Dear Professor Goodacre,

We would like to thank the Editor and the reviewers for providing extremely constructive feedback on our manuscript entitled: “*Heritability and developmental plasticity of growth in an oviparous lizard*”.

We have now carefully considered all the comments and revised our main manuscript accordingly. Below, we provide a line-by-line response (in ‘blue’) to each of the comments raised by the Editor and three reviewers (in ‘black’). When helpful, we have pasted the section of our manuscript where we have made edits to provide clarity to what we have done to address comments. Additionally, we have included a document to aid in the review process that documents most of the changes we have made (in Track Changes).

We believe that our revised manuscript is significantly improved and hope that you now find it suitable for publication in *Heredity.*

Sincerely,

Fonti Kar and Daniel Noble

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Editor (Comments to the Author):**   
  
Dear Fonti Kar and Daniel Noble   
  
Peer review of your manuscript "Heritability and developmental plasticity of growth in an oviparous lizard" is now complete and we invite you to revise your work based on the reviewers’ reports   
  
The reviewers are unanimous in their positive assessment of the novelty and value of your work. Nonetheless, they highlight a number of concerns regarding the methodology and interpretation of results. Specifically, the reviewers are concerned regarding the change of housing during the experiment, which may confound the obtained results. Similarly, the reviewers comment on the low power to detect multiple confounding factors with the current experimental design.   
  
Please find the detailed comments from the reviewers attached below and we look forward to receiving a revised version of your manuscript.   
  
Best wishes,   
  
Bastiaan Star

**Response:** Thank you Professor Star. We have now significantly revised our manuscript in accordance with the suggestions provided by each of the reviewers. Please see the line-by-line response below.   
  
**Referee #1 (Remarks to the Author):**  
This manuscript focuses on understanding the effect of temperature on the growth curves of lizards. The authors tested if a temperature treatment changed the average growth curves, but also how it may affect the sources of variation among-individuals in these trajectories. They analyze their experiment in a quantitative genetics framework in the context of GXE interactions. They found that lizards raised in colder temperatures where on average larger, but found no support for a strong treatment effect on the average curves or the relative contribution to the variation in mass of the different variance components. However, they find that the different sources of variation change through development regardless of treatment. The questions that were addressed in this study are interesting and the statistical analyzes were thoroughly performed. However there are some formatting issues that, at points, made it difficult to understand the results and need to be addressed. I also have some general and some specific comments that I hope the authors find useful.

**General comments**

What are the implications of differences between treatments in the growth trajectory to understand adaptation to extreme temperatures. For instance is the optimal growth trajectory dependent on temperature? Why is changing variance across the developmental trajectory important for adaptation? This may only be relevant if selection is age specific? The link between the experiment and the adaptive potential will be more clear if the expected relationship between the mass growth trajectory and fitness is explained more clearly.

**Response:** Good point. Mass is important for survival in lizards (e.g., Sorci and Clobert, 1999; Warner & Andrews, 2002), but we have little insight into how patterns of selection on body size change through age in most species – this one included.

*References*:

Sorci, G. and Clobert, J. (1999) Natural selection on hatchling body size and mass in two environments in the common lizard (*Lacerta vivipara*). Evolutionary Ecology Research, 1: 303-316.

Warner, D.A. and Andrews, R.M. (2002) Laboratory and field experiments identify sources of variation in phenotypes and survival of hatchling lizards. Biological Journal of the Linnean Society, 76: 105-124.

**Detailed comments**

**Abstract**

Line 23-24. Only at the beginning or through the whole development?

**Response:** Thanks. This wasn’t clear. We have now revised as follows:

*“Our results show that lizards reared in cold developmental temperatures had consistently higher mass across development compared to lizards that were reared in hot developmental temperatures.”*

Line 26-27. But in the results is stated that there are no ”significant” differences.

**Response:** Yes. We have indicated here that these are not statistically significant.

**Introduction**

Line 49: Chevin et al. (2010) is a relevant reference here.

**Response:** Thanks, we agree. We have now added Chevin et al. (2010).

Line 83: ”ratio nature” is confusing

**Response:** Agree. We have removed this and simplified the sentence. It now reads:

“*Importantly, environmental comparisons of heritability have been criticised because they mask changes in the relative contributions of non-genetic and genetic variance*”

Line 84. Hansen et al. (2011) is relevant here.

