Speciation through selection on mitochondrial genes?

Astrid Groot based on reviews by Sabine Haenniger and Heiko Vogel

A recommendation of:

Marion Orsucci, Yves Moné, Philippe Audiot, Sylvie Gimenez, Sandra Nhim, Rima Naït-Saïdi, Marie Frayssinet, Guillaume Dumont, Jean-Paul Boudon, Marin Vabre, Stéphanie Rialle, Rachid Koual, Gael J. Kergoat, Rodney N. Nagoshi, Robert L. Meagher, Emmanuelle d'Alencon, Nicolas Nègre. **Transcriptional differences between the two host strains of Spodoptera frugiperda (Lepidoptera: Noctuidae)** *(2020), bioRxiv, 263186, ver. 2 peer-reviewed and recommended by Peer Community in Evolutionary Biology.* 10.1101/263186

Open Access

Whether speciation through ecological specialization occurs has been a thriving research area ever since Mayr (1942) stated this to play a central role. In herbivorous insects, ecological specialization is most likely to happen through host plant differentiation (Funk et al. 2002). Therefore, after Dorothy Pashley had identified two host strains in the Fall armyworm (FAW), *Spodoptera frugiperda*, in 1988 (Pashley 1988), researchers have been trying to decipher the evolutionary history of these strains, as this seems to be a model species in which speciation is currently occurring through host plant specialization. Even though FAW is a generalist, feeding on many different host plant species (Pogue 2002) and a devastating pest in many crops, Pashley identified a so-called corn strain and a so-called rice strain in Puerto Rico. Genetically, these strains were found to differ mostly in an esterase, although later studies showed additional genetic differences and markers, mostly in the mitochondrial COI and the nuclear TPI. Recent genomic studies showed that the two strains are overall so genetically different (2% of their genome being different) that these two strains could better be called different species (Kergoat et al. 2012). So far, the most consistent differences between the strains have been their timing
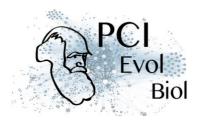
of mating activities at night (Schoefl et al. 2009, 2011; Haenniger et al. 2019) and hybrid incompatibilities (Dumas et al. 2015; Kost et al. 2016). Whether and to what extent host plant preference or performance contributed to the differentiation of these sympatrically occurring strains has remained unclear. In the current study, Orsucci et al. (2020) performed oviposition assays and reciprocal transplant experiments with both strains to measure fitness effects, in combination with a comprehensive RNAseq experiment, in which not only lab reared larvae were analysed, but also field-collected larvae. When testing preference and performance on the two host plants corn and rice, the authors did not find consistent fitness differences between the two strains, with both strains performing less on rice plants, although larvae from the corn strain survived more on corn plants than those from the rice strain. These results mostly confirm findings of a number of investigations over the past 30 years, where no consistent differences on the two host plants were observed (reviewed in Groot et al. 2016). However, the RNAseq experiments did show some striking differences between the two strains, especially in the reciprocally transplanted larvae, where both strains had been reared on rice or on corn plants for one generation: both strains showed transcriptional responses that correspond to their respective putative host plants, i.e. overexpression of genes involved in digestion and metabolic activity, and underexpression of genes involved in detoxification, in the corn strain on corn and in the rice strain on rice. Interestingly, similar sets of genes were found to be overexpressed in the field-collected larvae with which a RNAseq experiment was conducted as well. The most interesting result of the study performed by Orsucci et al. (2020) is the underexpression in the corn strain of so-called numts, small genomic sequences that corresponded to fragments of the mitochondrial COI and COIII. These two numts were differentially expressed in the two strains in all RNAseq experiments analysed. This result coincides with the fact that the COI is one of the main diagnostic markers to distinguish these two strains. The authors suggestion that a difference in energy production between these two strains may be linked to a shift in host plant preference matches their finding that rice plants seem to be less suitable host plants than corn plants. However, as the lower suitability of rice plants was true for both strains, it remains unclear whether and how this difference could be linked to possible host plant differentiation between the strains. The authors also suggest that COI and potentially other mitochondrial genes may be the original target of selection between these two strains. This is especially interesting in light of the fact that field-collected larvae have frequently been found to have a rice strain mitochondrial genotype and a corn strain nuclear genotype, also in this study, while in the lab (female rice strain x male corn strain) hybrid females (i.e. females with a rice strain mitochondrial genotype and a corn strain nuclear genotype) are behaviorally sterile (Kost et al. 2016). Whether and how selection on mitochondrial genes or on mitonuclear interactions has indeed affected the evolution of these strains in the New world, and will affect the evolution of FAW in newly invaded habitats in the Old world, including Asia and Australia – where, so far, only corn strain and (female rice strain x male corn strain) hybrids have been found (Nagoshi 2019), will be a challenging research question for the coming years.

