Dear Olivier Roux,

Thank you for your decision on the manuscript entitled “Relationship between weapon size and six key behavioural and physiological traits in males of the European earwig” we submitted for recommendation in PCI Zoology.

We have addressed all the reviewers’ comments. In particular, we have added information to better detail why we measured each of the six traits, we have added a figure to clarify the timeline of the experiment and the associated variation in sample size, and we have added power analyses to confirm the robustness of our approach. We have also edited the text to take account of all the minor suggestions and corrections made by the reviewers.

Overall, we believe that these changes have significantly improved the clarity, robustness and impact of our study. We are therefore very grateful to the three reviewers for their constructive comments and suggestions.

We have also updated the link to the repository of our R script and dataset as we have included the new power analysis in the script.

A detailed point-by-point response to the referees’ comments is provided below.

Sincerely,

Joël Meunier, on behalf of all co-authors

Revision round #1
Decision for round #1 : *Revision needed*
Revision required

Dear authors
Your manuscript has been reviewed and reviewers are enthusiastic about the topic. However, they raised concerns about the methodology and the interpretation of the results. The choice of traits measured also needs to be justified so that the reader can better understand their relevance to this study. The weaknesses of the study should also be better highlighted.
Looking forward to your revision.
Kind regards
Olivier Roux
by ***Olivier Roux****, 22 May 2024 10:20*
Manuscript: https://doi.org/10.1101/2024.03.20.585871
version: 1

> We have carefully revised the manuscript, incorporating changes that we believe address their concerns and improve the overall quality of our study.

==========

Review by anonymous reviewer 1, 29 Apr 2024 13:20

In this manuscript, authors measured six behavioural and physiological traits in the European earwig under the hypothesis that having long forceps comes with some behavioural and/or physiological cost and/or strength for males. However, authors could not find any associations at least in six behavioural and physiological traits. According to the result, authors questioned that long forceps functions as "high quality signal". All of experiments are well described and their limitations are also clearly mentioned. I think this manuscript can be published in the journal. In discussion section, authors carefully interpreted their results and significance with enough previous studies. Also, authors clearly stated limitations of their study and explained alternative possibilities.

Thank you.

My only concern is that authors did not mentioned about ecological cost (such as predation rate and flight ability) of forceps. I think it is better to add possible ecological cost and/or benefit of forceps in discussion section.

This is a good point. Unfortunately, there is not much information available on the potential costs of carrying long forceps in this species. We have therefore edited the discussion to emphasise that our results call for more research on this topic: "They also call for further research to quantify other potential costs of carrying long forceps in this species, for example in terms of predation rates and ability to fly (Crumb & Eide, 1941)." (L319-321)

==========

Review by anonymous reviewer 2, 21 May 2024 09:00

This paper explores the relationships between the size of extravagant forceps in male earwigs and several behavioural/physiological traits supposed to be linked with fitness (namely: locomotor performance, boldness, aggregation behaviour, survival under harsh conditions, sperm quantity and survival to pathogen exposure). The rationale for this study is that forceps length can be an indicator of male quality, but also that the forceps length can trade-off with other fitness-linked traits. I found the study interesting in its aims, but I have several questions/comments about the experimental process.

1- The choice of measuring certain traits should be better justified in relation to their importance on animal fitness. For example, in what locomotor activity is important for earwigs? Is sperm count important? (the long duration of copulation mentioned to justify this measurement can also be due to mate guarding, therefore having no relationships with sperm count) …

Sorry for the lack of information. We have edited the introduction to better justify the importance of the traits we have measured: "We then measured their level of expression of three important behaviours, their survival rate in two distinct harsh conditions and their sperm quantity. The first behaviour was their locomotor performance (Cheutin et al., 2024), which reflects the ability of males to walk long distances (to forage, hide or find a mate) while carrying long and heavy appendages. The second behaviour was their likelihood to flee after a physical disturbance (i.e., boldness), which shows how males react when disturbed by a predator attack (Thesing et al., 2015). The third behaviour was their propensity to aggregate with conspecifics. This is an important parameter in this gregarious

species, as adults typically live in groups of up to several hundred individuals and social isolation can have detrimental effects on their physiology (Kohlmeier et al., 2016; Van Meyel & Meunier, 2022). We also measured their survival rate in a harsh condition where they were isolated with no access to a food source for 31 days, and then their survival rate after exposure to the common entomopathogenic fungus Metarhizium brunneum (Vogelweith et al., 2017). Finally, we measured the level of sperm storage in seminal vesicles in each male, a parameter that is often important in the context of male-male competition for mating (Shuker & Simmons, 2014).“ (L102-116)

