Round #1

Author's Reply: Dear Troy Day, Stewart Plaistow and one anonymous reviewer,

Thank you very much for having read and corrected the manuscript. We carefully considered your suggestions and we hope our responses and modifications suit your expectations.

Two figures have been added: Figure 1 now describes the different treatments and number of individuals, and Supplementary figure 1 represents the experimental protocol. The statistical analyses have also been reviewed: only 4 traits are now investigated to limit false discovery rate. A multivariate approach was used to extract a single morphological trait instead of investigating 6 morphological traits separately. These statistical modifications did not change our previous results and interpretations. We also better explained the reproduction procedure which was not clear in the precedent version, and could limit the pseudo-replication concern.

We hope this new version of the manuscript is clearer and acceptable.

Sincerely yours,

Juliette Tariel, Sandrine Plénet and Émilien Luquet

Decision
by Troy Day, 2019-05-06 16:42
Manuscript: 10.1101/589945
Revision

I think this is an interesting paper that demonstrates quite nicely the potential complexity of transgenerational plasticity. The reviewers also felt that the paper makes a useful contribution to this important and growing area of research and they provided a number of very helpful comments. However, they also raised a couple of potentially serious concerns that I think need to be addressed. One has to do with the statistical analysis and false discovery – perhaps taking multivariate approach as suggested would help to alleviate this concern to some extent. The second concern though has to do with pseudo-replication. It is not entirely clear to me how best to deal with this issue but it is clearly something that needs to be addressed (both reviewers mention this point).
I had two other questions:
(1) How repeatable do the authors think their results are? Given the seemingly varied and complicated patterns of response as a function of past environment, one wonders if the patterns would remain consistent across multiple experiments.

> We are confident that we will get the same results if we were able to run multiple experiments in the same experimental conditions (genetic variation, parameter variations, etc.), which is impossible! However, if we rerun a similar experiment, we can be sure that we will get similar take-home messages because the strength of our paper is to not (over-) interpret specific effects but to show and discuss an overall pattern (1/ stronger effect of the current (offspring) environment compared to the effect of past environments, 2/ effects of the parental and/or grand-parental environments that can interact with the offspring environmental effect and 3/ these

effects can depend on the nature of traits considered).

(2) Given the results it is difficult to imagine how these responses could be adaptive. One can always make up stories but that isn't very satisfying. It would be nice if the authors could provide their thoughts on the sorts of experiments that might be done to test whether these kinds of patters are adaptive or simply "noise".

> You are right, and this is what we mean in the draft (L 259-260). The best way to test if the multigenerational responses are adaptive will be to exposed snails (from our different environmental histories) to a real (lethal) predator and to compare the survival of these different lineages (see Auld & Relyea, 2011). One of the co-author already done this kind of selection experiment in the past (Teplitsky, Plenet, & Joly, 2004). We added this perspective in the discussion (L 260-263).

It would be helpful to have legends on Figure 1.

> Done page 9 now on Figure 2.

Also, it should be stated that the vertical dashed line separates the offspring treatment groups.

> Done L 206.

Reviews

Reviewed by anonymous reviewer, 2019-04-28 12:38

In this study the authors investigate how exposure to predators of grandparents, parents, and self affects several behavioural and shell traits in an aquatic snail. This was determined by running a full factorial design under controlled laboratory conditions. We still know little about transgenerational plasticity, especially for generations earlier than parental ones, so this is an interesting and useful study. The design is adequate, and the sample sizes seem sufficient (although it is not mentioned how many parameters are estimated per trait – see comment below). However, I think the statistical analyses can be improved in several ways and there is especially a risk of overinterpreting the results due to inflated significance under multiple testing. I therefore make several suggestions that I think will be relatively easy to implement and that hopefully will lead to improved and more robust results. This may also change the interpretation and key messages of the study.

L 50: there are a few minor english grammatical mistakes throughout the text, e.g. here faced with or facing

> Done L 65

L 60: move "to ...cues" to the end of the sentence

> Done L 78

L 85: any idea why this might have happened? How could this have influenced your results?

> Not really. We always observed such an effect across ca. 5 generations when we initiate experimental lineages from wild individuals, which is likely a modification of trade-offs in resource allocation according to the laboratory conditions.

This does not influence our results because all snails were exposed to the same laboratory conditions (except the treatments) and here we compared the response of snails within the same (F3) generation.

