We are glad the Editor and Reviewers found the paper to be well written and of general interest.

**Editor**

The question of how temperature during development affects the thermal physiology of reptiles is an interesting one, especially in times of climate change. Both reviewers agree on this point. However, they differ considerably in their recommendations. Reviewer 1 criticises the sample size in particular, and the authors should now be able to demonstrate that the size is sufficient for their conclusions, or they should include additional data. Since both reviewers acknowledge that the paper is very well written and deals with an interesting topic, I would recommend that the authors be given the opportunity to thoroughly revise their paper.

**Response**: While our sample sizes appear small, they are comparable to samples sizes used in over 50% (26 out of 40) of the studies included in our meta-analysis. Smaller sample sizes in this area of research are not uncommon given that critical thermal limits push animals close to their physiological tolerance limits – pushing them much higher results in death. As such, there are ethical constraints that need to be carefully considered, and it is unlikely we would obtain approval for larger numbers of lizards than the number used in our study. We recognise that this may be a limitation. Our limited sample size was a major reason why we also included a meta-analysis of existing studies given that this is known to improve power (XX). We believe it is re-assuring the meta-analytic results completely support our empirical findings and that of other larger-scale meta-analyses in this area (e.g., Pottier et al. 2022). We have nonetheless revised our manuscript to make explicit mention of these points. Please see responses to Reviewer 1 below.

**Reviewers' Comments to Author**:

**Referee: 1**

Comments to the Author(s)

This is a concisely written, 'hybrid' ms. Part 1 examines effects of developmental temperature on critical thermal maximum and thermal preferences of an Australian lizard. Part II is a meta-analysis of impacts of developmental temperature on reptiles. Both are of general interest, as the vast majority of "acclimation" (plasticity) studies have looked treatments involving adults, not of developmental temperatures. [Note: the main exceptions are studies examining temperature dependent sex determination and the temperature dependence of body size ("temperature size rule".] The authors conclude that developmental temperature effects have minor impacts on their study species in particular, and on reptiles in general, and thus they suggest that behavioral or evolutionary responses will more likely to enable responses to changing climates.

The basic experimental design in Pt. I is to take groups of eggs and to assign them to 1 of 4 treatments (2 temperatures, 2 yolk provisioning), then rear the individual in a common garden, and test CTmax and Tpref months later.

A different design would involve adding a post-hatching treatment. This would enable one to see whether lizards reared at low temperature perform better at low post-hatching temperature than do lizards reared at high temperature (and vice versa). This latter design informs whether the developmental response is adaptive (see Huey et al. '99). However, your treatment has the advantage in that the hatchling were able to thermoregulate as they wished and so is probably more likely reflects what would happen nature, where hatchlings can (often) thermoregulate.

**Response**: This is a fascinating experimental design. We agree that it would be a fantastic future study. However, as noted by Reviewer 1, we had a different question, and our design likely reflects what would happen in nature.

As will be evident below, I have some concerns. I was unclear whether you actually measured body temperature, which is traditional in Tpref and CTmax studies. For Tpref, you measured skin temperature but didn't show whether that matched cloacal temperature. For CTmax, I'm not sure what you measured (see below).

In any case, what matters in terms of the questions posed here is whether developmental temperatures (and egg supplies) impact "indices" of CTmax and Tpref. They do not.

One weakness is power. Each of the 4 treatments had only 10 hatchlings, which came from different dams and unknown sex. Dam and sex effects will inflate the variance, and samples of 10 limited power. The meta-analysis sample sizes are (I suspect) too few given the number of categories. I like the questions being asked, but I'm skeptical that you have sufficient power to address these questions.

**Response**: While our sample sizes appear small, they are comparable to samples sizes used in over 50% (26 out of 40) of the studies included in our meta-analysis. Smaller sample sizes in this area of research are not uncommon given that critical thermal limits push animals close to their physiological tolerance limits. As such, there are ethical constraints that need to be carefully considered, and it is unlikely we would obtain approval for larger numbers of lizards than the number used in our study. We recognise that this may be a limitation. Our sample size was a major reason why we also included a meta-analysis of existing studies given that this is known to improve power substantially (XX). We believe it is re-assuring the meta-analytic results completely support our empirical findings and that of other larger-scale meta-analyses in this area (e.g., Pottier et al. 2022). We have nonetheless revised our discussion noting the sample size limitations of our empirical data.

that said, asking the question is important, as it might encourage more studies on developmental temperature effects. Towards the end of this review, I make several suggesting for you to consider -- I think highlighting these issues will help others to advance these questions.

