

Decision

by Astrid Groot, 2018-08-01 11:55

Manuscript: <https://doi.org/10.1101/263186>

Major revision

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

Dear editor, we wish to apologize for the long time it took us to revise our manuscript. You will find below, in blue, the point by point responses to the reviewers' concerns about our manuscript. One of the biggest weakness of our study is the lack of replication for some RNAseq data conditions, making the interpretation about diet adaptation far reached. We agree with the reviewers that this part needs to be taken with precaution. We thus decided to tone down the biological interpretation of this part and focus the study about the transcriptional difference between the strains. Indeed, our analysis seeking to find constitutive differences between *Spodoptera frugiperda* strains compares 4 datasets versus 4 datasets and is confirmed by qPCR on 50 candidates, by data from other laboratories as well as data from individuals from natural populations. This, we believe, makes our claim about the transcriptional difference of the mitochondrial genome very robust. We hope you will consider the revised version of our manuscript to be recommended by PCI Ecology & Evolution.

Reviews

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-C (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.

In accordance with this remark, we have developed the discussion on this part (line 148):

“This lack of preference for their preferred host plant is surprising because *S. frugiperda* is a species subdivided into two strains according to the host plant on which the individuals were found preferentially (i.e. sf-R on *Oriza sativa*, *Bermuda grass*, *Cynodon* spp. and *Medicago sativa* whereas sf-C consumes mainly *Zea mays*, *Sorghum* spp. and *Gossypium hirsutum*; Pashley 1986). The question of qualifying them as two distinct species as already been raised (Dumas et al. 2015). However, although two variants are defined, *S. frugiperda* is mainly qualified as a polyphagous species found on about 100 different host plants belonging to 27 different families (Pogue 2002). Despite these host plant preferences observed in natural populations, both strains were sampled on the same plants (Juárez et al. 2012). About 19% of sf-R individuals are present on maize and 5% of

sf-M individuals are present on various herbaceous plants (Prowell et al. 2004). This lack of striking female preferences could be accentuated by working on laboratory strains, used for several generations to lay on filter paper”

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 goal of our study was indeed, naively, to correlate phenotypic variations with transcriptional variations. While laboratory strains provide a robust material for standardizing conditions, it is not always clear if phenotypic variations reflect naturally occurring variations. For example, regarding 'survival', we provide in Supplementary data (see below, Figure S3) the survival curves obtained on artificial diet for both strains. Depending on whether we rear them, on the same diet, we can have as low as 10% survival or as high as 76%. We could not assess in this study what is the survival percentage of *Spodoptera frugiperda* natural populations in open fields or whether survival is a function of diet type. That is why in our study, we emphasize solely the difference between strains and diet since they were all cultured in the same way and can control each other. We try to relate the differences observed in survival (mainly of sf-C on 'corn' diet) with transcriptional expression of genes associated to growth and digestion and not to stress.