L 91: might it be easier and shorter to reflect the experimental design, including treatments and sample sizes, in a figure? If possible, also tracing the relatedness of the individuals, i.e. family membership.

> We added two figures: Figure 1 page 9 shows the experimental design with explicit sample sizes and Supplementary Figure 1 represents the experimental protocol. Our experimental design did not allow to follow the pedigree (the different families are merged at each copulation step to generate the next generation).

L 93: this is a non-native crayfish species – is the response of your snails as large and in the same directions as for native crayfish? Provide some results and references on previous studies showing responses in your snail using this predator treatment.

> Both the snail *P. acuta* and the crayfish *Procambrus clarkii* are invasive from North America in France. We added this information L 88-90 and L 128-129. This crayfish species and others (e.g. *Orconectes limosus*) coexist with *P. acuta* in their natural and invaded areas. In the *P. acuta* literature, different crayfish species are used (*Promcambarus acutus*, Auld & Releyea 2011, *Procambarus simulans*, Beaty 2016) and the defenses induced are always similar and in the same directions.

In France, there are currently only three native species (all endangered and very rare) and none lived in the same habitat than *P. acuta*.

L 99 and 100: reverse the order of these two sentences

> Done L 132-136

L 103: weighed, and omit weight. Dry or wet weight? Snail + shell? Or just the shell (which makes more sense)?

> We used the total wet weight (body + shell) and gently dried snails with paper towel before weight measurement (information added L 137-138). To distinguish body and shell weights required to kill the snails. As we need snails to generate the next generation, we could not have this information.

L 104: photograph

> Done L 138

L 106: so is that the thickness of the rim of the aperture, not the shell width?

> It was indeed the thickness of the edge of the aperture (information modified L 143)

L 113: it might be a good idea to standardise your variables before modelling, both for estimations purposes (improving collinearity issues) and interpretation (all estimates on the same scale)

> We scaled weight, shell thickness and all morphological variables (information added L 158-159).

L 116: why not just run the full model, and provide all t-, F- or Chi square- and p-values? That provides more information (e.g. for posterior meta analysis), and prevents biased estimates in the remaining variables. See the paper by Forstmeier and Schielzeth cited below.

> As suggested, full models (F- or Chi square- and p-values) are now provided.

L 117: how is the relationship between weight and the other variables? I would guess it would be exponential with an exponent close to 3, if so then don´t take the ln of weight, but the cube root as a covariate. Check for all your models that the residuals have the desired distributions.

> You were right, the best fit between weight and morphological variables was an exponential with an exponent close to 3, the cube root of the weight is now used as a covariate.

L 118: I lost track of the genetic relatedness between the individuals in your design, but isn´t it necessary to model relatedness in each generation (current family, parental family, grandparental family), i.e. as several nested random variables? Even if these higher levels are not significant, I would include them anyway as they take care of pseudo-replication in the design.

> Our experimental design did not allow to keep track of the whole pedigree. To generate F_2 and F_3 generations, we merged snails from each of the 15 maternal families (copulation

groups) losing their identity. At each generation, we only have information about the maternal family used in analyses as a random genetic variation effect. These information were now added L 96-97 and L 104-105.

Having said that, how many parameters are you estimating? Per parameter you should have at least 5 to 10 data points to obtain reliable estimates.

> We should have reliable estimates. In a full model, we are estimating 7 parameters (3 main effects and 4 interaction terms). Our total sample size is N = 375 (N / parameters = 375 / 7 ≈ 53). If you refer to the Figure 1 (page 9), you will see that we have ca. 40-80 points per estimate (meaning individuals per combination of environments at the 3rd generation), except for two environmental histories where we have only 9 and 12 individuals.

Another issues is the large number of traits and parameters that are tested in this study (many interactions per trait, and many traits), which could increase the probability of obtaining significant results by coincidence. The authors should study and implement ways to control for this, for example via false discovery rate. There are R packages that implements several options to control for inflation of significance. As an exercise, simply randomise the data of your response variables (but according to the structure of your design, so not across families for example) and re-run the models, and see if you get any significant effects (I predict you would). See e.g. Cryptic multiple hypotheses testing in linear models: overestimated effect sizes and the winner's curse by Wolfgang Forstmeier and Holger Schielzeth.

> To avoid this issue, we reduced the number of tests by reducing our analyses to only 4 traits (crawling-out behaviour, shell thickness, weight and shell morphology). We used a principal component analysis (a vector approach) to extract a global morphological response instead of 4 traits (L 149-155).

