This study by Bacot, Haberkorn *et al.* focuses on a genomic duplication encompassing multiple P450 genes suspected to be associated with insecticide resistance in French Guinana *Aedes aegypti* mosquitoes. The authors used a wide array of approaches (bio-assays, molecular quantification through ddPCR, RNA interference, experimental evolution and both long and short-read sequencing) to assess the resistance levels associated with this duplication, its fitness cost, and uncover its genomic architecture. This study is thorough in its investigation and highlights the challenges faced when identifying new copy number variants associated with xenobiotic resistance, most of all in a genomic environment as challenging as the *A. aegypti* genome.

Overall, this article is well written and concise despite the breadth of analyses performed, but in our opinion a few key points need to be addressed carefully.
The results would greatly benefit from few clarifications, the framework imposed by the impossibility to study the effect of the massive duplication independently from the 1011Met allele really complexifies the understanding of the analyses performed here. Because of various inconsistencies across the analyses we have the feeling that we were not able to fully assess the quality of the study. For instance some F1 crosses are reported in Figure 4 quite late in the manuscript but not in the first paragraph that refers to supplementary file 1. Why couldn't the authors use the results of the first analyses? The authors define the resistance as semi-dominant while the mortality in F1 crosses in Figure 4 is above 80%. More surprisingly the mortality in the F2 crosses is also close to 80% when one could have expected it to decrease because of the presence of resistant homozygous genotypes. These kinds of discrepancies occur in several instances in the text (see below about Fig. 1) and would deserve to be discussed to help the reader understand why it could be expected.

Another major concern comes from the fact that the writing implicitly suggests that the massive duplication encompassing the CYP genes provide resistance while, and as nicely acknowledged by the authors, they obtained contrasted results regarding this point. For instance, the analyses presented in figure 1 are pretty convincing about it (but see major comment below) but there is no further support for it in the other analyses. In figure 4, the slight difference in copy number between F2D (dead, 4) and F2S (survival, 3.3) let's suppose that the dead ones were globally heterozygous for the duplication (mean number of copies 3.5 in the F1 crosses). How could the authors then rule out that the survivors are simply mostly homozygous for the Met too? Here again, the frequency of the Met increases between the

F2D and the F2S. In absence of individual genotyping showing that the Met and the duplication segregated independently (despite the fact that they are on different chromosomes) it is hard to discriminate the respective effects.

The order of the analyses and the choice of sub-heading does not help either and we suggest that the authors revise the structure of the results; we list below the sub headings in order of appearance in main text:

- Deltamethrin resistance is associated with P450 activity
- The resistance phenotype is autosomal and semi-dominant
- The resistance phenotype is associated with a duplication affecting multiple P450s
- Genomic architecture of the duplication
- The P450 duplication is associated with deltamethrin survival
- Multiple P450s carried by the duplication may contribute to deltamethrin survival
- The duplication is hardly retained by selection in presence of the 1011Met kdr allele

Additional comment:

- Providing a figure early in the main text presenting the different lines and crossing design used in the different analyses would make the results easier to follow.
- Figure 1 is essential in showing the implication of P450s duplication in resistance. It is the only clear support for it through the whole study. However, there seems to be a strong discrepancy between the IR13 resistance status indicated in table 1 and the results shown in fig. 1. IR13 is described as providing low resistance (ca. 95% mortality rate) in the table but is later on referred to as susceptible in the main text. Surprisingly, the same line, IR13,displays a ~45% mortality rate without PBO in fig. 1, despite a lower exposure to the insecticide. Even if it is not statistically significant, IR13 shows a reduced resistance when exposed to PBO too. Could the authors specify early in the manuscript the genotype of the various strains/lines regarding the presence of the duplication too? Could the authors consider moving Fig. 1 and the corresponding section after figure 2?
- In fig. 2, the presence of the fifth overtranscribed P450 gene is scantly discussed. Was the position of this gene investigated after the *de novo* assembly? Would it be possible that this gene is mis-positionned and is in fact a part of the duplication encompassing the four other CYP genes? Knowing that this gene show both a higher transcription

level and an increase in gene dosage, how could the author exclude its effect on the resistance phenotype.

- The authors report that the two resistant lines show a ca. 2 - fold increase in DOC for the duplicated region but these data are not shown. The heatmap let room for interpretation for the actual change in relative DOC. Would it be possible that the IR13 line is in fact heterozygous for the duplication ? Could the authors provide a table with the actual normalised relative depth for each of the CYP trapped into the duplication to the depth of the corresponding CYP for both "susceptible strains".

Minor commentaries:

- Is it common (or even meaningful) to list second co-equal contributions, or is it a typo and these authors are all first -co-equal contributions? We have the feeling that such a level of details in the authorship does not match with the publishing model put forward by PCI, but we'll let the recommender decide upon that.
- L154 IR13 is either called susceptible or slightly resistant (see Tab. 1) please harmonise.
- L154 "Among the 11078 genes detected", what is meant by "detected"? Does it mean DE genes in total or total number of genes with enough transcript sequenced in all the lines? Also please provide the total number of genes in *A. aegypti* to give a reference frame.
- L 195 "Indeed, both short read and long read data revealed that this Ile1011Met mutation …" This is the first instance where short reads are mentioned, the results would gain in clarity if a little bit more of methods was added here, or consider moving the methods before the result section.
- L 232 "(Fisher test p value = 0.035)". This is surprising, please add the value of the statistics and the degrees of freedom for all the tests performed.
- L 238 replace "segregated with a high dose" with "exposed to a high dose"
- L. 244 and 249: please provide statistics on the copy number differences and Met allele frequencies between survivors and dead individuals.
- L. 254 please define N≥5
- Figure 4: The distribution of error bars is surprising given the difference in sample size (113 vs 15), please check.

It would be nice to have direct access to the SRA repository through the use of Hyperlink.