We have also edited the last prediction of the introduction, to emphasize that the long duration of mating may be due to more sperm transfer, but also other mating strategies: “We also predict that 6) long-forceps males would produce more sperm and contain more sperm in their seminal vesicles than short forceps males. This last prediction would be consistent with the longer duration of copulation reported for the long-forceps males (Kamimura, 2014), even if this longer duration may reflect other male mating strategies, such as mate guarding” (L123-125)

2- L. 121-123: Prior to experiments, animals were maintained in groups for 3 months, so that it is highly probable that the males used were not virgin... What is the impact on sperm count? (how to control that all males copulated or not before measurements?)

This is correct, the males used in this experiment were not virgin. We believe that it is not an issue for our measurements, as we measured sperm count about two months after they have been separated from females, which allow every male to recover from sperm production independent of their previous mating rate. To clarify this point, we have edited the method section to mention that “Three months later (i.e., at the end of the reproductive season), we removed the females from all the terrariums to mimic their natural dispersal. One month later, for each population, we visually selected the 30 males with the longest forceps and the 30 males with the shortest forceps (Körner et al., 2017) and isolated them in individual Petri dishes (diameter 5 cm) for …” (L135-138) and that “This counting occurred about two months after the males were separated from the females, which is probably long enough for the males to rebuild their sperm reserves, regardless of their previous mating rate.” (L196-199)

3- During the whole experiment, males were isolated and not fed. I would like the authors discussing this harsh condition. First, why is it so? Why not feeding them? Second Is there not a risk that the weakening of the animals will level down all their behavioral performances or sperm count, thus erasing all potential variation?

Earwig are very resistant to starvation and we have previously observed that earwigs can live very well for several months without having access to a food source (J Meunier, pers. Obs.). In our experiment, we starved them for two main reasons. The first one was to determine whether this apparent resistance to starvation was indeed common to earwigs from different populations and with different forceps length. The second was to make sure that good rearing conditions would not mask any potential investment trade-off between forceps length and other life-history traits. The general male behaviour we observed in this study did not appear to be altered compared to the male behaviour we generally observe under normal

rearing conditions, although we have no data (published or unpublished) to support this claim.

We have edited the text to explain why we starved the males for 31 days: "We then kept the isolated males without access to food from the day they were isolated until 31 days after their isolation (Figure 1) to test whether resistance to both starvation and social isolation (i.e., harsh environmental conditions) were population and/or forceps-length specific, while ensuring that good rearing conditions did not mask any potential investment trade-offs between forceps length and other life history traits." (L150-154)

4- Two trait were measures after one-month isolation, during which only 15 to 50% males died, variable among groups and populations. There is therefore a risk for a differential selection of best individuals. Such a selection may confound sperm count and survival to pathogen infections.

This is a good point. We have added this information in the method section on sperm production: "It should be noted that sperm storage was measured in males that survived 31 days in isolation without access to food, so that this represents the sperm storage of individuals best adapted to these two stressful conditions." (L213-215) and on survival to pathogen exposure :"As with sperm storage, survival after pathogen exposure was measured in males that survived 31 days in isolation without access to food, so that it represents the survival rate of individuals best adapted to two these stressful conditions." (L233-236)

5- The sample size gets smaller and smaller as time goes by. I suggest running a power analysis. I guess that owing the variation observed, sample size is too small to conclude firmly. At least there is no strong effect.

As suggested, we have run power analysis of each of the statistical models used in this study. The values are presented in the new table 1 and ranges from 0.155 (sperm production) to 0.447 (survival under harsh conditions). We have also changed our wording throughout the manuscript to emphasise that we did not detect an effect, or that our experiment did not allow us to detect an effect, rather than stating that there was no effect (e.g., L26, L259, etc). We believe that reporting the power of our analyses and providing this nuanced wording should clarify for readers the strength of the non-significant effects. Thank you for the suggestion.

6- In Table 1, numbers in the last line are not consistent with those provided in the text (Where 39 individuals are given for sperm count and 52 for survival to pathogens).

There was an error in the former Table 1. We apologise for this. We have corrected these numbers in the new Figure 1.

7- L. 239. Aggregation score is at 0.05, so this result should be considered and discussed.