References

[1] Dumas, P. et al. (2015). Spodoptera frugiperda (Lepidoptera: Noctuidae) host-plant variants: two host strains or two distinct species?. Genetica, 143(3), 305-316. doi: 10.1007/s10709-015-9829-2
[2] Funk, D. J., Filchak, K. E. and Feder J. L. (2002) Herbivorous insects: model systems for the comparative study of speciation ecology. In: Etges W.J., Noor M.A.F. (eds) Genetics of Mate Choice: From Sexual Selection to Sexual Isolation. Contemporary Issues in Genetics and Evolution, vol 9. Springer, Dordrecht. doi: 10.1007/978-94-010-0265-3_10
[3] Groot, A. T., Unbehend, M., Hänniger, S., Juárez, M. L., Kost, S. and Heckel D. G.(2016) Evolution of reproductive isolation of Spodoptera frugiperda. In Allison, J. and Cardé, R. (eds) Sexual communication in moths. Chapter 20: 291-300.
[4] Hänniger, S. et al. (2017). Genetic basis of allochronic differentiation in the fall armyworm. BMC evolutionary biology, 17(1), 68. doi: 10.1186/s12862-017-0911-5
[5] Kost, S., Heckel, D. G., Yoshido, A., Marec, F., and Groot, A. T. (2016). AZ-linked sterility locus causes sexual abstinence in hybrid females and facilitates speciation in Spodoptera frugiperda. Evolution, 70(6), 1418-1427. doi: 10.1111/evo.12940
[6] Mayr, E. (1942) Systematics and the origin of species. Columbia University Press, New York.

[7] Nagoshi, R. N. (2019). Evidence that a major subpopulation of fall armyworm found in the Western Hemisphere is rare or absent in Africa, which may limit the range of crops at risk of infestation. PloS one, 14(4).   doi: 10.1371/journal.pone.0208966

[8] Orsucci, M., Moné, Y., Audiot, P., Gimenez, S., Nhim, S., Naït-Saïdi, R., Frayssinet, M., Dumont, G., Boudon, J.-P., Vabre, M., Rialle, S., Koual, R., Kergoat, G. J., Nagoshi, R. N., Meagher, R. L., d'Alençon, E. and Nègre N. (2020) Transcriptional differences between the two host strains of Spodoptera frugiperda (Lepidoptera: Noctuidae). bioRxiv, 263186, ver. 2 peer-reviewed and recommended by PCI Evol Biol. doi: 10.1101/263186

[9] Pashley, D. P. (1988) Current Status of Fall Armyworm Host Strains. Florida Entomologist 71 (3): 227–34. doi: 10.2307/3495425

[10] Pogue, M. (2002). A World Revision of the Genus Spodoptera Guenée (Lepidoptera: Noctuidae). American   Entomological   Society.

[11] Schöfl, G., Heckel, D. G., and Groot, A. T. (2009). Time-shifted reproductive behaviours among fall armyworm (Noctuidae: Spodoptera frugiperda) host strains: evidence for differing modes of inheritance. Journal of Evolutionary Biology, 22(7), 1447-1459. doi: 10.1111/j.1420-9101.2009.01759.x

[12] Schöfl, G., Dill, A., Heckel, D. G., and Groot, A. T. (2011). Allochronic separation versus mate choice: nonrandom patterns of mating between fall armyworm host strains. The American Naturalist, 177(4), 470-485. doi: 10.1086/658904

Revision round #2

2019-12-04
Dear Marion et al,

Thank you for revising your manuscript. Both reviewers agree that the manuscript has improved considerably, however, both still have important comments that should be addressed before your manuscript can be accepted. Can you please revise your manuscript according to these new comments, and can you indicate in your manuscript where you made changes (e.g. by marking this text in red), as well as in a cover letter, where you detail how you addressed each of the comments of both authors exactly? This helps the reviewing process considerably.