Traditionally, multiple comparisons are not taken into account when interpreting the p-values of three-way ANOVA. We explored our results for each trait after p-value corrections (e.g. Bonferroni $p < \frac{0.005}{7} = 0.007$ or Holm procedure) and overall it did not change the effect significance. The only change would be the E1 x E3 interaction for the shell thickness ($p = 0.04$ without correction) but without consequence on our interpretations. Indeed, the strength of our paper is to not (over-) interpret specific effects but to show and discuss the complexity of an overall pattern. We thus decided to not add the corrected p-value to lighten the analyses.

In view of my suggested improvements in the statistical analyses, and thereby possibly changing the results and interpretation of the study, I have not in detail reviewed the discussion. Nonetheless, the discussion is multifaceted and balanced, and in that sense is acceptable. Figure 1 and table 1 are also good in their design.

Reviewed by Stewart Plaistow, 2019-05-01 18:10

Review of manuscript Tariel *et al.* -

The hermaphroditic gastropod, *Physa acuta* has previously been used extensively to document predator (*Procambarus clarkii*) induced phenotypic plasticity. However, the effect that predator cues have over multiple generations has not previously been investigated. In this study, Tariel *et al.* used a multi-generational, factorial experiment to test the hypotheses that predator-induced defences in *Physa acuta* accumulate across generations and/or are influenced by predator generations. Tariel *et al.* concluded that multi-generational effects were sometimes observed, but they were not cumulative and depended on the trait considered. I found that the paper was generally well-written and well-referenced and tackles a question that is currently very interesting with respect to recent attempts to understand

how the integration of genetic and non-genetic cues influence phenotypes. However, there is a major pseudo-replication issue that hasn't been accounted for that would I think seriously jeopardise the conclusions of the study. If the pseudoreplication issue can be accounted for, I think a multivariate statistical approach might enable a clearer interpretation of the findings.

Major comments

1) My main issue with the paper as it is at the moment is that offspring from each family in each generation were reared together in the same tube for the first 28 days. As a result, they cannot be treated as independent observations even if they were then reared separately from day 28 through until day 35. This is a real problem because it means that any phenotypic differences we see in the F3 offspring are not just a result of multiple generations of predator cues (or not), they might also be the result of 'within tube' effects that are caused by the different phenotypes caused by predator cue effects in each generation. Since we already know that predator cues do induce phenotypic differences in this species, this could easily be confounding and makes understanding how multigenerational effects relate to each other especially difficult. One way around this issue might be to do a further experiment testing whether the phenotypes of offspring from a single family are any different in predator cue and no predator environments when reared together for 28 days and then separated; or separated from the start. If the phenotypes were equivalent, you might then argue that 'within tube' effects are non-existent and can be ignored.

> We agree that offspring are not real independent observations as they were reared in the same vials (one per family per treatment) for the first 28 days. However, a 'within tube' effect in each generation could not cause a generational effect because to produce the next generation, we merged one snail from each of one tube (forming 6 copulation groups of 15 snails / treatment) and then we randomly selected 15 new offspring families (from 90 in total). This means that even if there was a significant 'within tube' effect, it was diluted at each reproduction and cannot influence the phenotypic difference we observed (or at least bring noise).

2) This paper tested the effect of factorial multigeneration predator cues on individual traits even though many of the traits concerned are likely to be part of a phenotypically integrated anti-predator defence. While I can appreciate that the behavioural trait is binomial response variable and might require an independent analysis (although a % response isn't), the other traits could have been analysed using either MANOVA, perMANOVA or a phenotypic vector approach so we can understand how multigenerational cues influence the overall antipredator response (see these references: (*1-3*)). This approach is advantageous because it enables us to better understand how the different components of an anti-predator response co-vary with each other. It reduces the total number of tests you need to do, thereby reducing the possibility of type II errors. Moreover, it helps us to avoid the rather unsatisfying conclusion that the existence of grand-parental, parental and offspring environment effects depends upon which trait you look at.

> To avoid this issue, first, we used a principal component analysis (a vector approach) to extract a global morphological response instead of 4 traits (L 149-155). A single morphological trait was enough to describe more than 95% of the total variance in morphological traits. We then reduced the number of tests by reducing our analyses to only 4 variables (crawling-out behavior, shell thickness, weight and shell global morphology).