**Response:** We have now added Hansen et al. (2011).

Line 90. Why the non-genetic sources? While heritability may be dependent on the environ- mental variance, the expected evolutionary changes is determined by the genetic variance. Maybe develop this argument better?

**Response:** Good point. We have re-worded this sentence. It now reads as follows:

*“The expression of genetic variation under different developmental environments can thus influence the evolutionary potential of fitness related traits.*”

Line 111. Is there any evidence that selection during development is age dependent, why is it important for adaptation that the genetic variance changes through development?

**Response:** Unfortunately, we do not have data on age-dependent selection in this species. It would be important to understand how genetic variation changes through development, and how this is affected by early environment because it informs on evolutionary potential at specific ages if selection were to operate – higher genetic variation would imply that, if selection were to operate, it would be more likely to lead to an evolutionary response. We have now made this clearer.

*“As such, the various sources that influence body size variation (genetic, environmental, maternal) are predicted to vary across ontogeny and temporal approach is needed in order to evaluate age-specific evolutionary potential of body size – higher genetic variation at a given age would imply that, if selection were to operate, it would be more likely to lead to an evolutionary response.”*

Line 127. Is the prediction related to the release of cryptic variance, if so, maybe refer to it here.

**Response:** Yes. We have now made this clearer and added a reference.

“*In addition, we predicted greater amount of genetic variance under higher developmental temperatures, after controlling for non-genetic sources of variance, as higher temperatures may release ‘cryptic genetic variation’ (Rowiński and Rogell 2017)*”

**Methods**

It was not clear the source of the parents. Where they born in the wild? Where the different source populations very different?

**Response:** Agree. Sorry, this was unclear. We collected all adults (i.e., parental population) from the wild. These wild caught adults were then used in our breeding design. We think this was unclear because the start of the sentence implied that we set up a colony across 2015-2017. We have now re-worded. While we collected from a few different sites within Sydney to ensure that local sites were not overly sampled (in accordance with our permits), sites across Sydney are uniform in their genetic structure (i.e., there is high gene-flow) and they form a single lineage (Chapple *et al.* 2013. Diversity and Distributions, 19: 134-146). We have made this clearer now:

“*We established a breeding colony of adult L. delicata (nfemales = 144, nmales = 50) using wild individuals collected across five sites throughout the Sydney region between August and September 2015. While we collected from five different sites in Sydney, biogeographic data suggests high gene-flow across the Sydney region (Chapple, Miller et al. 2013).*”

Line 134. Why did you chose this breeding desing. Is there are reason for only generating paternal half-sibs and not maternal half-sibs?

**Response:** Great question. We choose this design because creating paternal half-sibs is easier than maternal half-sibs given that females in our colony only really produced a single clutch in a year. We have now stated this more clearly:

“*We choose a paternal half-sib design because maternal half-sibs are difficult to generate given that females in our colony only produced a single clutch in a year (see below)*.”

Line 154-155: Were all reproducing individuals caught in the wild? Or were some dams and sires in the 2017 trials born in the lab in 2016? This could potentially influence maternal effects. What was the proportion of individuals breeding in two breeding seasons?

**Response:** All were caught from the wild. We have revised our methods to make this more clear.

Line 197: 94% of females had been sired by a single male? Do you mean offspring instead of females?

**Response:** Yes! Thanks for catching this mistake. We have now corrected ‘females’ to ‘offspring’.

Line 196-199: How can you have sperm retention if there is only one male in the box?

**Response:** It is possible because females can store sperm from previous breeding attempts. This is common in lizards. In other words, even though a single male is present, some eggs may have been fertilised from sperm stored from previous copulations in the wild (for example).

Line 231. Non-informative for what, priors on the random effects are rarely non-informative for all the variance, covariance and correlation of a model. Specific?

**Response:** We agree. We have now removed this statement.

Line 233. If posteriors are very skewed, the median provide a better central tendency measure.