Thanks   in   advance.
Kind   regards,
Astrid

Reviewed by *Sabine Haenniger*, 2019-12-04 11:15

Overall, the restructuring of the ms and addition of the field sample investigation has greatly benefited the ms. It has very interesting data and I enjoyed reading it. I thank the authors for taking my recommendations and comments into account and answering them to my satisfaction. I would like to see the ms published, after some minor revisions. My new comments arise despite the authors answering my previous comments, because the ms was completely remodeled. The story is very dense and especially the part of the differentially expressed genes is difficult to disentangle as a reader and reducing it to a single mitochondrial gene seems oversimplified and is not well justified. I can't really understand, which role the numts play or do not play and how exactly they are differentiated from their "parents" in the mt. This part needs to be simplified, so the reader can really follow the line of arguments and data.

Here are a few additional, more specific comments:

line 221ff: While I agree that both strains have not been in contact with rice compounds for many years, I would be very careful about the "corn compounds". The corn flour, originating from the cobs, is very different from the foliage that the insects encountered in the experiments and often in the field. I'd be surprised if the insects could adapt to defensive corn plant compounds from feeding on corn flour.

line 347ff: The mentioned figure does not exist and I fail to see which of the existing figures this could be. This is really a shame, because it does not become clear, how exactly co1 stands out over the rest of the differentially expressed genes. This set of genes, consistently overexpressed in both lab and field insects, sounds very interesting, yet there is no information about it, which would be very valuable. Co1 is still just 1 gene from the mitochondrion; many more are needed to make an impact on the energy household. Also, it is just 1 mitochondrial gene compared to the mentioned 76 + 73 genes which might be nuclear, so I find the statement that selection mainly acts on the mitochondrion too bold.
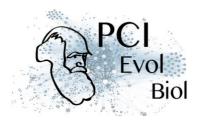
Figures: a homogeneous color code for all Figures would be great, i.e. to always use the same color for the rice strain in every figure. These colors should maybe not be the ones from Figures 1, 2, S2 (too radiant). The bars in these mentioned figures are too wide, use up too much space. Figure 2 should have the strain over each panel; font size is different in E+F compared to A-D. Figure 3 not suitable for red/green color blindness.

Reviewed by [Heiko Vogel](#), 2019-12-04 08:34

The authors have adequately addressed the majority of the reviewers comments. Particularly by toning down the biological interpretation of the gene expression analysis of the different strains on host plants/diet (which was not really well replicated), and by omitting any final conclusions regarding the adaptation to plants and instead focusing the study on the transcriptional differences between the strains, the manuscript became clearer and more straightforward. I also do agree with the authors that with now 4 replicates (well, actually strains on different diets and of different origins), there is the possibility to identify constitutive, fixed gene expression differences between strains, which are also unaffected by diet. This being said, the revised manuscript almost reads like a new paper, since the main focus is shifted to a larger extent. It was very hard to read the revised manuscript in the light of all the changes made, since there was no tracking of the changes.

However, what I still do not understand (maybe I am missing something here) is the argument related to differences in COIII/COI expression between the strains. The authors state that:".. With this level of differentiation between the strains, with the mitochondrion being a central organelle for cellular metabolism it is rather intriguing that we also find the most different transcriptional difference being associated to mitochondrion. Thus we do not think it is too far reached to state that two different types of mitonchondria, functioning differently, might indeed be the main selective event between the strains that explain both their behavioral differences (host plant preference) and their genetic divergence..." Well, I do think that it is actually rather far-fetched to connect differences in mito gene expression with host plant preferences. The authors make a very far-reaching and rather bold statement. How exactly would steady-state elevated ATP levels relate to behavioral differences, such as host plant preferences? Even if you argue that much more data is needed to address this, one could at least propose some kind of mechanism for this suggested causality.

