Round #2

Author's Reply:

Decision

by Astrid Groot, 2019-12-04 11:22
Manuscript: https://doi.org/10.1101/263186

Major revision

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

Dear editor,

please find below our answers to the reviewer's comments. As per your suggestion we also attach a tracked changes document of the main manuscript, since it required some amount of rewriting. We hope this new version will be satisfactory to both reviewers.

Best,

Nicolas Nègre & Marion Orsucci

Reviews

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.

In return, we thank both reviewers for their constructive comments and recommendations that helped improve the paper.

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.

As suggested also by reviewer 2, we rewrote some parts of the paper to make it clearer. In particular, we want to make it clear that numts are not expressed and serve here the purpose of identifying differential expression in small portions of the mitochondrial genome.

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.

Correct, we modified our wording in this section.

line 347ff: The mentioned figure does not exist and I fail to see which of the existing figures this could be.

Sorry for the error in the referencing of the figures. It should be Fig. 4B of course instead of 7B.

This is really a shame, because it does not become clear, how exactly co1 stands out over the rest of the differentially expressed genes.

Fig. 4C orders the sf-R specific genes on the left, by hierarchical clustering. The “gene” corresponding to the longest branch is indeed a numt, so is the “gene” corresponding to the second longest branch. We specified this in the relevant sections.

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.

We added two Supplementary Tables (S4 and S5) with this list of genes and their manual annotation, and a discussion of these genes in the appropriate results section.

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.

It is true. But what is striking to us is that, at the genetic level, the differentiation between sf-R and sf-C occurs mainly at the level of mitochondrial genome (see Fig. 3 of Gouin et al. 2017). In the same paper, the authors, some of which are also part of this present study, say: “The fst of mtDNA is 0.938, while that of nuclear DNA is only 0.019”, which means, almost complete differentiation at the mitochondrial level, but almost none at the nuclear level. Thus, it is striking to us, that, by asking the same question at the transcriptome level, our top candidates also correspond to mitochondrial genes. This suggests, at the very least, a direct effect of mitochondrial genetic variation on its function. We can then safely hypothesize that it must be that mitochondrial function upon which selection acted to produce the two different strains. And since these two strains are phenotypically different mainly with respect to their host-plant, we conclude that there may be a functional link between mitochondrial function and plant adaptation. We would like the community to consider our hypothesis, so we stand by it. However, we reworded the text to make it clear that it is -in our sense- a valid hypothesis but not a definitive conclusion.

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).

We tried to develop a consistent color code throughout the manuscript. It’s a shame if we failed. We went over the figures once more and made them more consistent and we have also tried to use color combinations that are colorblind accessible in most of our figures.

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.

Done.

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.

We apologize for not making our changes clearer. This time, we provide a change tracked version of the manuscript to ease the reviewing process.

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.

As mentioned in our reply to reviewer 1, we really wish to expose the possibility that host-plant associated divergence in this species can be linked to different functioning mitochondria. We realize it is a bold conclusion, so we modified the text to make it clear that it is only a hypothesis based on our current understanding of phenotypic, genomic and transcriptomic divergence between the host-strains of this species. We exposed several possible mechanisms linking mitochondrial function with host-plant association in the discussion section (lines 400 to 424) with relevant literature. One mechanism might be that a reduction in mitochondrial function prevents the sf-C strain to colonize plants with lesser nutritional values such as rice and grass but not plants with higher energetic values such as corn and beans. This is reflected in the wild, since sf-R can frequently be found in corn fields while the converse is less frequent.

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.

In figure S16, we show the level of reads from RNAseq aligning to the COI *numt*. On this figure, we can see the original prediction ‘GSSPFG00006578001’, but we think this prediction is

incorrect and in fact represent a numt - noted in pink - corresponding to a fragment of COI (a simple blastn search confirms it). This *numt* has no feature of a 'regular' gene. It is lacking a well-defined ORF and the reads mapping (in gray) show a high coverage on a short sequence corresponding only to the mitochondrial insertion and not flanking regions.

That said, we still don't understand why only this region of COI is not transcribed (Fig. 4D). But we know that mitochondrial RNAs are heavily post-processed after transcription and this region may be excised. To confirm this hypothesis, we would have to re-sequence several *S. frugiperda* mitochondrial genomes, coupled with their expression to understand the molecular process at play.

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.

We carefully reworded our text so that there is less confusion, even though *numts* are difficult to explain. In clear, our gene annotations contain some *numts*. These *numts* should not be transcribed because they are just mitochondrial fragment insertions. Thus, when we detect differential expression of *numts*, what we are really seeing is differential expression of the mitochondrion. The mitochondrial genome is transcribed as one long RNA (actually, 2, one in each direction) from a single promoter region, usually corresponding as well to the replication origin. It is then post-processed in the different mRNAs. We are certain that the reads alignment come from the mitochondrial RNAs because of their high coverage (several 1,000s rpm; Fig. S16B), on a very short sequence (~150 bp) with sharp boundaries (Fig. S16A). In addition, we detected some SNPs on these *numts*, representing the divergence between our reads and the reference sequence. Those SNPs correspond to the mitochondrial genome sequence, not the *numts* (compare the different SNP composition on Fig. S16A between sf-C and sf-R) while we confirmed by PCR + sequencing that the *numt* sequence is the same on both strains.

We hope, it is now clear in our paper that *numts* are not expressed but only reveal parts of the mitochondrial genomes of both strains that are differentially expressed (Fig. 4D).