Dear Dr Aubret,

Thanks for your decision on our manuscript formerly entitled 'The successful invasion of the European earwig across North America reflects adaptations to thermal regimes but not mean temperatures' and now entitled "Thermal regimes, but not mean temperatures, drive patterns of rapid climate adaptation at a continent-scale: evidence from the introduced European earwig across North America".

We are happy to see that both referees and you are enthusiastic about our work. We have followed all suggestions and addressed all comments. In particular, we have completely changed the angle of our study to position it in the climate change framework. We have also addressed the few concerns about the interpretation of your results (phenotypic plasticity versus genetic) by adding a new paragraph in the discussion on how the use of a common garden experimental setup allowed us ruling out the direct effect of phenotypic plasticity on our results. Finally, we have addressed all the questions and followed the other suggestions that have been raised by the two referees throughout the manuscript.

Overall, we would like to thank the two referees for the time they spend reviewing our manuscript and providing such important and insightful inputs. We are convinced that our manuscript is now of much better quality and of a broader scope.

You can find a point-by-point reply to the comments of each reviewers below, as well as a track-change version of our manuscript attached to this submission.

Sincerely,

Joël Meunier & Jean-Claude Tourneur

Additional comment from the managing board:
As indicated in the 'How does it work?' section and in the code of conduct, please make sure that:
-Data are available to readers, either in the text or through an open data repository such as Zenodo (free), Dryad (to pay) or some other institutional repository. Data must be reusable, thus metadata or accompanying text must carefully describe the data.
-Details on quantitative analyses (e.g., data treatment and statistical scripts in R, bioinformatic pipeline scripts, etc.) and details concerning simulations (scripts, codes) are available to readers in the text, as appendices, or through an open data repository, such as Zenodo, Dryad or some other institutional repository. The scripts or codes must be carefully described so that they can be reused.
-Details on experimental procedures are available to readers in the text or as appendices.
-Authors have no financial conflict of interest relating to the article. The article must contain a "Conflict of interest disclosure" paragraph before the reference section containing this sentence: "The authors of this preprint declare that they have no financial conflict of interest with the content of this article." If appropriate, this disclosure may be completed by a sentence indicating that some of the authors are PCI recommenders: "XXX is one of the

PCI Evol Biol recommenders."

To follow the guidelines, we have added the two requested sections (data availability L393-394; and conflict of interest L401-402) at the end of the manuscript. Moreover, the complete data set and R script have been archived in the open data repository Zenodo (https://doi.org/10.5281/zenodo.2652192).

Reviewed by Eric Gangloff, 2019-02-25 09:07

Tourneur and Meunier present an excellent data set of life-history observations across 19 populations of an invasive insect species across North America. The scope of the data is impressive and have the potential to offer important contributions to our understanding of variation in life-history traits in response to environmental conditions. In its current form, however, the manuscript significantly oversteps in its conclusions about adaptation to novel environments and contains some important areas that require further explanation or justification. I hope that the two major concerns below, as well as minor concerns and editing suggestions, help to improve the manuscript for publication.

The major story of the paper is that variation in life-history traits is related to differences in temperature regimes and that this represents an adaptation to novel environments in North America. However, no information is presented on the native habitats of this species and how these might differ from those in North America. Such data are necessary to demonstrate that these are in fact novel adaptations rather than responses to climates already present in the native range. At the minimum, it would be useful to demonstrate that at least some of these climates are outside of the native range (for example, is the Montreal climate beyond the range experienced by this species in its native range?). If such a comparison is not possible, then I don't think that this paper demonstrates this response to temperature as an adaptation to novel environments. Nonetheless, the paper is certainly interesting and valuable in demonstrating the covariance of life-history traits with climate patterns.

Thanks for this comment. We believe that this issue stems from our unclear and misleading definition of "novel environment" in the manuscript. Sorry for that. Reviewer 1 is correct when stating that our study cannot provide any conclusion on the capability of the European earwig to adapt to environmental conditions that are absent in its native range (i.e. reviewer's understanding of "novel environment"). However, we believe that our study provides clear and robust conclusions on the capability of the European earwig to change its life-history traits in response to the broad set of encountered conditions after it has been introduced in a new continent (our definition of "novel environment"). To limit the risk of misunderstanding, we have edited the text to change the term "novel environment" by "novel location" or "novel population", when appropriate, as well as have changed several sentences to clarify that our conclusions are drawn in the context of colonisation of new locations. Furthermore, we are convinced that the new framework of our manuscript (i.e. adaptation to climate change) helps limiting this risk of misunderstanding.