Likewise, some aspects of the supposedly higher expression levels of COIII are strange: why was only a fragment of COIII (and not the complete mRNA) overexpressed - and how exactly does this fragment corresponded to a numt? The authors argue that "..(numts) sometimes confound gene prediction because they contain the open reading frame (ORF) sequence of the original mitochondrial gene. However, numts are usually not transcribed, lacking the promoter region sequence..." I would actually argue that in RNAseq data the occurrence of "genome background transcription" is frequent, where non-coding parts of the genome are represented in the RNAseq data. The authors also state that "In the case of the COIII-numt, the differential expression we measured comes from messenger RNAs of mitochondrial origin, whose reads also align on the numt region. In practice, numts

show differences of expression at the level of mitochondria..." I am puzzled - first of all because numts now actually do show differences in expression levels? The authors further argue that "Two numts in particular, corresponding to fragments in the mitochondrial genes COI and COIII are clearly differentially expressed in sf-C compared to sf-R in all the RNASeq datasets we analyzed" Again, the authors contradict themselves, since they on the one hand argue that numts are not expressed, but on the other hand state that two numts corresponding to the mitochondrial CO genes are. In case the authors are certain that the observed expression differences in COIII/COI are not from numts (which would require a re-phrasing of the text passages): how have the authors concluded that the sequence reads are from the mitochondrial genome and not from a numt? I honestly cannot follow their line of arguments given the provided information. Again, maybe I misunderstand what the authors meant to really say, but if this is the case, they have to carefully reword the use of numt expression versus mitochondrial gene expression.

Author's reply:

[Download author's reply (PDF file)](#)

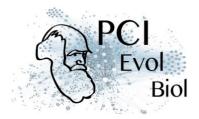
Revision round #1

2018-08-01
Dear authors, Two reviewers have given extensive feedback to your manuscript. Please revise your manuscript according to these comments, questions and suggestions, and add a letter in which you detail your answers to each point raised by both reviewers. Thanks in advance

Preprint DOI: https://doi.org/10.1101/263186

Reviewed by Heiko Vogel, 2018-05-10 13:23

Being a serious agricultural pest, the polyphagous lepidopteran herbivore Spodoptera frugiperda has attracted a lot of attention for the past decades. S. frugiperda consist of two strains, the Sf-R (rice) strain, which is preferentially found on rice and other grasses, and the Sf-R (corn) strain, preferentially found on corn and other crop plants, which could even represent incipient species. Addressing strain differences, it has been shown that pre- and post-zygotic reproductive isolation mechanisms exist between the strains, with viability losses in hybrids as well as differences in reproductive behavior (i.e. timing of mating). Despite the significant number of studies addressing differences between the two strains, it is unclear whether host plant specialization/adaptation has driven strain divergence and might eventually result in different species. Here the authors addressed questions related to genomic (i.e. transcriptomic) plasticity as well as constitutively different transcript levels between strains when exposed to different plant diets.

Comments It is striking that S. frugiperda – depending on the strain tested – tends to lay equal or even higher numbers of egg clutches on the cage net compared to the plants present. While this is also known from other moth species kept in the lab, and might not be that relevant for the major focus of the manuscript, it would still be important to discuss these findings, especially in the light of host plant preferences observed in field situations. Furthermore, what I did find interesting in this context is that the assessment of larval weight gain on the rice, corn and artificial diets showed a clear difference between Sf-C and Sf-R. While both strains performed similarly well on their preferred host plants and the artificial diet (which is assumed to be optimal for both strains?), Sf-R larval performance (measured as larval weight) on the non-preferred host plant corn was higher, while another trait (survival) was lower. Overall I found the larval survival rates to be very low, which is a factor worth discussing. Again, this might not be such a new finding, but it is relevant in the context

of gene expression analysis, since differences in growth rate and potential stress levels should be reflected at the transcript level.

The first paragraph of the section on "Gene expression in RT experiment" is a bit too simplified in my opinion. The authors state that "When confronted with different host plants, polyphagous insects will express a combination of genes that will ensure their optimal fitness .." While this is generally assumed, I am quite certain that broader evidence for this statement is rather lacking. Although herbivorous polyphagous larvae usually respond to different host plants with a plastic transcriptional response, regulating mRNA levels of many genes, it is unclear whether all or most of these responses are directional and the herbivore indeed benefits from such regulation on a specific host plant. Thus, there should be a more balanced discussion of the biological significance and directionality of gene expression on different host plants beyond such a simple statement. Methods, data interpretation & discussion
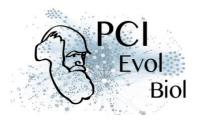
The reaction-norm also depends on the initial genetic polymorphism in response, and this may have been reduced in each of the S. frugiperda strains after (how many) generations on artificial diet. Based on the methods section is not completely clear to me how long Sf-C and Sf-R strains have been reared on artificial diet before being used in the corn and rice host plant experiment. Likewise, the authors state that "We collected 4th instar larvae of the second generation on native and alternative plants.." Does this mean that they only kept these strains in the lab for 2 generations? And were those reared for these two generations on their respective host plants or on artificial diet? And how might this (long or short term) rearing each diet influence the host plant switching and thus the obtained results?

