The authors here used sequence data from multiple populations and/or species of *Heliconius* to analyze how the presence of a polymorphic supergene affects the demography and diversity of the population. The supergene is responsible for the pattern on the butterfly wing, which is a trait under negative frequency-dependent selection. The species carrying this supergene is characterized by a lack of population structure and a higher genetic diversity. Overall, the results obtained do not contradict the idea that the adaptive introgression of a supergene leads to negative frequency dependent selective selection for wing pattern, the phenotype controlled by the introgressed supergene.

Overall, the paper is quite clear and well written.

Major comments:

- The result section is extremely short (34 lines!).

- Both the Fst (and pi and dxy) calculations and the admixture seem to include the whole genome, and in particular the inverted region and chromosome 17 in general. Given the selective pressure on this region, I wonder whether it should not be excluded or analyzed separately. In particular, inversion may accumulate mutations faster than the rest of the genome, therefore biasing some of the measurements.

- I am concerned about the choice of individuals for the G-PhoCs analysis. Indeed, the authors claimed in the method that "we selected two individuals per taxon or population, retaining those with the highest sequencing depth (see TableS3)." Yet based on table S3 provided in bioRxiv, this is not the case. There are 9 individuals from the *Numata* taxa, in the "amazonian" subgroup that have higher depth than the first individual picked, and 10 than the second individual picked. In addition, picking individual with the highest coverage is likely to create a bias towards population that have more samples (assuming everything else equal). A higher coverage means a better chance to detect the polymorphism and correctly call SNPs (see Jiang et al. BMC 2019 for example). Indeed, when looking at the correlation between the mean sequencing depth of the 2 individuals per species and the final estimates of Ne as reported on Fig3, I obtained a correlation of 0.8785. I wonder whether the authors could pick individuals to minimize the variance in coverage across species. Alternatively, given how many samples (especially for *numata*) are available, the authors could estimate how sensitive to resampling (for a given sequencing depth or when varying it) the Ne estimates are.

- For Figure 3A, some populations for Numata (Alto-Mayo, Pongo and Venezuela) are not displayed (without this being mentioned in the methods). They all have single individual, yet based on Table S3, they were used for admixture and PCA analysis, and two of them were used to calculate Fst. If having only one individual is not good enough to measure pi, then it should also not be good enough for Fst.

- Figure 3B, I understand trying to maintain a comprehensible figures, but I think it would be interesting to have the complete figure with migration as a supplement. Gene flow plays a key role (as pointed out by the authors l. 249) yet there is no quantitative information in the main manuscript at all about it. In addition, Table S6 is rather difficult to read, with no indication of which estimates are considered significant. I believe that

the results should be integrated in the main manuscript, especially given the current shortness of the result section.

- I wonder if this lack of isolation by distance has been observed in other systems with negative frequency dependent selection. Finding other examples would strengthen the results found here.

Other comments:
For the PCA analysis, the first component captures 9% of the variance and corresponds to the difference between the Atlantic and Amazonian pop. The second component captures 6% of the variance and correspond to the difference between Guiana and the other Amazonian populations. Yet this difference does not appear in the Admixture nor the Fst analysis. I wish this difference between analysis was further discussed.

L123-124: Reference?

L126: I would be cautious about the use of "adaptive introgression of a balanced polymorphism", since there is no evidence that at the time of introgression, the new arrangement was already under frequency-dependent selection. Mate choice could have evolved afterwards.

L130-132. I do not understand this sentence.

L176-178: How were the 15 *numata* individuals from Peru chosen?

L180: how were the SNP chosen?

L181: this is the wrong citation- the cross validation is presented in the 2011 paper.

L206: is there a particular reason for this choice of 30kb? Does it correspond to a LD decay threshold?

L208: I am not sure exactly what migration bands are supposed to be.

L271-273: I am not sure what the authors referring too here. Based on Figure 3A, pi varies by an order of magnitude.