**Response**: Thank you! We agree that these are important questions, and we hope that it will provide some guidance for future studies.

Miscellaneous suggestions:

line 49 cite Sinervo's egg manipulations here

**Response**: Thanks. We have now cited Sinervo here. See line 49.

line 59 probably true for animals, but not for plants (where most research is focused on moisture, nutrients, light)

**Response**: Thanks. We have now reworded to fit reviewer’s point. See Line 58

line 78 how many females?

**Response**:

line 80 Haphazardly not randomly

**Response**: Thanks. We have reworded. See line 82.

So you had 10 eggs per treatment, correct? a better design would have been to take 4 eggs from each female and assign them to the 4 treatments -- even better would be 2 eggs to each treatment. By not including a "dam" effect in your design, you add variance associated with dams, which reduces power to detect treatment effects.

**Response**: Sorry, this was not clear, we did in fact use a design suggested by Reviewer 1. Eggs from the same clutch (i.e., dam) were randomised across the four treatments when possible (*L. delicata* lays an average of 4 eggs but can range from 3-8 eggs). In other words, we have a ‘split clutch’ design and ‘clutch’ was included as a random effect in our models. We have now made this clearer in our methods.

line 81 while it is common for researchers to use fixed temperature treatments (I've done so myself), it is far from ideal. Soil temperatures show daily cycles of temperature and often change during development. More importantly, why did you pic 23 ° and 28 °C temperatures? Has anyone monitored egg temperatures of Lampropholis in the field? As is, a reader has no a priori way of knowing whether these temperatures are at all ecologically relevant for these lizards, or perhaps are stressful.

**Response**: This is an excellent point. Thanks for raising it. We agree that fluctuating treatments are more realistic, but most studies use constant incubation treatments (See Noble et al. 2018). Interestingly, meta-analyses suggest that fluctuating temperature treatments don’t often result in major differences to constant temperature incubation regimes (See Raynal et al. 2021).

Regardless, our temperature treatments were not constant. We are sorry that we did not make this clear enough in our manuscript. Temperatures fluctuated +/- 3 degrees Celsius around 23 and 28 C. We have now added this in our revised manuscript (See lines XX).

With respect to our choice of temperatures. Egg temperatures have been monitored in *L delicata* nests naturally (see Cheetham et al. 2011). Our choice of incubation temperatures are at the extremes of the natural temperatures. We have now cited this paper and made this point clearer in our revised manuscript, see lines XX.

If you used a large number of developmental temperatures, the extreme would be pathological, and some temperature range would be optimal for physiology. But we don't know whether 23 and 28 are within that range, or whether 28 °C is in fact stressful.

**Response**: Please see our response above. These temperatures are at the extremes of naturally occurring nest temperatures. We don’t believe our treatments were pathological as we did have high hatching success (~90%) across our treatments.

line 85 Were groups from mixed treatments, or single treatments? If the latter, then you have a "group" effect (i.e., the treatment and post treatment environments are confounded.

Also, group housing can lead to social dominance effects.

Regal, P. J. (1971). Long term studies with operant condition techniques of temperature regulation studies in reptiles. Journal de Physiologie, 63, 403-406.

**Response**:

line 89 Was the order of measurement independent of treatment group? I hope so.

**Response**:

line 104 I suspect some reviewers will object to this wide 'phylogenetic' grouping. Also, were only 15 species represented? If so, you have very few species per taxon, suggesting limited phylogenetic power.

**Response**:

line 114 define ARR

**Response**: The definition has now been provided. See line 119-121.

line 115 what is "study"? does that mean some papers reported on more than 1 species?

**Response**: This is correct, some studies had multiple species or studies often had more than two temperature treatments. We have now clarified this on lines 124-126.

line 126 Good to have checked for publication bias. (syntax bad here -- publication bias wasn't using a funnel plot.)

Note: I don't do meta-analyses so I can't comment on the methods, but the text conveys the impression that the authors were careful here.