What I also found unclear is the description of samples used for RNAseq, since these seem to not match the respective descriptions in the Methods section. Furthermore, in the Methods section the description of samples used for RNAseq is puzzling and it is not clear from the descriptions provided how many larvae were pooled for each biological replicate in the end. This section should be reorganized to provide a more logical flow and make it clear what material was used for which approach and in what numbers.

If I am not missing something essential, the RNAseq experiment consisted of only two replicates for each strain on the corn diet and only a single replicate for all other treatments & strains. This is certainly not sufficient for performing any meaningful statistical analysis (or any at all) and I would therefore be very careful with respect to data interpretation. This being said, most of the results related to gene expression changes discussed subsequently can only be used to formulate working hypotheses – and do not allow any final data interpretation. One of the figures showing the PCA analysis of each of the Sf-C and Sf-R strains on the three diets seems to show a clustering of the two corn replicates. However, since only a single replicate was used for two of the three diets, in my opinion the PCA results are unreliable. This is even more so since on corn, larvae of both strains performed best (at least when considering developmental time and larval weight), such that one could assume that biological replicates might cluster well. However, the same might not be true for the rice and the artificial diet treatments, since especially on rice general stress levels might be higher and a stronger transcriptional response might be much more variable between replicates. This could only be verified by using biological replicates. Although the authors subsequently used qRT-PCR to verify expression changes of their Top 50 candidate list (why picking the Top 50 and not, say Top 200?), they used the exact same samples (2 replicates each for both strains on corn and a single replicate for the other experiments). So in essence they are verifying that the non-replicated qRT-PCR approach confirms the non-replicated RNAseq results.

For the analysis of enriched GO terms (using Fisher´s exact test) the authors used an FDR-corrected p-Value cut-off <0.1. It would be important to understand the rationale for using such a rather relaxed constraint for the identification of enriched GO terms. Were the results non-significant when using a different cut-off value, such as <0.05?

Regarding the findings related to differential expression of a mitochondrial gene (COIII) derived from a mitochondrial sequence integrated into the nuclear genome (numts), the authors argue that the differential expression ".. comes from polyadenylated RNAs of mitochondrial origin, whose reads also align on the numt region." They further argue that this suggests that "Sf-C has a major difference in energy production at the

mitochondrial level compared to Sf-R". For a number of reasons I find this interpretation of the data a bit premature. First, the differences in COIII transcript levels could also be the results of differences in background transcription of COIII-numt, since in order to detect transcripts even in poly(A)+ enriched mRNA preparations, polyadenylation is not an absolute requirement, as evidenced by the occurrence of substantial levels of contaminations with bacterial (poly(A)-) transcripts under certain circumstances. Second, finding differences in COIII transcript levels does not automatically imply higher energy production levels, since other parts of the machinery should also show elevated expression levels. An alternative explanation to what is proposed by the authors (i.e. that the cytochrome oxidase gene "may be the original target of selection between strains") is that the number of COIII-numts as well as the respective expression of these mitochondrial sequences integrated into the nuclear genome differ between species – but have no impact on mitochondrial energy production.

Broader relevance Although in general the authors are on the right track, they fall a bit short in really addressing the questions formulated in the introduction by not performing truly replicated experiments. Therefore, it was very difficult for me to accept all the arguments on changes in gene enrichment and expression differences, as associated with a selective process of adaptation without backing them up with much more solid, truly replicated gene expression data. These finding may just reflect changes that are not associated with enhanced or reduced performance and might as such not be under selection. The broader relevance of these results to other groups, namely polyphagous herbivores, when switching among food items is not considered in much detail. Would the relatively limited short term changes in gene expression levels observed here be expected in animals that switch much more frequently among host plants, even within generations?

Reviewed by Sabine Haenniger, 2018-05-22 21:41

Download the review (PDF file)

Author's reply:

Dear editor, we wish to apologize for the long time it took us to revise our manuscript. You will find in attachment a point-by-point responses to the reviewers as well as our amended manuscript.

Download author's reply (PDF file)