is really being compared is a situation in which amphibians and amniotes are widely separated from one another by any extinct tetrapods, rather than whether amphibians specifically share a relationship with lepospondyls, to the exclusion of amniotes (ie what is implied in the LH topology).

314- the data are unpublished, but I did do this in my dissertation (Olori, 2011), which might be a good starting point, at least in terms of source material. I never published those results because of all of the concerns and problems regarding ossification sequences well discussed by the authors here.

352 – clever subtitle, but first we need to revisit whether lepospondyls are monophyletic (unfortunately this problem seems to keep recurring every few years). The following discussion is weird, given that no data actually exist for lepospondyl cranial development, other than the fact that it is very early relative to temnospondyls.

I am happy to review future versions if the authors plan to continue work on the study. I think with the major issues addressed the paper would be a nice contribution to the literature and a great jumping off point for future use of sequence data in phylogenetic studies, as the authors suggest. I definitely agree with their assessment of the potential for these types of data.

[Download the review](#)