We have edited the results and discussion parts, accordingly: "Regardless of forceps length, males from Cinais were less likely to flee after a physical disturbance and less gregarious than males from Valence" (L266-268), "These findings were consistent between the two populations, although some of the traits measured were population specific: males from Cinais were generally bolder, less gregarious and (only if they carried long-forceps) had a better chance of surviving in harsh

conditions than males from Valence" (L283-286) and "While we found no overall difference between short- and long-forceps males, our data reveal population differences in terms of males' boldness, aggregation level and resistance to starvation. Cinais had males that were generally bolder, less gregarious, as well as males with long forceps that survived food deprivation better than Valence." (L323-326).

8- L. 252-253. There is a significant interaction between survival and population (Table 2). Therefore there is no "no effect". Please be consistent with your own analysis.

We were talking about main effect, but we agree that it was confusing. We have changed the text accordingly: "Contrary to predictions, our experiment did not allow us to detect an association between forceps length and locomotor performance, boldness (i.e., likelihood to flee after a physical disturbance), aggregation behaviour, sperm production, and male survival after pathogen infection. These findings were consistent between the two populations, although some of the traits measured were population specific: males from Cinais were generally bolder, less gregarious and (only if they carried long-forceps) had a better chance of surviving in harsh conditions than males from Valence." (L280-286)

To conclude, in the discussion, before discussing the background, I would like the authors acknowledging the potential weaknesses of their study that could confound the results (points 2-5 before). While discussing the limitations of most previous experiments is fine, it would be fair to discuss the limitations of the present one.

We have edited the abstract, results and discussion to clarify the potential limitations of our study and emphasise that we did not detect an effect of forceps length on the measured traits rather than stating that there was no effect. In the abstract: "Contrary to our predictions, we detected no main association between forceps length and the traits measured." (L26-27). In the results: "Overall, we detected no main association between forceps length and the six traits measured (Table 1)." (L259-260). In the discussion: "Contrary to predictions, our experiment did not allow us to detect an association between forceps length and locomotor performance, boldness (i.e., likelihood to flee after a physical disturbance), aggregation behaviour, sperm production, and male survival after pathogen infection." (L280-283). More generally, we believe that the clarifications we have made in the manuscript regarding the points raised above will also help readers to recognise the potential limitations of our study.

==========

Review by Luna Grey, 09 May 2024 00:41

In regards to "Relationship between weapon size and sex key behavioural and physiological traits in males of the European earwig", I think that this paper presents an interesting set up to assess whether forceps length is actually a reliable signal for male quality. However, I feel that the methods could be clearer as to the timeline of the tests. Ideally, each of these tests would have been done prior to them being starved. Additionally, I would like a bit of clarification on why some of these traits were chosen. Some of the traits make sense (boldness, resilience to

pathogens, etc.) but some I don't quite see the rationale for (e.g. aggregation). This is not to say there is none, but it should be stated explicitly in the text.

Thank you for these supportive comments. We have carefully edited the text to clarify the timeline of the test (we have added a figure for this) and to clarify the rationale for selecting the measured traits (see detailed response below). Thank you for the detailed review and important insights into the manuscript.

Line 23 – the sample size is a bit misleading here as I assumed that out of hundreds of individuals you chose these 120 individuals. I think this can be made clearer by saying something to the effect of: "selecting 60 individuals with the longest and shortest forceps from two populations and the…."

We have edited the sentence, accordingly: "We sampled hundreds of males from two populations, selected 60 males with the longest and shortest forceps from each population, and then measured locomotor performance, boldness, aggregation behaviour, survival under harsh conditions, sperm storage, and survival after pathogen exposure." (L23-24)

Line 24 – It would be interesting to explain why you chose these traits

We have edited the sentence to explain that these traits were important: "Here, we tested whether forceps length is associated with six important behavioural and physiological traits in males of the European earwig." (L21-22). However, it is difficult to provide more detail in the abstract due to space constraints. We have therefore edited the introduction to describe each of these traits and provide details of their importance in male fitness: "We then measured their level of expression of three important behaviours, their survival rate in two distinct harsh conditions and their sperm quantity. The first behaviour was their locomotor performance (Cheutin et al., 2024), which reflects the ability of males to walk long distances (to forage, hide or find a mate) while carrying long and heavy appendages. The second behaviour was their likelihood to flee after a physical disturbance (i.e., boldness), which shows how males react when disturbed by a predator attack (Thesing et al., 2015). The third behaviour was their propensity to aggregate with conspecifics. This is an important parameter in this gregarious species, as adults typically live in groups of up to several hundred individuals and social isolation can have detrimental effects on their physiology (Kohlmeier et al., 2016; Van Meyel & Meunier, 2022). We also measured their survival rate in a harsh condition where they were isolated with no access to a food source for 31 days, and then their survival rate after exposure to the common entomopathogenic fungus Metarhizium brunneum (Vogelweith et al., 2017). Finally, we measured the level of sperm storage in seminal vesicles in each male, a parameter that is often important in the context of male-male competition for mating (Shuker & Simmons, 2014). " (L102-116)