**Response:** Agree, but the mean and median in most cases were very similar. We therefore report on the mean.

Line 243: ”Overall. Heritability” has a misplaced period

**Response:** Thank you! Fixed.

Line 251. For people not familiar with WAIC, can you refer to the “standard” difference threshold.

**Response:** A great point. Thanks for the suggestion. We have switched to LOO (Leave-One-Out cross-validation approach) for model comparisons as it’s more robust than wAIC – although the decision on what model is best supported in our case is not greatly impacted either way and in the supplement we present both.

STAN allows for uncertainty to be propagated when calculating LOO (or more specifically the predictive density differences (elpd\_diff) between models), which is why an SE can be calculated. As a rule, elpd\_diff values, calculated by comparing predictive power of two models of less than 4 mean that models are comparable. If the elpd\_diff is larger than 4 then determining whether one model is better than the other should be compared to the standard error of elpd\_diff (<https://mc-stan.org/loo/articles/online-only/faq.html#se_diff> and https://mc-stan.org/loo/articles/online-only/faq.html#ref-Sivula+etal:2020:loo\_uncertainty) (see also Sivula et al. 2020). If the SE is much smaller than the difference, then the model with the smallest elpd\_diff is preferred.

We have now made this clearer in our revised MS, adding the following section:

“*We fitted models with varying complexity in their random effects and used Leave-One Out (LOO) cross validation to compare model fit and select the model with best predictive performance. Using LOO, the expected log pointwise predictive density for a model can be calculated, and these can be used to compare model performance – by calculating the difference between expected log pointwise predictive density of various models. Differences of less than 4 mean that models are comparable (*[*Sivula et al, 2020*](#_ENREF_76)*). For differences greater than 4, then the standard error (SE) of the differences in expected log pointwise predictive density should be compared. If the standard error of the differences are much larger than the point estimate of the difference then the model closer to zero is preferred (*[*Sivula et al, 2020*](#_ENREF_76)*). The difference in LOO between models can be used for model selection, and in our case, gave similar results to model selection using Watanabe–Akaike Information Criterion (WAIC) (Table S1)*”

Reference: Sivula, T., Magnusson, M. and Vehtari, A. (2020) ‘Uncertainty in Bayesian leave-one-out cross-validation based model comparison’, arXiv:2008.10296.

Line 279. Gavrilets & Scheiner (1993) is a relevant reference.

**Response:** Thanks. We have added this reference.

Line 291. Applying the logarithm of the response variable is already a way of mean stan- dardizing the variance. Therefore inferences of the variance components can be linked to evolvability as mean scaled variances (e.g. Hansen et al., 2011)

What about sex differences?

**Response:** Thanks. We are not completely sure if Reviewer 1 wants us to mention the point about mean-standardising or would like us to take a different approach. We used the CV to mean-standardise the variance so that we could compare it to previous studies which adopt similar approaches. In regard to sex, animals are not easily sexed until after they are sexually mature so we have not looked at sex differences.

**Results**

314. The statement in the abstract about treatment differences in the variances should be toned down?

**Response:** Thanks. We have indicated a lack of statistical significance for this finding in the abstract.

Line 317-322. Something is wrong with the figure legend.

**Response:** We have now fixed the legend.

Line 332. Heading out of place.

**Response:** Fixed. Thanks.

**Figures**

Age was z transformed but graphs are presented with natural scale. How was the back transformation done?

**Response:** Age was back-transformed using the following equation:

Age\_raw = z\_Age\*SD(Age) + Mean(Age)

The mean and SD of age is provided in the legends.

Most legends are out of place.

**Response:** Sorry! We have now corrected these legends.

**Tables**

Line 390. Density of what?

**Response:** This was indeed unclear. We have now re-worded as follows:

“*∆ELPD represents the expected difference (on log scale) in predictive density for a new dataset estimated from cross-validation.*”

**Discussion**

Line 437. ”have others” change to ”others have”

**Response:** Thanks! Changed.

What are the consequences of choosing the wrong type of growth curve? Line 493. But there are lots of paternal half sibs right?