My second major question is around the conditions in which animals were kept in the lab. Few details are provided other than that these conditions were those of Montreal. So were

animals kept outdoors, subject to natural weather variation? Or in climate chambers mimicking natural conditions? Given that this is likely one of the coldest regimes experienced by this species in North America, using this as a common garden condition provokes questions about the observations: the design exposes insects native to different climates to one of the extreme climates. In other words, these results don't represent what insects do in the field in their (newly established populations), but rather how they respond to an extreme condition (Montreal's climate). While this does not invalidate the results, this would seem to indicate that the data do not address the question of variation across climate regimes, but rather response to an extreme climate. In my opinion, this is a major issue with the experimental design that would need to be addressed fully before publication.

This is an excellent point. We apologize for not having provided such detailed information on the rearing conditions. We kept each field-sampled group of individuals outdoor (under a shelter) from the date of their arrival in Montreal until mid-October. At that time, pairs of adults were setup and then maintained in a climate chamber at 10±1 °C until the end of the experiment. We have added this information in the main text L144-145, 155-158).

The temperature used in our common garden setup is not extreme regarding the temperature range of the other populations (Table S1). In particular, our pairs were setup at 10°C, which is close to 9.5°C - the overall median temperature of the 19 sampled populations. We have added this information in the main text (L155-158).

With this information, it is therefore unlikely that our results only reflect how individuals react to extreme thermal conditions. Nevertheless, this interesting comment raised an important point that we previously forget to mention in the text: Common garden experiments do not preclude the possibility of genotype-by-environment interactions on the measured life-history traits, which can in turn affect the strength and nature of the reported patterns of adaptation. To address this important point, we have added a new paragraph in the discussion, in which we present the benefits and limits of common garden experiments, as well as emphasize that future studies may explore the role of rearing temperatures on the patterns of adaptation reported in tour study (L347-368).

In addition to these major points, I offer some small editing suggestions that I hope will improve future versions of the manuscript:

Thanks a lot!

Line 25: The definition of 'thermal regimes' (as presented in line 60) would be useful to include in the abstract as well.

Thanks for the suggestion. We have completely rewritten our abstract to present our result in the context of global warming. Nevertheless, we still use the term "thermal regimes" and have therefore edited the abstract to provide its definition. In particular, the new sentences is "… this species modified 10 of the 13 measured life-history traits in response to the encountered thermal regimes, defined as a variation of temperatures between seasons or months (here winter-summer and autumn-spring temperatures)." (L24-27)

Line 31: There are numerous reviews and studies demonstrating this in insects and other

invertebrates as well.

We have rewritten the beginning of the introduction to present our result in the context of global warming. We therefore do not focus on the phenomenon of invasive species anymore and the corresponding sentence has been deleted.

Line 42: Change 'in' to 'to'

Done.

Line 47: Change to "British Columbia" (not hyphenated)

Done.

Lines 51-52: It would be helpful to provide just a little bit of detail about why this species is considered a pest. What are its negative effects?

The European earwig is considered as an agricultural pest due to its consumption of soft fruits and young leaves in orchards. Because the new framework of our study has shifted to adaptation to global warming, we decided to keep the focus on the invasive history of the European earwig across North America, while removing all information on its potential effects as a pest. For this reason, we would like to not add these details in the main text.

Lines 54-55: Change 'drove' to 'driven' (or consider re-wording sentence)

The sentence has been deleted in the new introduction.

Line 56: Change to "Elderberry" (no hyphen)

Done.

Line 63: Change to "life cycle" (two words, no hyphen)

Done.

Line 65: The phrase 'depending on the population' is a bit ambiguous. Suggest changing to "with variation among populations" as this is more precise (and I think captures the intended meaning)

Thanks for the suggestion. This is exactly what we intended to mean. We have changed the sentence accordingly: "The life cycle of this species generally starts with the emergence of new adults in late spring to early July, with variation among populations." (L92)

Line 113: Suggest changing "grounded" to "lined" or changing wording

We have followed this suggestion and changed the word "grounded" into the word "lined" (L154)

Line 127: I don't think 'homogenous' is the word to express this idea. Do you mean that the data are unimodal, indicating a single continuous distribution of traits?