Line 37 – This sentence is a little awkward, I would put the examples of extravagant structures in parentheses. "take a variety of forms (such as: antlers, horns, spurs, fangs and tusks), and work to enhance the male's fighting ability and/or attractiveness to females (Emlen, 2008).

Changed, accordingly: "These weapons and ornaments are typically large and extravagant morphological structures that can grow on different parts of the male's

body, take a variety of forms (such as antlers, horns, spurs, fangs and tusks), and work to enhance the male's fighting ability and/or attractiveness to females (Emlen, 2008)." (L38-41)

General – first reference to a species should include the authority of the species.

This is a good point. The European earwig is a complex of species which may have different names and belong to different clades. We have therefore provided both information in the text: "All these individuals belong to *Forficula auricularia* Linneaus, 1758, also called *Forficula auricularia* clade A (González-Miguéns et al., 2020)." (L131-133). As the authority names of the vast majority of the species we cite in the manuscript are not given in the cited papers, we have chosen not to speculate on them and to stick to the information given in the cited papers.

Line 50 – the plural of prey is prey

Changed. "This cost can arise from the fact that carrying heavy, bulky weapons (or ornaments) makes males less mobile and more visible to both predators and prey (Oufiero & Garland, 2007)." (L50-51)

Line 67 – Do you know if male earwigs have hyperallometric growth to their forceps? I think adding this information may help give context or explain the trends you are seeing, particularly about using forceps length as a proxy for male quality.

Unfortunately, this information is unknown in the European earwig (or any other earwig, to the best of our knowledge).

Line 78 – "or to interrupt mating by non-copulating males" is a bit confusing. May I suggest "or to interrupt mating individuals by non-copulating males". Originally it sounds as though the earwigs are interrupting their own copulation.

Thanks. We have followed this suggestion: "In the European earwig *Forficula auricularia*, male forceps are also used in male-male contests as a weapon to deter competitors prior to mating (Styrsky & Rhein, 1999) or to interrupt mating individuals by non-copulating males (Forslund, 2000, 2003; Walker & Fell, 2001)." (L77-80)

Line 81 – I would change "females do not seem to select their mate on the basis of forceps length" to "females do not seem to select their mate exclusively on the basis of forceps length" as I am sure that forceps length does play some role in female choice, even if a very small one.

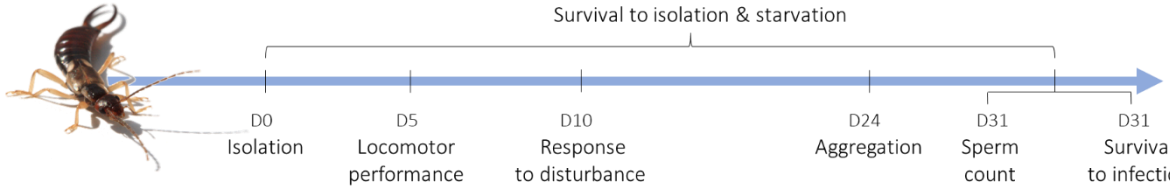
Good point. We have added "exclusively" in the sentence, as suggested: "Although forceps are involved in male courtship (Walker & Fell, 2001), females do not seem to select their mate exclusively on the basis of forceps length (Forslund, 2000, 2003; Radesäter & Halldórsdóttir, 1993; Walker & Fell, 2001)." (L82-84)

Line 102 - I think these are all good traits to test individually. But I would like some rationale as to why you chose these traits in particular.

We have edited the introduction to describe each of these traits and their importance in male fitness (see above) (L102-116)

Line 160 – The setup for this section was a bit confusing. I think that having a diagram might help make this clearer.