**Response:** Good question. For the data that we had, there would probably be little consequences of choosing the wrong growth curve because our data should be accurately described with a linear or quadratic parameter given that we did not measure much of the ‘plateau phase’ of growth. This takes longer given that lizards have indeterminant growth and will continue to grow (albeit more slowly) throughout their lives. In response to the second query about half sibs we have now added more details about the pedigree and percentages of maternal and paternal half sibs (See Table S11). Also, see comment to Reviewer 3 below.

Line 560-562. Reference?

**Response:** We have now added a reference to Charmantier and Garant (2005) and edited this sentence to hopefully reflect more clearly the point we would like to make.

Line 573: Cold-reared lizards had larger variance in incubation time. But what about the mean incubation time?

**Response:** We provide the details on mean incubation time in the results section:

“*On average, the incubation period for the ‘hot’ treatment was 29.36 days (SD = 2.17, range = 15 - 49) days and 48.48 days (SD = 4.18, range = 25 - 56) for the ‘cold’ treatment*.”

The mean incubation time is not as relevant to the discussion point being made here compared with the variance. The key point is that, because development time is non-linear with an asymptote being reached at higher temperatures, colder incubation conditions that fluctuate (as was the case in our study) will experience greater variability in the speed of development. This is expected to result in variable rates of resources acquisition and lead to higher variances in hatching time and body size at hatch.

Line 556-557. The fact that variance in body mass coincided with changing from single individual habitats to five individual habitats suggests that intraspecific competition may have triggered a response attributed to previous experiences by mothers. For instance see Marshall (2008). One of the concluding remarks is that there are changes in the mean body mass but not in the genetic variance across treatments, but what would be the population-wide consequences of (environmentally driven) changes in the mean body size in response to temperature?

**Response:** This is an interesting point. Thank you for pointing us to Marshall (2008). This is a very relevant study. We’ve now included this possible explanation in the discussion. In terms of what the population-wide consequences might be to environmentally driven changes in body size are hard to say given that we conducted a lab-based study, and only measured one trait. We would anticipate that larger body size would result in higher survival but given that we did not measure these we prefer not to be too speculative. Especially given that there may be trade-offs with other traits relevant to fitness.

**References**

Chevin, L.M., Lande, R. & Mace, G.M. (2010) Adaptation, plasticity, and extinction in a changing environment: Towards a predictive theory. PLoS Biology, 8.

Gavrilets, S. & Scheiner, S.M. (1993) The genetics of phenotypic plasticity. vi. theoretical predictions for directional selection. Journal of Evolutionary Biology, 6, 49–68.

Hansen, T.F., P ́elabon, C. & Houle, D. (2011) Heritability is not Evolvability. Evolutionary Biology, 38, 258–277.

Marshall, D.J. (2008) Transgenerational plasticity in the sea: Context-dependent maternal effects across the life history. Ecology, 89, 418–427.  
  
**Referee #2 (Remarks to the Author):**   
  
I have read the manuscript Heritability and developmental plasticity of growth in an oviparous lizard by Kar and colleagues, where they address to which extent different incubation temperatures affect heritability, maternal effects and growth trajectory in skinks. In general, the sample sizes are good, the methods well described and the statistical framework solid. There is, however, a few major concerns that have to be properly dealt with in order to enable interpretation of data. The most serious concern is that hatchlings are housed in individual cages, but at a specific time point moved to a joint cage. After this, the heritability and maternal effects increased. This could, however, be due to that both heritability and maternal effects become more visible when competition and social stress is present as the lizards are housed in the same cage. The data will therefore need to be analyzed separately for the period in single cages and the period being co-housed in order to avoid this to confound the findings. 

**Response:** Thanks for these comments. The point about social housing is a good one. We did already mention this being a possible reason for changes in variance across age in the discussion. See the section of the discission below.

We have considered the comment about analysing the data separately before and after social housing carefully. However, we are not sure how analysing the data separately will avoid the “confound” between estimates of genetic variance and growth. It will only decompose sources of variance across age before and after social housing. Such an approach is not expected to tell us anything different than our current models that estimate changes in variance across age because we are already detecting the changes after social housing (although we are happy to get more clarification from R2 on this point!).

We agree it would have been good to keep housing consistent, but this was logistically not possible given space constraints. In addition, and as pointed out by reviewer 3, power is possibly an issue. Fitting these models with less mass data on only a fraction of the growth curve will likely inflame this issue further.