Sorry for the mistake. We have edited the sentence to make it clearer. The new sentence is now "Although these subspecies were not considered in our analyses (our data were collected before the publication of these genetic analyses), the continuous distribution (unimodal data) of the life history traits measured across populations

(Figures 2 to 4) suggests an absence of species-specific values regarding these measurements." (L168-171)

Line 141-142: Without information on the age at collection, I don't think that these observations can be used as a measure of longevity. As detailed later (Lines 280-282), these adults may have hatched at different times, thus confounding any measure of longevity of field-collected individuals of unknown age. While this is acknowledged, the potential to bias results and interpretation is not addressed.

This is true. To avoid misinterpretation, we have changed "longevity" by "experimental survival duration" throughout the manuscript. We have also edited the text to clarify its difference with adults' longevity. The new sentence is "Although our measurement of survival duration does not necessarily reflect adults' longevity, as individuals could have different age at field sampling (see discussion), it nevertheless provides important insights into the period at which males and females of each population die during the year." (L186-189)

Line 152: There appears to be a word missing after 'a priori' – maybe 'assumptions of data structure'?

We apologize for this lack of clarity. The sentence has been edited into: "To reduce dimensionality of co-varying temperatures in our data set while characterizing potential thermal regimes of each population without a priori definitions of their composition, we then conducted a Principal Component Analysis (PCA) on the set of 12 mean monthly temperatures per population (Table S1)." (L196-199)

Lines 173-175: These methods are unclear. The cbind command adds a vector as a column in a data frame. I don't understand how this relates to this analysis.

When analyzing ratios, one potential issue is to not take into account the number of individuals on which each ratio has been calculated. For instance, a ratio of 0.2 based on 10 individuals should weigh less on the final result than a ratio of 0.2 based on 10000 individuals. To control for this issue, one option is to enter the numerator and denominator of the ratio directly into the statistical model. This can be achieved with the command *cbind* in R. Here, the *cbind* command allows us to enter two columns as a response variable in the GLM: one with the number of iteroparous females and one with the number of semelparous females. This is a common statistical procedure to compare ratios calculated on different samples sizes (see the book R in action, by R Kabacoff, Manning edition)

We have edited the text to make it clearer. The new sentence is now: In the GLM, the response variable was the ratio of iteroparous females per population, which was entered using the command cbind in R (to weight each ratio by the sample size of its population) and fitted to a binomial error distribution corrected for overdispersion." (L219-222).

Lines 175-176: The meaning of this sentence is unclear. Does this mean that dependent variables were centered around a mean of 0? Where they also standardized to a SD of 1? This requires clarification.

Sorry for the lack of clarity. We did not centre our data. Our sentence indicates that for every statistical model, the response variable was the mean value of each

measured trait per population. In other word, we used one value per population and this value was the average of all the measurements performed in the given population. This allowed us to limit the risk of Type I errors based on very large data sets. Note that the results are qualitatively the same when we entered every data within population (instead of the means) as response variable, and interpreted the results based on effect sizes (again to limit Type I errors). We have edited the sentence to make it clearer: "In all our statistical models, the response variables were the mean values of each measured trait per population" (L222-223)

Line 178: What was considered non-significant?

We have edited the sentence to clarify what was considered as non-significant. The new sentence is: "They were also checked for homoscedasticity and residuals normality, as well as simplified stepwise by removing all non-significant interaction terms (all $p > 0.05$)" (L223-225)

Lines 178-180: Was the FDR adjusted to provide an experiment-wide Type I error rate of 5% (P = 0.05)? This is generally the case and seems to be here, but should specified.

Yes, this is exactly the purpose of our correction. We have added this information in the main text (L225-227)

Line 177: Change "residuals normality" to "normality of residuals"

Done.

Lines 232-236: This sentence seems to conflate plasticity in response to immediate environments with canalization within a population. The argument is that higher summer temperatures will accelerate physiology, which leads to earlier reproduction. However, the experiment was conducted under common garden conditions, which tests for either canalization or the effect of previously-experienced environments. This point also applies to the conclusions drawn in lines 244-247.