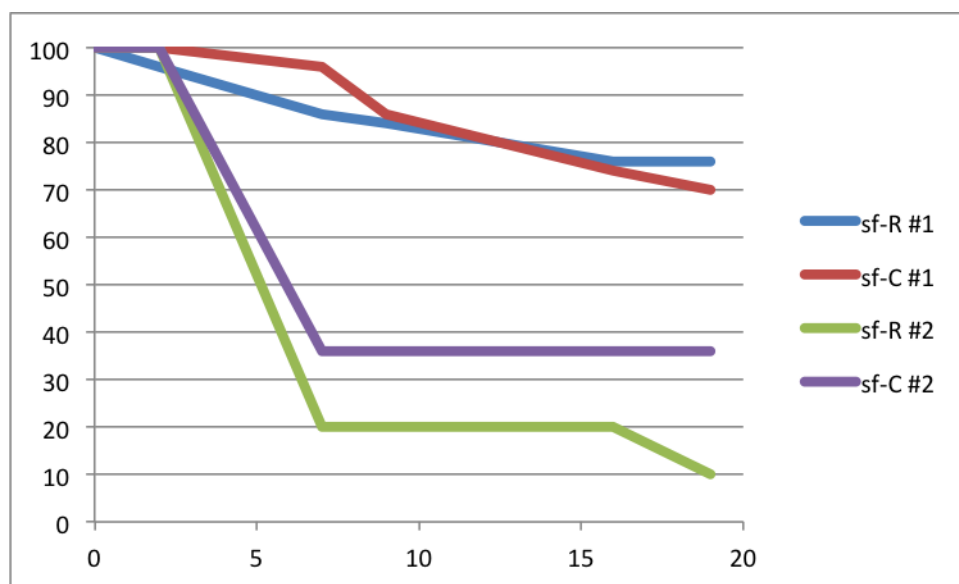


Figure S3: Survival from egg hatching for 50 individuals reared on artificial diet with low (exp. #1) or high (exp. #2) hygrometry.

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.

In the mentioned sentence, we simply wanted to indicate that in response to different environmental stimuli, the set of expressed genes can change and/or be adjusted to permit a better survival of individual. But we agree with the reviewer that it is impossible to know if the responses are directional

and advantageous on the native host. Indeed, our experiment allowed us to quantify gene expression and thus we showed that different genes-set are involved for the two strains and also that their expression changed according the host plant. But we cannot say if this has directly resulted in a change of fitness. The introduction of this paragraph, **line 227**, has been modified to modulate our thought:

" When confronted with different host plants, polyphagous insects will respond by expressing different sets of genes, some of them can be associated to a better adaptation to the host plant. Such adaptation genes in insect are known to be involved in chemosensory, digestion, detoxification and immunity processes among others (Simon et al. 2015; Celorio-Mancera et al. 2016). In order to understand if the two *S. frugiperda* strains express different adaptation genes to host plant diet, we performed RNA-Seq experiments from the larvae of the RT experiments"

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?

We clarified the Methods section. Our laboratory strains have been maintained in captivity and reared on artificial diet from a source population since 2001 for sf-C and 2012 for sf-R (information added line 444):

"We used individuals from the two strains of S. frugiperda: corn (sf-C) and rice strain (sf-R). Those strains were seeded with around 50 pupae sampled in Guadeloupe in 2001 for sf-C and in Florida (Hardee County) in 2012 for sf-R."

They correspond to the now published reference genome (Gouin et al. 2017). From these populations, we took a batch of eggs that we reared on artificial diet in the experimentation platform. Then we subdivided the progeny on the three different diets so that they would have the same ancestral background and thus could be compared. We modified the sentence in Methods section to facilitate the understanding (see line 511):

"From a batch of eggs reared on artificial diet, we subdivided the progeny on the three different diets (corn plant, rice plant and artificial diet). A total of 80 larvae (which hatched the morning of the experiment) were deposited in each cage."

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.

We modified the Methods section to clarify which samples have been sequenced (line 559):

"We collected 4th instar larvae of the second generation on native and alternative plants, corresponding to offspring of the larvae used to estimate the different components of fitness (survival, weight and developmental time)."

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.

Due to problems in the maintenance of the experiment on rice (very low survival), we were only able to sequence one replicate. We fully agree with the reviewer that is insufficient for robust statistical analysis such as linear model to fully characterize the reaction norms.

In contrast, the PCA permit to have a global idea of the relative positions of each samples. As we have few samples no statistical analyzes could be conduct on the PCA statistics.

But we agree with the reviewer that the part of our study describing transcriptional plasticity in function of the diet might not be fully supported. That is why we initially presented interpretation on the Top 50 candidates for each condition, because at this level of fold change, this is a conservative number of candidates and limits the false positive. Note, however, that DESeq2 can be used with low number of replicates although the analysis is not statistically robust, we have exploited our dataset to highlight some genes that may have a role in the specialization to the host plant. We are aware that these genes will have to be confirmed by other studies (silencing, DE in natural populations, etc...). That said, we decided to tone down the biological interpretation of this part, remove the results and conclusions regarding the adaptation to plants and instead focused the study on the transcriptional differences between the strains. Indeed, our analysis seeking to find constitutive differences between *Spodoptera frugiperda* strains compares 4 datasets versus 4 datasets and is confirmed by data from other laboratories as well as in natural populations and is confirmed by qPCR. Having a different diet parameters in those RNA-Seq actually help finding constitutive gene expression differences. This, we believe, makes our claim about the transcriptional difference of the mitochondrial genome very robust.