Given the above, and responses to R3, we have now tried to be clearer that social housing was necessary due to logistical constraints. We have also made mention in the discussion that this could be an important component driving changes in variances at this point in age as Reviewer 2 correctly points out.

Methods:

‘*For logistical reasons, at approximately nine months, hatchlings were housed in groups of five in opaque bins with the same measurements as the adult enclosures. We pseudo-randomised individuals to each shared enclosure while maintaining a similar number of individuals from each treatment. Social housing conditions may result in additive genetic and maternal effects becoming more apparent because of competition and social stress that may drive greater variation among individuals. Our modelling approaches that estimate changes in variance components across age should be able to detect any changes brought about by the release of variation (see below).*

Discussion:

“*Genetic contributions to body size are expected to vary throughout ontogeny (Lynch & Walsh, 1998). Selection pressures on body size are likely to increase at critical life stages, such as at birth or at sexual maturation, thereby reducing genetic variance at certain ages (Rollinson & Rowe, 2015). On the contrary, we found that additive genetic variance of mass was very low upon hatching but slowly increased by the end of the first year. This result parallels those seen in big horn sheep (Réale et al., 1999), soay sheep (Wilson et al., 2007) and ladybird beetles (Dmitriew et al., 2010). While the underlying cause of this pattern is not well established, it coincided with changes in the social environment (shared housing). This suggests that perhaps competition for resources (basking sites or food) may orchestrate changes in genetic variation (Dmitriew et al., 2010; Hoffmann & Merilä, 1999).”*

**Major comments**   
  
The experimental design confounds the effects of socially interacting with other animals, where stress and competition arise, with time after hatching as the housing regime is changed in the middle of the experiment. This makes it impossible to distinguish between effects of timing. This has to be properly dealt with, including a thorough reanalysis of data where data is split in “single cage” and “5 individuals in the cage” data sets analyzed separately. The discussion and results and abstract will need to be rewritten based on these findings. 

**Response:** As pointed out above, we believe that our models are already distinguishing the changes in variance across age ‘before’ and ‘after’ social housing conditions because we are explicitly estimating changes in variance across age. In fact, in the discission we point to this as a major reason for why variance estimates change because those changes coincide with the move to social housing conditions. We have made this clear in the discussion:

“*This result parallels those seen in big horn sheep (Réale et al., 1999), soay sheep (Wilson et al., 2007) and ladybird beetles (Dmitriew et al., 2010). While the underlying cause of this pattern is not well established, it coincided with changes in the social environment (shared housing). This suggests that perhaps competition for resources (basking sites or food) may orchestrate changes in genetic variation (Dmitriew et al., 2010; Hoffmann & Merilä, 1999).”*

What is cold and what is hot? While extremely stressing temperatures have been shown to give rise to a release of additive genetic variation, would this be expected for temperatures within the natural range experienced by the species? The motivation for the temperatures of interest and how extreme they are in relation to those in the literature should be introduced already in the introduction and materials and methods, and the findings should be discussed in relation to how extreme the temperatures in studies with similar or different results are. 

**Response:** A fair point. The temperatures chosen are at the extremes of temperatures experienced by animals in nature as we have indicated in the methods:

*“The ‘hot’ treatment was exposed to a mean temperature of 29ºC whereas the ‘cold’ treatment was exposed to a mean temperature of 23ºC. Both incubators fluctuated +/- 3ºC over a 24 hour period around these mean temperatures to simulate natural nest site temperature variability. These treatments represent the temperature extremes of natural nest (~ 2 standard deviations above and below the mean - ~27 ºC) sites for L. delicata (Cheetham et al., 2011)*.”

It is difficult to compare whether these are extreme relative to other species in the literature given the very different thermal experiences of each species. We have now added more clarity around this point in the methods:

“*We chose these temperatures because were expect thermal environments to become more extreme and variable in the future making it of interest in knowing how the expression of genetic variation is likely to manifest in abnormal thermal conditions. While it is challenging to determine if an environment is ‘stressful’ or not without data on egg mortality (Roelofs, Morgan et al. 2010), we viewed this as atypical of what is commonly encountered in nature.*”

The presence of less than 1% multiple paternity could potentially give rise to spurious patterns in the animal model, it would be great to investigate if omitting these outliers change the findings. 