This is an excellent point. Thanks for having spotted it. We are sorry for our wrong interpretation. We have profoundly edited the discussion to correct this mistake. We now explain that our results could reflect a plastic response, but that our common garden setup precludes from such an interpretation. We then conclude that our findings overall suggest that the observed changes in the timing of first reproduction and females' reproductive strategy may have first emerged as a plastic response to the thermal constraints of the different localities, then diverged between populations through canalization to ultimately become inherited traits – all this in less than 100 generations. (L291-300)

Line 252: The phrase "decreased together with autumn temperatures" is unclear. Suggest changing to "decreased with decreasing autumn temperatures" (I think this is the intended meaning).

Changed. Thanks for the suggestion. (L303)

Line 298: If the first documented introduction was in 1907, shouldn't this be "a bit more than a century"?

True. Nevertheless, we have decided to focus our new conclusion on the number of

generations (to avoid confusion and strengthen our take home message), which is less than a 100 (our field sampling finished in 1996). We changed the sentence into To conclude, our results demonstrate that the spread of the European earwigs across North America came with important changes in their life-history traits and life cycle, and that these changes emerged in a maximum of 100 generations" (L369-371)

Line 319: This closing phrase is awkward. Suggest "…to which the present study contributes."

The closing phrases has been changed, but its final part has been conversed and corrected, accordingly. (L388-389)

Table 1, Fig. 1: Change "Ashville" to "Asheville"

Done. We also corrected it in table 1.

Table 2: It would be useful to provide Eigenvalues of each PC as well.

We have added the Eigenvalues of each PC to Table 2

Table 3: The adjustment used for P-values doesn't seem to make sense. For example, a range of raw p-values from 0.012 to 0.025 all result in corrected P-values of 0.041. Furthermore, some of the P-values don't change at all.

Our correction follows the original Benjamini-Hochberg (FDR) procedure. This is a step-by-step procedure, which transforms each P-value in function of its rank of statistical significance within the entire data set and allows providing an experiment-wide Type I error rate of 5%. This procedure can be done using the command p.adjust() in R. Because it is based on ranks, the corrected p-values can be similar even if the non-corrected p-values are different. Note that this procedure is different from the mean FDR (MFDR), which simply consists in transforming the alpha threshold using a specific formula. Although both FDR and MFDR provide qualitatively similar results, we do prefer the use of the original FDR procedure as it provides readers with both original and transformed p-values, while keeping the traditional p>0.05 interpretation criteria. Having said that, we understand that the results provided in the table can be misleading and have therefore edited the legend to provide the required information: "Note that FDR correction transforms each P-value in function of its rank of statistical significance in the data set, which can lead to similar corrected p-values." (L555-557)

Reviewed by Ben Phillips, 2019-02-25 09:09

General comments

This manuscript reports an impressive sampling effort in which life-history traits were measured across 19 populations of earwig in North America. The authors examine whether variation in life-history across populations is correlated with aspects of the thermal climate experienced by each population: mean temperature and seasonality of temperature. They find some fascinating correlations between life-history and thermal climate; correlations with the seasonality of temperature, but not with mean temperature.

The manuscript is well written and reported, and I see no major flaws with the science. There is an issue of interpretation that needs thought, however, which is how much we expect the measured traits to reflect plasticity (e.g., simple constraints) versus genetic differences. My reading is that, because the animals were collected as juveniles and grown in lab conditions for most of their development and all of their reproductive lives, life-history variation here would most likely reflect genetic differences. If plasticity is involved, it might be through maternal effects rather than direct environmental constraints.

We agree that the former version of our manuscript did not discuss the determinants of the reported associations, and did not emphasize (enough) that our common garden experimental setup allows disentangling the effects of plasticity versus genetic (or early life experience). In accordance with this comment (and the one of the first reviewer), we had deeply edited the abstract, introduction (predictions) and discussion, in which we have specifically added a new paragraph on this point (L347-368)

The other place where I think the manuscript can be improved substantially is by adjusting the focus. This is not just a dataset that speaks to invasive species, but it has important relevance to climate change and adaptation to climate, generally. I feel like the authors are missing a big opportunity here by not linking their work to that literature in thoughtful ways. It is increasingly appreciated, for example, that species adjust their phenology in response to climate change. Here you have some fascinating data on precisely that, with time scales relevant to impending climate change, and a reasonable argument that this is adaptation, not just constraint. Much of the literature looking at adaptation in animals in response to climate change has focussed on physiological traits (see, for example, some relevant examples on *Drosophila* in Australia in work by Sgro and Hoffmann), this study would be a relatively uncommon example of a study that examines life-history traits.