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?

Sorry for the typo, it was indeed 0.01.

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.

We modified the text to make this claim more clear. First, we emphasize that numts are not supposed to be transcribed (see <https://en.wikipedia.org/wiki/NUMT>). Indeed, they are small DNA fragment of mitochondrial origin but they lack promoter or regulatory elements. It is the case for COIII, RNAseq reads alignment do not overlap with the surrounding genomic context (see SFig. 17A). Since numts are not transcribed, these detected alignments must come from corresponding part of the

mitochondrial genome. The mitochondrial genome is transcribed by a mitochondrial RNA polymerase (coded by the nuclear genome and consisting of only one subunit - there is a gene in *S. frugiperda* corresponding to mtRNAP: GSSPFG00025242001-RA) from one single promoter on the mt genome, overlapping with the origin of replication. The mtRNAP can transcribe two DNA strands corresponding to both directions of transcription. These 2 preRNAs are then processed to produce the mt-tRNA and mt-mRNA that will be translated in the cytoplasm or the mitochondrial matrix. Many of these mt-mRNAs are polyadenylated and thus can be found enriched in our RNAseq experiments, as any other mRNA. That is why we realigned our RNAseq to the mt genome. This confirmed that indeed Col and ColIII were differentially expressed between the two strains.

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.

As stated previously, numts are not transcribed, and are not supposed to have a function other than any other type of sequence insertion. We agree with the reviewer that this differential expression of COI may be compensated by other parts of the machinery, but it is a hypothesis that reducing levels of COI might decrease (but to what extent, we do not know) the ATP synthesis yield. It is natural to ask what could be the phenotypic impact of mitochondrial transcription differences, but we make it clear in the text that it is only a hypothesis, not tested in the present study.

We also make clearer, we hope, our rationale for the mitochondrial genome being the main target of selection between the two strains. In our previous study on genomic differences between the strains, we show that, by resequencing natural populations, the mt genome have an F_{st} of 0.938 and cluster on a phylogenetic tree as two sister species (as confirmed by many other studies using Col -or other mt genes - as a marker) - see Fig. 3b of Gouin et al. 2017. The nuclear genome is much less differentiated, with an F_{st} of 0.019 between the strains. 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 mitochondria, 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.

Gouin, A., A. Bretaudeau, K. Nam, S. Gimenez, J.-M. Aury, B. Duvic, F. Hilliou, et al. 2017. “Two Genomes of Highly Polyphagous Lepidopteran Pests (*Spodoptera frugiperda*, Noctuidae) with Different Host-Plant Ranges.” *Scientific Reports* 7 (1). doi:10.1038/s41598-017-10461-4.

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 findings 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?

By removing all the interpretation of data on plant adaptation, we indeed can not conclude on transcriptional plasticity of polyphagous insects. We hope we convinced the reviewer that refocusing the paper about differences between strains provides a more solid ground from our dataset and that the transcriptional resources and results presented are worth being published.

The study investigates host-plant related differences between the two strains of *S. frugiperda* in

- a) oviposition preferences
- b) larval performance
- c) gene expression (RNAseq) on different plant diets

Overall, the topic of this study is very interesting. However, I do have some major concerns as stated below, that need to be addressed by the authors before I could recommend this manuscript for publication.

Results and Discussion:

Oviposition trials: Was the exact position of egg masses recorded, i.e. did the females sometimes lay eggs close to the plants but on the net and sometimes really far away from the plants? Maybe the rice strain is not so “dumb” if the egg masses in the vicinity of the plants were counted separately? In nature there is no net close to the plants: if plant volatile cues advertise oviposition sites, there are oviposition sites “everywhere” and it wouldn’t matter too much to lay egg masses a bit next to a specific plant. I’m missing an interpretation of the data, why do they lay on the net? Were they in the lab for too long, so a preference is lost because it doesn’t give an advantage and is not selected for? Is oviposition triggered by the volatiles and then the female lays the eggs just anywhere? It’s a generalist after all.