That is an excellent idea. Thank you. We have added a diagram showing the timeline of the experimental design and the evolution of the sample size over the course of the experiment as Figure 1, which you can find below:



Long forceps	Cinais	N=30	30	30	18	9	8
	Valence	N=30	29	29	28	10	18
Short forceps	Cinais	N=30	30	30	22	10	12
	Valence	N=30	30	30	24	10	14

Line 169 – I think this line can be simplified by rewriting it as: "We recorded whether …"

We have followed this suggestion: "We recorded whether the tested male was in the chamber next to the group of conspecifics (yes or no) and repeated this measurement by taking pictures …" (L186-187)

Line 176 – I think this should be moved up, perhaps at the beginning of the "Behavioural measurements" section (Line 144). There you can explain the timeline of your assessments and at the beginning of each test you monitored how many individuals were still alive and that you used this as an assessment of survival in harsh conditions.

We have edited the Methods section (as suggested below) to include a full section on survival measurements (L217). We have thus added (and expanded) the information on measuring males under harsh conditions in this new section:

"We measured male survival under two types of harsh conditions. The first type of harsh condition was the absence of any food source (starvation) combined with social isolation, which is known to have detrimental effects on this gregarious species (Kohlmeier et al., 2016; Van Meyel & Meunier, 2022). We assessed the male survival rate under these conditions by recording whether each of the 120 males tested was still alive on day 31 after isolation (Figure 1).

The second type of harsh condition was exposure to pathogens. We measured survival rate after pathogen exposure in the 52 males that were still alive on day 31 and were not used to measure sperm storage (Figure 1). We exposed each male to spores of the entomopathogenic fungus *Metharizium brunneum* (formerly *M. anisopliae*)…" (L217-225)

Methods – While reading I was having a bit of a hard time understanding the timeline of when you did your tests and I think having a section detailing this would improve that understanding. This could be the first paragraph where you explain the behavioural measurements. Here you could outline when you were doing each test, when you were starving the earwigs and when the last test was conducted.

Sorry for the misunderstanding. We have added a figure (Figure 1) to clarify the timeline of our experiment (see above).

Line 186 – I think you mean "dissecting microscope"?

Yes. We have changed "binocular" with "dissecting microscope" (L201)

Line 195 – "mal" should be "male"

Changed. (L210)

Line 198 - I think this should be its own section as otherwise the change in subject is too abrupt.

Excellent idea. We have followed this suggestion and (as explained above) made a new section focusing on how we measured survival in harsh environments and after exposure to pathogens. (L217-236)

Line 234 – The "(not short)" in "long (not short) forceps" doesn't seem to clarify the point you want to make as clearly as possible. I think adding explicitly that this trend was not found with in males with short forceps is important.

We have edited the sentence accordingly: "There was also no difference between males with short and long forceps in terms of survival under harsh environmental conditions, although males with long forceps survived better when they came from Cinais compared to Valence and that this trend was not present in males with short forceps (Figure 3; interaction in Table 1; pairwise comparisons: Long forceps Cinais vs Valence: Z = -3.311; P = 0.005; Short forceps Cinais vs Valence: Z = -0.611; P = 0.929)." (L262-266)

Line 269 - I believe that this paper will help make this argument as this shows that there are alternate morph for this species, albeit very difficult to identify. "Tomkins, J. L., Kotiaho, J. S., & LeBas, N. R. (2005). Matters of Scale: Positive Allometry and the Evolution of Male Dimorphisms. The American Naturalist, 165(3), 389–402.

Excellent point. We have added this reference to the text to support our argument: "This could be the case with the European earwig (Tomkins et al., 2005)." (L298)

Line 298 – I agree with all the possible reasons for this, but I would like to add that environmental conditions may be an important factor as well. Valence is a closed and forested area, the males from this region may be naïve to this type of perturbation.

We believe that reviewer is talking about Cinay, which is a forest, and not Valence, which is a very large cultivated orchard located in the middle of many other orchards. Nevertheless, we have added the possible difference in environmental conditions in the text: "Instead, it could reflect population idiosyncrasies that have affected their development, such as climatic conditions (Valence is warmer than Cinais), environmental conditions and/or exposure to phytosanitary products (e.g. Valence is a cultivated orchard, whereas Cinais is an uncultivated forest edge), or population-specific genetic background." (L329-332)

Line 317 – I think "growing longer forceps" might be more concise.

Changed. "While the results of these studies suggest that males can gain fitness benefit from growing longer forceps (Forslund, 2000, 2003; Radesäter &

Halldórsdóttir, 1993; Styrsky & Rhein, 1999; Walker & Fell, 2001), they contain two important limitations." (L347-351)