**Response**: Thank you! Yes, we have tried to conduct the analysis as carefully as possible.

line 134 So none of your hatchlings died? That's a good sign that rearing conditions were tolerable.

**Response**: Hatching mortality was very low.

line 136 Personally, I would like some mention of methods of CTmax and Tpref in the text. There's big debate on CTmax methods, and a reader should not have to find the Supplement to learn the basics of your methods. With Tpref, you probably had multiple measurements/individual, but did you analyze means/medians for each individual or did you pool all measurements?

**Response**: We agree with this point the reviewer makes, and we have provided more details in the collection methods and our definition for each thermal trait in the methods section of the actual manuscript. The section now reads as follows:

*“Briefly, after undergoing a 24-hour fasting period, animals were transferred into individual lanes of a thermal gradient (5◦C to 55◦C) to measure Tpref. A FLIR T640 thermal camera was used to take thermal images of all lanes every 15-minutes over an eight-hour observation period. Tpref was defined as the mean skin surface temperature (on the neck) over the eight-hour observation period. Given the small size of lizards (i.e., XX snout-vent length and 1.2 grams) we assumed skin surface temperature reflected body temperature, which has been shown for many small lizards (Garrick 2008). For CTmax we followed the same fasting period used for Tpref experiments. Here, lizards were placed in falcon tubes in a water bath for 5 min at a temperature of 30◦C. The water temperature was increased to 38◦ C at a rate of 1◦ C/min. We used a control falcon tub with a thermal couple attached to the bottom of the tub where lizards were positioned to recorded the temperature of the tube surface which we took to be the temperature experienced by the lizards. CTmax was defined as the temperature at which an individual lost their righting reflex (for further details in collection methods, see Supp.).”*

Fig. 1 suggests no obvious treatment effect, and the statistics support that. But you have very limited power (only 10/treatment). Incidentally, why didn't you include sex in your statistical model? Can you sex hatchlings?

**Response**: We did include sex in our model. Please see Table 1 and line 107 and supplement. Sex was also not significant.

line 147

Fig. 2 what does "overall" mean"? Fig. 2D shows that sample sizes are small-- 2 turtles, 2 snakes, 10 lizards. Fig. 2C shows that you have 1 species of tropical something. I realize that that is what is available, but my feeling is that any conclusions from this meta-analysis are premature. You just have no good sampling (not your fault) and no power. The utility of a meta-analysis here is to highlight the paucity of information

**Response**: Thanks for catching that. “Overall” has now been clarified in figure 2 caption.

line 153 What about behavior? In any case genetic adaptation and plasticity need not be competing.

**Response**:

line 155 On "anticipatory" see Padilla and Adolph

**Response**:

line 162 I think you are overstating the pattern here, as studies with Drosophila (papers by Gilchrist, Crill, Huey, and others) have shown effects of developmental temperature on heat tolerance and other traits. The MacLean study used flies from stock centers, which have been maintained (often for decades) at fixed temperatures. I'm skeptical of their results, though obviously they are note.

**Response**: Thanks for pointing this out. We were unaware MacLean et al., flies were from stock centres. The *Drosophila* reference, MacLean et al., has now been removed from our statement. The other studies mentioned (i.e., Gunderson, Deutsch, and Pottier et al. ), one being a meta-analysis, explicitly show adults had limited plasticity. See line 175.

Seems to me that you need to add a paragraph discussing the limitations of your approach. First, you apparently know nothing about developmental environments of these eggs in nature, so your treatments may not be ecologically relevant. Basic natural history information is always needed (though almost always absent!). But you can encourage it. You hint at this in lines 169-170, but you don't concede ignorance of nest conditions for your species. Second, fixed temperatures in particular may yield misleading results -- there is a growing literature on the impacts of fluctuating temperatures. Third, power -- you have limited power in your experiments and in your meta-analyses. Call attention to this as it might encourage more studies of developmental effects.

**Response**: Natural history information of selected treatments for *L. delicata* has been provided. See comment above. See lines 86-89.

line 176 Janzen's classic paper ('67 Am Nat) should b cited here

**Response**: Great suggestion. This reference has been added. See line 189.

line 173 Interesting suggestion. I don't recall any studies that have examined this.