Unfortunately, we did not record the exact position of each egg masses but they were distributed over the entire cage without a preferred pattern (i.e. some were on the top, others on the floor and on the sides of the cages). Thus, the rice strain seems really not to choose to lay on the different plants because the egg masses were clearly not found close to the plants

Moreover, I don’t think that the egg laying is triggered by the presence of volatile plant compounds because females lay as many eggs in population cages in the laboratory on filter paper. In addition to that, I think it is indeed possible that the strains may have lost the ability to discriminate the plant volatile cues because the selection on this trait has been relaxed by keeping them in the laboratory.

Otherwise, we agree with the reviewer about the fact that *S. frugiperda*, although known for having two host strains associated with different ranges of plants, is a generalist species. In any case, laying on a substrate that is not a plant can be considered as a “dead end” because the mobility of the larvae is limited and therefore by laying on the net causes a decrease of fitness due to a higher mortality of the offsprings.

To explain these points, we added a discussion point, line 148)

Performance trial: Figure 2 A+B no legend, so can’t be interpreted. No error bars for the weight. Data representation makes it difficult to compare between the strains (as they are shown in different graphs). Axes not labelled. Please revise this figure completely.

We did the changing as recommended Figure 2.

We have also replaced the Figure 1 by Fig. S1 on which the dispersion measurements are presented allowing a better understanding of the differences between the two strains in term of global fertility (Fig 1A), and the similarity of both strains concerning the choice of laying sites (Fig 1B-C).

Performance trial: How did the rice strain survive on artificial diet? (or both strains, for that matter). That should be the baseline. Else one can’t compare the strain’s survival on corn or rice and state that one is doing better than the other. To be clear: If the rice strain survival on artificial diet in that generation would have been 5%, a 10% of survival on corn would actually be outstanding. But we can’t know, because this info is not given. The basic principal of having a control in scientific studies...

We do not think the artificial diet should be considered a “control” but rather a reference point. The artificial diet is not found in the wild and has been optimized for providing maximum health to lab population. Of course we were expecting whole plant to have lower survival rates, probably recapitulating natural conditions as well. In fact we tried to measure survival on natural diets. Depending on rearing conditions, survival could vary from 5% in conditions close to the plant experiments to 80% if we try to really care for each larva individually.

Gene expression: “extracted RNA from larvae of the RT experiments” – There is no data other than the RNAseq data from these larvae, right? So it’s not larvae from the RT experiments, it’s larvae that were fed on different plants. This sentence gives the illusion that the RNAseq samples are from the SAME individuals that performed on the plants in the previous RT experiment, but that’s not the case.

The RNA-seq data came from the larvae from the RT experiment. The life history traits were measured on the individuals of the first generation on the plants and the sequenced individuals came from the second generation on plants. In other words, the extraction and sequencing was done on the offsprings of the measured individuals.

Our explanations were not clear in the first version. Thus, we modified the main text line 559:

"We collected 4th instar larvae of the second generation on native and alternative plants, corresponding to offsprings of the larvae used to estimate the different components of fitness (survival, weight and developmental time)"

Gene expression: "we could perform for each strain two replicates on the corn diet but only one replicate for the rice diet and the artificial diet" – one replicate? (This does not become clear in M&M) This would be nice for a preliminary experiment but not for a study you want to publish.

We are aware of the problems associated to have only one replicate. Unfortunately, the experience is difficult to perform due to the low survival on rice for both strains.

It is obvious that our results on plant adaptation are not supported by statistical models but the fact of researching whether the most DE genes found in the RNA-seq on laboratory strains are also differentially expressed in the individuals from the field allows us to cross presumptions and to validate these genes as being involved in the specialization of variants to different host plants. Which is why we refocused our manuscript around this one question.

Material and Methods:

Laboratory strains: The strains used for the preference-performance assays are laboratory strains, in the lab for 4 or 10 years, originating from rather small populations (30 or 50 individuals) from different geographic regions. No details are given if the seeding pupae were collected from the field or obtained from a different lab (more likely when it is referred to pupae), where they had been in rearing on artificial diet for even longer. The absence of plant cues during rearing in the lab means relaxed selection on host-plant related (preference-performance) traits. Please discuss how the generations in the lab could influence your results.