**Response:** It’s unclear what R2 means by these being outliers. From the perspective of the model, they are not. The 1% of multiple paternity just means that the number of clutches sired by multiple males was small. We don’t think that it would be wise to exclude these as this was part of the breeding design and are contributing information relevant to the estimation of variance components.

**Minor comments**  
  
L26-28 Are these significant differences or not?

**Response:**  Good point. We have now clarified:

“*On average, additive genetic variance, maternal effects and heritability were higher in hot developmental temperature treatment, however these differences were not statistically significant*”

L28 This statement cannot be made given the current confusion with housing conditions 

**Response:** It is true that we do not know the mechanisms driving the increase (e.g, something intrinsic vs social housing), but it did increase with age. Either way, we agree this needs clarification and we have now modified as follows:

“*Heritability increased with age, whereas maternal effects decreased upon hatching but increased again at a later age which could be driven by social competition or intrinsic changes in the expression of variation as animals growth*”

**Introduction:**   
  
Motivate the choice of the egg stage for studying these effects. 

**Response:** Thanks. We agree. We have now added the following sentence:

***“****For many oviparous (egg-laying) organisms, early life stages are particularly sensitive periods because many species do not provide parental care that would shelter embryos from environmental insults.”*

Add information on sex determination in relation to temperature for readers that do not know much about skinks 

**Response:** Thanks. We agree. We have now revised the following sentence to capture this point.

*“Regardless, phenotypic plasticity represents a promising immediate solution for threatened populations by allowing them to better track adaptive optima and persist providing the population experiences environmental conditions they have experienced in the past”*

When suggesting that plasticity might be a way to adapt to novel conditions, it is important to phrase it to reflect that adaptive plasticity typically acts within the range of conditions that are naturally experienced by the populations 

**Response:**

Add information on how extreme the conditions experienced were in relation to the range of conditions the organisms are exposed to in nature for the studies that are referred to that give contrasting patterns of additive genetic variation under stress 

**Response:** Thank you. See our comment below. We have now added more detail to the methods section as this seems to be the most appropriate place for these details.

**Methods**   
  
Include a discussion on how extreme the temperatures are in the experiment (e.g. in SDs or other measure that enables standardized comparisons), why these temperatures were deemed interesting, and how they are likely to affect the outcome in terms of release of additive genetic variation. 

**Response:** We have now included a bit more detail on the choice of temperatures in the methods. Cheetham *et al.* 2011 provide thermal data of nest sites. These sites were shown to have a mean temperature of 27ºC with an SD of approximately 2ºC; hence why we chose 23 and 29 C. We were specifically interested in manipulating incubation temperatures at the extremes on naturally occurring temperatures as we were interested in understanding how more extreme temperatures are likely to affect the release of additive genetic variation. We have revised as follows:

“*These treatments represent the temperature extremes of natural nest (~ 2 standard deviations above and below the mean - ~27 ºC) sites for L. delicata (Cheetham et al., 2011). We chose these temperatures because were expect thermal environments to become more extreme and variable in the future making it of interest in knowing how the expression of genetic variation is likely to manifest in abnormal thermal conditions*.”

L505-511 this information should be given already in the introduction and materials and methods 

**Response:** Thanks. We have added this to the methods where we feel it is most appropriate in the “*Developmental Temperature Manipulations*” section.

L531 competition and stress are expected to change, leading to larger differences. See major comment, I suggest reanalysing these periods separately. 

**Response:**

L536 This will need to be rephrased after reanalysis. 

**Response:**

L571 I think constrained has other connotations not suitable here 

**Response:** Completely agree. We have changed this to “comparable” and deleted constrained.

L581-582 I think the large differences in eclosion time and how they might interact with size and affect fitness is a bit neglected and could be mentioned here. 