**Response**: Thanks! We’re glad Reviewer 1 found this interesting.

\*\*\*\*\*\*\*\*\*

Supplement

line 8 Tpref is NOT the optimal body temperature. See Martin, T. L., & Huey, R. B. (2008). Why suboptimal is optimal: Jensen's inequality and ectotherm thermal preferences. American Naturalist, 171, E102-E118.

**Response**: Corrected. We appreciate the literature suggestion. See line 8.

line 19 Check for consistency over the 8-hours. Brattstrom (decades ago) found (if I remember correctly) that lizards often take some time in a gradient before they "settle in".

You measured dorsal skin temperature, not body temperature. Did you ever check whether skin temperature = cloacal temperature? These are moderately large lizards (8 to 10 cm), so skin temperature may be different from cloacal.

**Response**: Actually, these are small lizards. Mean mass at measurement was only 1.3g (~4cm snout to vent length). While it is possible that skin temperature differed from body temperature lizards of this size are expected to equilibrate with environmental temperatures quickly (i.e. heat/cool rapidly given the environmental temperature; *See* Garrick 2008). We have revised our manuscript to be more explicit about this and provide better justification where we discuss the methods in more details. The revised section reads as follows:

“*A FLIR T640 thermal camera was used to take thermal images of all lanes every 15-minutes over an eight-hour observation period. Tpref was defined as the mean skin surface temperature (on the neck) over the eight-hour observation period. Given the small size of lizards (i.e., XX snout-vent length and 1.2 grams) we assumed skin surface temperature reflected body temperature, which has been shown for many small lizards (Garrick 2008).”*

line 27 are you sure they reached 30 °C? what is the time constant for a lizard of this size?

**Response**:

line 34 Suggests that you used water temperature as index of body temperature. But with a fast heating rate, Tb might lag, especially in relatively large individuals. You can easily test this -- assuming you have access to some lizards.

**Response**:

From Fig. S1, I'll guess the "operative" temperature inside the tube reflects water temperature and also air temperature (tube + air circulating from above the water).

Sorry I'm confused as the legend is ambiguous. Did you put a thermocouple probe inside the Falcon tube or inside a lizard in the Falcon tube? I've been assuming you were using Twater as your estimate of CTmax, but not I'm not sure what you measured. Sorry if I missed an explicit statement.

**Response**:

Even if you used Twater (or Tair inside the tube) as your metric, that should be ok for detecting developmental temperature effects, unless the size range of lizards was substantial. The temperature lag of a large lizard might be substantial.

but if you did not actually measure Tb, then I strongly encourage you not to call this CTmax, as you did not measure body temperature.

**Response**: We used Tair inside the tube as indicated above. The size range of the animals was minimal because they were all the same age (average mass = 1.3 g, SD = 0.27 g). Importantly, the mass of animals across the treatments did not differ (see Table below), so there should not be a body size effect on heating rate across the treatments. Our methods for measuring “CTmax” follow much of the literature on lizards (See XX). As such, we have opted to keep it as CTmax but we have been clear that not measuring body temperature

|  |  |  |
| --- | --- | --- |
| **Treatment** | **Mean** | **SD** |
| A-23 | 1.24 | 0.289 |
| A-28 | 1.34 | 0.363 |
| C-23 | 1.31 | 0.286 |
| C-28 | 1.27 | 0.193 |

Has anyone measured CTmax of Lampropholis the traditional way? Check Spellerberg's old papers, Greers, and John-Alder and Bennett. And check GlobTherm.

**Response**:

Inconsistent capitalization of titles in Literature cited.

**Response**:

Some relevant references:

Huey, R. B., D. Berrigan, G. W. Gilchrist, and J. C. Herron. 1999. Testing the adaptive significance of acclimation: a strong inference approach. American Zoologist 39:323-336.

Levins, R. (1968). Evolution in changing environments. Princeton University Press. # Levins laid out the conceptual literature on developmental effects.

Paranjpe, D. A., Bastiaans, E., Patten, A., Cooper, R. D., & Sinervo, B. (2013). Evidence of maternal effects on temperature preference in side-blotched lizards: implications for evolutionary response to climate change. Ecology and Evolution, 3, 1977-1991.

Padilla, D. K., & Adolph, S. C. (1996). Plastic inducible morphologies are not always adaptive: the importance of time delays in a stochastic environment. Evol. Ecol., 10, 105-117.