From our reference genome paper (Gouin et al. 2017) Supplementary Information:

"The *S. frugiperda* laboratory strains have been seeded with 30 to 50 pupae in 2000 and 2010 for the corn and rice strain respectively. Since then they were reared in laboratory conditions (on an artificial diet [Poitout 1974], at 24°C with a 16:8 photoperiod and hygrometry of 40%). The individuals that seeded the corn strain came from Guadeloupe whereas those that seeded the rice strain came from Florida (Gift of Dr Meagher)."

We added these informations about the strains in Methods session line 443:

*"We used individuals from the two strains of *S. frugiperda*: corn (sf-C) and rice strain (sf-R). Those strains were seeded with around 50 pupae sampled in Guadeloupe in 2001 for sf-C and in Florida (Hardee County) in 2012 for sf-R."*

We agree with this point, it is possible that the selection is relaxed for the developmental and preferences traits. We added the following discussion point regarding LHT measurements:

- line 148: *"This lack of striking female preferences could be accentuated by working on laboratory strains, used for several generations to lay on filter paper"*

- line 219: *"Once again, as noted in the plant preference, the absence of plant cues during laboratory breeding over several generations could have allowed a relaxed selection of host plant characteristics. Moreover, the artificial diet is based on corn flour and therefore Sf-R has not been confronted with rice compounds for many years. Sf-R has therefore been able to adapt to certain compounds of corn explaining that differences between these two plants are not detected"*

As for the RNA-Seq experiment, that is also why we went even further and combined data from controlled conditions as well as natural populations. We actually discuss this point by saying in the last paragraph (line 429)

"Traditionally, two different RT strategies were used, either by using colonies from natural populations or long maintained laboratory colonies and each approach has its pros and cons. Working with laboratory colonies allows one to control for genetic background variations as well as environmental conditions. But in turn, they might be subject to genetic drift or adaptation to

the artificial diet used to maintain them. Here, we show that by combining the two approaches, we revealed a smaller set of genetic events that could explain the differentiation of the two strains."

Laboratory strains: Also, no information is given about how the strain identity was diagnosed when rearing started. Was there a test for hybridization? The only genotyping I can see was done on the mitochondria in the field collected samples. If the lab strains are also only discriminated based on the mitochondria, in the worst case the lab strains could theoretically both be hybrid populations that only differ in the mitochondria (but are wildly mixed for chromosomal markers), then maybe it's no wonder you find the main differences in the mitochondria.

Upon acquisition (see above) the strain identity was confirmed by both Col and FR repeat diagnostic markers. This diagnosis has been repeated when we sequenced the genome of both strains (Gouin et al. 2017). In the present manuscript, we again repeated the diagnosis of lab strains with not only COI and FR1 but also Tpi. The results can be visualized in Fig. S8A, S9A and S10A.

Also, the genome sequence paper of these two strains indicates notable genome-wide differences between the strains and the laboratory strains cluster with natural populations of the same strain (Fig. 3b in Gouin et al. 2017)

Age: No information is given on the age of the moths released in the oviposition cages. The age does have an influence on the mating and thus oviposition behavior, e.g. females don't mate in their first night after emergence and thus very young females will produce less eggs in 3 consecutive nights because they won't have mated in the first night. Please give this information.

We added this information line 469: " *All individuals released had emerged the night before the oviposition choice experiment.* "

In brief, we released moths that have hatched during the night before the start of the experiment. By consequence, all the individuals were the same age (the individuals had a maximum difference of 12 hours). Moreover, since we used pools containing at least 12 females released during 72 hours, the differences was minimized and as the number of individuals affected strongly the number of egg masses laid, we have conducted the statistical model using as the response variable " mean Fecundity" which is the number of egg masses divided by the number of females in the replicate.

By consequence, as all females are of an equivalent age, there is no bias introduced in our data set.

Climatic conditions oviposition: 4 replicates of each set-up were done under the same climatic conditions – what are these conditions?

This information has been added to the Material and Methods section (line 476):

"4 replicates of each set-up were done under the same climatic conditions, within the quarantine platform (22°C, 50% humidity, natural dark-light conditions - in November around 14h dark:10h light- with fluorescent light bulbs)."

