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.

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.

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

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).

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.

Minor comments

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

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

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

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.

Ln 197/8 – Doesn't make sense.

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

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).