**Referee: 2**

Comments to the Author(s)

The authors used a two-pronged approach to investigate whether developmental conditions influence the thermal physiology of reptiles. First, using the delicate skink (Lampropholis delicata), they conducted a lab experiment to determine whether developmental temperature and maternal resource investment (via egg yolk reduction) influenced two thermal traits, CTmax and thermal preference. They found that it didn’t. They then did a meta-analysis more broadly for reptiles, to see whether developmental temperature influences thermal physiology. Their results suggest that there is limited developmental plasticity in these thermal traits in reptiles.

Overall, the study is of great interest to the readership of Biology Letters. It is extremely well-written, and analyses were outlined in detail and were appropriate for the questions being addressed. However, I do have one key question regarding the study….why focus on CT max, rather than CT min? CTmax has been show to be evolutionarily conservative in ectotherms, without much response to selection or plasticity. Conversely, CTmin is more evolutionarily labile, and the authors may have been more likely to observe a response in this trait. Perhaps it would be beneficial to explain the choice of CTmax for this study.

**Response**:

In addition, I have outlined a series of more minor comments that the authors should address in their revision.

Minor comments

1. Lines 71-72. Please provide the common name for Lampropholis delicata as well.

**Response**: This has been provided in the methods. Line 80.

2. Lines 78-80. It would be better to refer the reader to the supplementary methods here, rather than in Line 84.

**Response**: This has been done to fit the reviewer's suggestion. See line 84.

3. Lines 78-80. Is there a reference that you can provide that outlines the housing conditions that the females were kept in prior to oviposition? If not, it would be great to add this information to the supplementary methods. Likewise for the egg incubation details.

**Response**: Yes, a citation for husbandry and egg incubation conditions has been added. See line 89-90.

4. Lines 134-136. This is repetition from the methods.

**Response**: Thanks. This redundancy has been removed. Line 149.

5. Line 137. A Tpref of 20.99 C seems unusually low for the species. Was this a valid measurement, or was there some issue with this individual(s) during the experiment?

**Response**: We agree with Reviewer 2 that this outlier seems biologically low for the species and was in fact a point we checked prior to submission. We have thoroughly interrogated the raw data (i.e., pictures from thermal imaging camera) but we are convinced that this is correct. We also checked our data to see if this individual had any issues prior to measurement, but there was nothing noted. We also checked if individuals collected that measurement day demonstrated low Tpref’s. They did not. We also checked whether it was a common issue with that specific lane, but all other individuals performed fine. We are not sure why this specific individual has such a low Tpref, but this is the data!

6. Lines 153-156. This paragraph is repetition. It would be better to focus on your key results, and how they address these competing hypotheses.

**Response**:

Editorial office comments to authors:

Please ensure that you include;

\*A description of your dataset in the Data Accessibility section as described on our website: https://royalsocietypublishing.org/rsbl/for-authors#question4

**Response**:

\*Due to its lack of permanence, we do ask that datasets of soon to be accepted articles are uploaded to other repositories other than GitHub, such as Dryad or figshare. Please upload your data into another repository and amend your data section accordingly to reflect this.

**Response**: Thanks. We have loaded the data up to Zenodo.

\*Please also ensure that your data is cited in your reference list as described here https://royalsociety.org/journals/ethics-policies/data-sharing-mining/

**Response**:

\*As long as they are accurate, your Ethics, Competing Interests, Data, and Author Contributions sections should be removed from your main manuscript and added to ScholarOne in the relevant sections instead (Step 6 in ScholarOne)

**Response**: Thanks. Removed.

\*Please also confirm in your cover letter whether all the figures are your own or whether permission has been obtained for their use

\*If you have any images that can be used to promote your article on social media (should this be accepted) please upload them as a supplementary file

\*Please do ensure that your supplementary files themselves include the title and authors of your main manuscript, and that you also provide an appropriate title and description for each of these on ScholarOne

\*Please upload a version with the changes clearly marked as a supplementary file

\*Please confirm whether your manuscript should be available via open access, as we note that this may be a requirement of your funder. Please also check if your institution has a Read & Publish deal by visiting our website here: https://royalsociety.org/journals/authors/read-and-publish/read-publish-agreements/

\*Further tips on how to approach revisions that you may find useful can be found on our blog post here: https://royalsociety.org/blog/2021/07/responding-to-decision-letters-tips/