Egg masses: In our own FAW rearing on artificial diet as well as in oviposition experiments on plants I noticed a huge variation in the number of eggs per egg mass. The number and size of egg masses can differ from female to female, depending e.g. on if she was disturbed when laying eggs and had to start again. Also, corn leaves provide a large area for oviposition and can support one egg mass that includes all eggs of one night, while rice or bermudagrass plants may call for multiple egg masses by the same female in the same night to fit all eggs in. I'm concerned that looking at the level of egg masses introduces a bias that is difficult to control for, as opposed to e.g. looking at egg numbers, especially when addressing the hatching proportion. An egg mass giving rise to 1 larva is counted the same as an egg mass giving rise to 100 larvae, yet this would be a dramatic difference in terms of fecundity and fitness.

Thanks for this comment. Indeed, the number of egg by egg masses can change a lot the results in term of fecundity and fitness. But our objective here was to record the laying preference in both strains, since *S. frugiperda* does not lay individual eggs but rather egg masses, we think that the number of egg masses is a sufficient measure because it allows to highlight the plants to which females are attracted. In this study as the number of egg masses laid on the plants was restricted, we do not think it has much influence on our interpretation.

Note that in Figure S1 B and C we counted the percentage of individual eggs giving rise to a live larva in each trial.

Performance assays replicates: While I acknowledge the amount of work that went into the two replicates of 4 modalities (= total 8 cages with 80 larvae each), I still think at least 3 replicates would be needed. Why did you not use 50 larvae in 3 replicates each? Were there room constraints, maybe? Please explain.

We agree that it is better to limit the number of individuals by cage and increase the number of replicates. Our problem was the space available. We had only one growth chamber. If we realized a replicate at different time, to increase the number of replicates we would have been confronted with difficulties of analysis due to temporal differences.

Performance assays: No details are given for the control on artificial diet. (Oh, it is, but not in M&M, in results, and with little detail).

We added the following information in the M&M section, line 536:

" In parallel, and as a reference point, we performed the same experimental design and measurements on standard rearing conditions on artificial diet (Poitout and Bues 1974). Two replicates of each strain on artificial diet have been set-up from the same batch of L1 larvae from our laboratory strains. Compared to plant conditions, rearing has been performed in a square plastic box with mesh filter for aeration and food supplied ad libitum. Since the rearing conditions differ significantly from plant assays, we considered those experiments as reference and not as control."

Performance assays: poorly described. E.g.: how often did you weigh the larvae?

We described more the performance assays (line 530).

To answer about the weight, at the beginning, we pooled the 80 individuals because individual larvae were too light to be detected by the scale. When the larvae reach the second larval stage, most of them had reached a weight sufficient to be individually measured. Thus, from the second larval stage, the weight was assessed every two days when the plants were replaced in the cages.

Sample preparation RNAseq: It does not become clear, how many larvae were in the end pooled for each treatment and biological replicate. Also, the subpools "having a good quality and quantity" were pooled – what is a good quantity and quality? That would be very valuable information to evaluate your data as well as for researchers planning similar experiments.

We clarified and detailed the protocol from line 570:

The samples from rice plant contained one sub-pool (i.e. 3 larvae) for all the other samples, 4 sub-pool containing 3 larvae were pooled after quality and quantity control, representing at the end 12 larvae by sample.

The quantity required for sequencing was 200 ng/μl but for each subpool, containing three individuals, we had a quantity between 1058 and 4225 ng/μl.

Concerning the RNA-quality, we used the 260/280 and 260/230 ratio, which measure the nucleic acid purity. Expected values are commonly around 2. If the ratio is appreciably lower than expected, it may indicate the presence of contaminants. In our samples the values varied from 1.35 to 1.98)

Sample preparation: Are the used insects offspring from the survivors of the performance assays or from the lab colony reared on artificial diet?

As mentioned above, the sequenced individuals come from survivors from the first generation of plants of RT. The appropriate information has been added in the text, line 559.

Sequencing: I'm not sure why the chance to create a decent number of biological replicates for the RNA seq was missed. The larvae were sub-pooled in multiple pools of 3 initially, and then combined again to make 2 biological replicates. I consider 3 biological replicates the lower limit for any scientific experiment. It is also not necessary to "obtain samples corresponding to the 2 biologic[al] replicates of the 4 experimental set-ups [of the performance experiment], as it's other larvae and the performance results can't directly be linked to the RNAseq results anyway (because it's different larvae, different plants, maybe adults reared on plants?, next generation, ...). This needs to be explained.

