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.

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.

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…

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.

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.

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.

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.

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 mergence 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.

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

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.

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.

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).

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

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.

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

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.

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 4^{th} 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.

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).

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

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.