We preferred to analyze 3 of these 4 variables with univariate linear mixed models, with weight as a covariate. We think that shell thickness and shell morphology should be analyzed separately

as thickness is not correlated to morphological variables after accounting for weight. Moreover, mixed MANOVA and phenotypic vector approach of Collyer & Adams do not allow to model our generational design (1/ mMANOVA only allows to test for environmental effects within each family and between families, which has no sense in our study (offspring within a family have the same grand-parental and parental environmental history); 2/ phenotypic vector approach cannot disentangle between the random effect of family and the interaction between grand-parental and parental environments (idem a family is characterized by a unique grand-parental and parental environmental history)).

3) I struggled to understand the experimental design. An experimental design figure would make the paper much easier to follow.

> We added two figures. Figure 1 page 9 now describes treatments and sample sizes; Supplementary Figure 1 represents the experimental steps.

4) You could make it clearer throughout how you are testing whether predator effects accumulate or not across generations. With your current model such effects would presumably show up as part of complex higher order interaction terms. It might be worth considering a simpler model with treatment fitted as 8-levels and specifically contrast phenotypes that had pure cues (e.g. PPP and CCC with those that experienced mis-matched environments). Your results in Figure 1, a,b,d clearly suggest that the biggest phenotype differences were often between PPP and CCC offspring (but this could ofcourse be because of the problem outlined in point 1).

> In our current model, we can already assess if predator effects accumulate or not. For instance, there were significant main effects (grand-parental and offspring environments) on behavior in the same direction with no interaction, their effects were then cumulative. One environment act independently from the other environment, but their effects can accumulate. In the case of an interaction, it rather means that the effects are not just cumulative, but could be antagonistic or synergetic.

5) Although you mention that this is a hermaphroditic snail, a brief section on the biology of *Physa acuta* in the methods might be a good idea given that so few people work on hermaphroditic organisms.

> Done L 88-90

Minor comments

Ln 12 – You use 'WGP' but it's not defined until the introduction.

> Changed to within-generation plasticity in the abstract L 18

Ln 35/6 – change to 'A few examples have shown……'

> Done L 48-49

Ln 60 – Re-phrase 'We exposed, to environments without and with predator-cues….'

> Done L 78-79

Ln 104 - There isn't enough detail here for anybody to be able to repeat your work. What height was the camera? What settings etc. Full detail required.

> Done L 139-140.

Ln 197/8 – Doesn't make sense.

> Beginning of the sentence *"As previous theoretical and experimental works (references),"* changed to *"similar to theoretical and other experimental studies"* and moved to the end of the sentence L 253-254

Ln 199-203 – this section is hard to follow and sounds contradictory.

> Sentences have been changed L 252-259 to clarify.

References

1. D. C. Adams, M. L. Collyer, A GENERAL FRAMEWORK FOR THE ANALYSIS OF PHENOTYPIC TRAJECTORIES IN EVOLUTIONARY STUDIES. *Evolution* **63**, 1143-1154 (2009).
2. S. J. Plaistow, H. Collin, Phenotypic integration plasticity in Daphnia magna: an integral facet of G × E interactions - Plaistow - 2014 - Journal of Evolutionary Biology - Wiley Online Library. *Journal of Evolutionary Biology* **27**, 1913-1920 (2014).
3. M. L. Collyer, D. C. Adams, Analysis of two-state multivariate phenotypic change in ecological studies. *Ecology* **88**, 683-692 (2007).

References

Auld, J. R., & Relyea, R. A. (2011). Adaptive plasticity in predator-induced defenses in a common freshwater snail: altered selection and mode of predation due to prey phenotype. *Evolutionary Ecology*, *25*(1), 189-202. https://doi.org/10.1007/s10682-010-9394-1

Beaty, L. E., Wormington, J. D., Kensinger, B. J., Bayley, K. N., Goeppner, S. R., Gustafson, K. D., & Luttbeg, B. (2016). Shaped by the past, acting in the present: transgenerational plasticity of anti-predatory traits. *Oikos*, *125*(11), 1570-1576. https://doi.org/10.1111/oik.03114

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, *6*(2), 65-70.

Teplitsky, C., Plenet, S., & Joly, P. (2004). Hierarchical Responses of Tadpoles to Multiple Predators. *Ecology*, *85*(10), 2888-2894. https://doi.org/10.1890/03-3043