As stated in the response to comments from previous referee, we acknowledge that missed chance to link gene expression variations with phenotypical traits. Ideally -and it was indeed our original intention- larvae from the same experimental set-up should have underwent both phenotypical measurements and gene expression measurements.

Clearly, the reviewers, and more likely the readers will find the experimental design presented in this manuscript, unable to provide rigorous and robust results regarding the specialization to plants in *S. frugiperda*. In response, you will see that we completely modified the manuscript that now revolves

around only one question : the difference between strains. To answer this question, we think we have enough data (4 RNA-Seq for sf-C and 4 RNA-Seq for sf-R) for the analysis to be statistically robust, which has been verified by independent qPCR on 50 genes and cross referenced with RNA-Seq data from natural populations as well as with RNA-Seq data from another laboratory (Roy, A., W. B. Walker, H. Vogel, S. Chattington, M. C. Larsson, P. Anderson, D. G. Heckel, and F. Schlyter. 2016. "Diet Dependent Metabolic Responses in Three Generalist Insect Herbivores *Spodoptera* Spp." *Insect Biochemistry and Molecular Biology* 71: 91–105. <https://doi.org/10.1016/j.ibmb.2016.02.006>). We hope that the reviewer will now agree that the results answer our question.

Natural populations collections: Please give the numbers of insects that were "sacrificed" (that's a weird word there) directly after collection from the field vs. the number that had to be reared on plant parts until 4th instar. I have a major concern with this: whole living plants are certainly defended differently than cut leaves. So differences you see between field samples and laboratory samples reared on host plants and alternative plants may be due to growing on defended plants vs undefended cut leaves. You don't control for that. What are the numbers of collected insects? How was the ratio corn-strain/rice- strain? How many were used in the end? DNA/RNA extractions (field collections): were these samples pooled? Not described at all.

In the end, only larvae killed immediately after collection were used for sequencing. Supp. Figure 8D shows the ratio of sf-C and sf-R per collection site. Tpi and FR1 genotyping protocols details were indeed missing. They have now been added in the methods section I.636

In the analysis we performed we never compared gene expression differences between lab conditions and field collection because the experimental conditions were just too different. Rather we combined the data to address only one question: "regardless of diet and or condition, what are the constitutive differences, at the expression level, between sf-C and sf-R"?

Genotyping: please give the primer names according to the referenced paper. Also, the Msp1 digest is the digest that is more difficult to interpret as the 50bp fragment is difficult to differentiate from left- over primers on the gel and a 50bp difference between sf-C and sf-R band can be difficult to see depending on the gel quality; usually both Sac1 and Msp1 digest are used. Please explain why you used only one (and the one that's more difficult to interpret).

We added in the Methods section more details about the primers used with respect to the cited papers I.636. Supplementary Figure 8 shows without ambiguity that our test with Msp1 is easy to interpret, especially with the use of an undigested control, as shown on the gels.

qPCR: primers? amplicon size? Why was the lab sf-C strain used as reference point?

We have added a Supplementary Table 3 containing the qPCR primers used and the amplicon length of around 50 nucleotides for SYBR green based quantitative PCR. As for the use of a reference point, our laboratory sf-C strain is the one we use for many other transcriptomic experiments in the lab and it was more convenient to always use it as a reference point. Of course the $\Delta\Delta C_t$ method is relative measurement and the choice of reference doesn't change the extent of variation, only the direction.

Author contribution:

JPB and MV produced the corn and rice plants used in the RT experiments – is this enough for an authorship? These were not modified plants, at least that wouldn't be transparent.

We thank the reviewer for the insight on authorship. In this case, producing a large scale, standardized production of both rice and corn plants corresponding to the need of our study (large amount, optimal density, organic, healthy plants of standardized age etc.) definitely required an important expertise and contribution to the project. The other authors agree that they are entitled to authorship. As for all authors JPB and MV read and approved the manuscript and are willing to accept responsibility and accountability of the work submitted.