Thanks a lot for this suggestion. We have followed it and reframed our entire study in the context of adaptation to global change. To this end, we have changed the title, rewritten the abstract, as well as the beginning of the introduction and parts of the discussion.

Specific comments

Introduction

L47: might be worth mentioning that dispersal is primarily human-mediated (I assume). Some of the basic setup here reminds me of Huey *et al.* 2000 Science 287:308-309.

This is a good suggestion. We agree that the dispersal of *F. auricularia* across North America is very likely to be human-mediated. However, we currently have no data to support this claim, and the capability of this species to fly could allow such long-distance dispersal without human intervention (even if long-distance flying capability remains unknown in this species). To avoid being speculative, we therefore would like to not mention this information in the main text.

Methods

L128-130: It is a shame you can't identify subspecies post hoc, because it could add substantial strength to the study by allowing you to examine whether there are convergent responses across the two subspecies.

We fully agree with this comment. Although all the animals involved in the present study were subsequently maintained in alcohol, they have all been lost in the 2000s when JC Tourneur get retired from the university.

L152: A word missing here? "PCA; an analysis without *a priori*".

Sorry for that. We have edited the sentence to clarify what a priori referred to. The new sentence is now: "To reduce dimensionality of co-varying temperatures in our data set while characterizing potential thermal regimes of each population without a priori definitions of their composition, we then conducted a Principal Component Analysis (PCA) on the set of 12 mean monthly temperatures per population (Table S1)." (L196-199)

L157, 161: these might be best expressed as capturing variation in seasonality, rather than a "trade-off". It seems both PC2 and 3 capture the degree of seasonality (PC2: summer-winter; PC3: fall-spring); this might be more reader-friendly terminology throughout.

This is an excellent suggestion. Thanks. We have edited the sentences accordingly: "The second component (PC2) revealed variation in seasonality between February on one hand, and June, July, and August on the other hand. In particular, high values of PC2 reflected populations with cold February (winter) and warm summer, whereas small values of PC2 reflected populations with warm February (winter) and cold summer. Finally, the third component (PC3) captured variation in seasonality between October and November on one hand, and April and May on the other hand." (L201-208)

L167: lm() in R specifies a linear model (OLS fitting procedure), not a generalised linear model (likelihoodbased fitting).

This has been corrected (L213). Sorry for the mistake.

Results

Nicely done.

Thanks.

Discussion

L224: the time-course of data collection is not relevant here, and has the potential to confuse. Suggest dropping reference to it. Instead, X clutches across 19 populations is the more informative metric.

Thanks for this suggestion. We have changed our sentence accordingly: "In particular, our data from 19 populations revealed that females changed their timing of first reproduction, their reproductive strategy and investment into egg production when facing different thermal regimes, while both males and females experimental survival duration varied accordingly." (L271-274)

L238: apostrophe of posession: "females' physiological.."

Changed.

L244: "Overall, these results suggest that changes in the timing of first reproduction and females' reproductive strategy did not evolve to better cope with novel thermal constraints,

but instead that they are simple by-product of these constraints." It is great that you consider the fact that there are simple energetic/metabolic constraints driven by temperature, but all your animals were collected as juveniles and raised in constant lab environment, correct? Given this, how much of the variation you see in the lab do you think you can ascribe to energetic constraints in the collection environment?

Sorry for the misunderstanding. Our individuals were almost exclusively field-sampled as adults, even if a few of them were old nymphs (i.e. last developmental instar). As a result, we believe that it is impossible to rule out that a (significant) part of the variation we see in the lab results from energetic constraints experienced by each individual during its early development in the collection environment.

We have edited the text to make it clearer. In the material and method section, we now explain "These individuals were mainly collected as adults using wooden traps …" (L139-142). In the new paragraph of the discussion, we also emphasize that "Our tested individuals were mostly collected as adults in the field, so that we cannot rule out that population-specific constraints occurring during their early developmental stages have had long-lasting effects on their life-history traits (English et al. 2016)." (L356-358).

270-274: fascinating!

Thanks a lot.