**Response:** We think Reviewer 2 is referring to ‘incubation time’ rather than eclosion time. ……

**Referee #3 (Remarks to the Author):**  
This is a straightforward MS presenting the results of a well designed and executed quantitative genetic study in a lizard species (delicate skinks). The authors tackle the important question of whether genetic variance differs across environmental conditions. They use a paternal half-sib breeding design and a randomized split-clutch experiment to estimate genetic and maternal effects of size and growth and conclude that there is no strong environmental effect on these variance components. A major strength of the work is the verification of paternity via SNPs, which allows the variance components to be estimated using a genomic relatedness matrix.   
  
In general, I believe that this study should be published in Heredity, but I think the authors should consider a few things in a revision before the MS is published. I detail a couple of substantive concerns below, followed by a short list of copy-editing suggestions. 

**Response:** Thank you for your very helpful, and quite constructive, comments. We have now tried to take on board as much of the suggestions provided. We have also explored in far more depth some of the issues raised. We detail these below.

The main concern I have with the study is the potential that the design does not have a great deal of power to separate maternal effects from additive genetic effects. Because it is a paternal half-sib breeding design, the power to estimate the additive genetic effect will come from having **multiple dams nested within sires**. **Maternal effects will be more easily estimated when there are larger families, multiple generations, and/or when there are multiple sires per dam**. Otherwise, there is the potential for a confound between the genetic and the maternal effect. Here, multiple paternity is low, and as far as I can tell, family size seems to be small. I notice that there are 144 dams but only 262 offspring. This means that the average family size is just under two. Then, when the offspring are separated into treatments, there seems to fewer than one offspring per dam in each treatment. Incidentally, it would be nice to have a table detailing pedigree statistics (sires, dams, family size, etc.) within each treatment. 

**Response:** Thank you. We agree. We have now added to the ESM a table (Table S11) that details offspring from each dam-sire pair for each treatment. We have also provided more summary information here and in the MS to better flesh out the details.

Comparing Tables S4 and S5 to Table S6 suggests that G and M may indeed be confounded. The M components are much larger (and G slightly smaller) in S4/S5 when the treatments are split than when they are both included. This suggests to me that G and M are confounded due to the lack of replication within dams in S4/S5, but that when all offspring are included, this is ameliorated somewhat, leading to a more reasonable value for M. There also seems to be a very strong correlation between slope and intercept for M, further suggesting that the model is overparameterized. 

I recommend that the authors consider comparing models that do not include maternal effects in their model comparison approach (Table 1). **If I am correct that there is a confound, I would expect G-only models to provide similar fit to G+M models**. 

**Response:**

As far as I can tell, Model 7 was selected based on the full dataset rather than the split data set (line 254). Is this correct? If so, the authors may be selecting a model to analyze their split dataset that has too many parameters. 

**Response:** This is an excellent point. We have now conducted the same model selection procedure for the subset (cold and hot) growth data. We provide the model selection tables in the ESM.

One of the findings is that the proportion of variance explained by maternal effects declines and rebounds. I believe that this conclusion hinges on analyses from the split data set (figure 3). If so, I would treat this conclusion with extreme caution given the potential confound mentioned above. I would also like to see how these curves look from the combined dataset. 

**Response:**

I found the statement noting the rarity of half-sibs in line 493 to be curious, as it seems to ignore paternal half-sibs, which should not be rare given the breeding design, and focus only on maternal half-sibs. I suggest editing this, as it has the potential to be misleading. That said, it would be nice to discuss the origin of the few maternal half sibs. Is this because females had stored sperm from previous matings in the wild, or because some females were mated multiply in the lab? 

**Response:**  Thanks. We now provide a full breakdown of the data in each of the treatments (Table S11), and both the number of maternal and paternal half sibs in the dataset. We have also edited this sentence as suggested.

I suspect that none of these considerations will affect the authors' main conclusion that G does not significantly differ across environments, but it would be nice to confirm that this result is robust to a model that does not include maternal effects. 

**Response:**

Minor copy-editing changes:   
30: Hyphenate "age dependent"

**Response:** Changed.

85: I think this should be "have" instead of "has" here, because the clause modifies "traits", not "heritability". Also needs a comma before "which".

**Response:** Changed.

393: "influenced" -> "influence"

**Response:** Changed.

412: "is" -> "are"

**Response:** Changed.

Supplement:   
35: "To avoid overfitting" is repeated

**Response:** Thanks for catching this. We’ve removed this duplication.