November 2022

Dear Editor,

We thank you as well as all three reviewers for positive and constructive comments. We have attempted to adress all comments and concerns in the most constructive ways. In particular this revised version now includes demographic inferences to explicitly estimate changes in effective population size and gene flow between the different populations.
Moreover we have also developed a model to test whether disassortative mating could lead to an increase in genetic diversity beyond the supergene and compared our results with those observed under assortative mating (the classical case in most *Heliconius* species with local adapation) and a null model of random mating. We hope our results now better address the hypothesis being tested throughout this manuscript.
High resolution figures as well as intermediary data are available at zenodo at 10.5281/zenodo.7319913
We also apologize for the delay,

Best regards,
Quentin Rougemont,
On the behalf of all the authors.

Editor's comments:
Revision required

This is an exciting paper that links mimicry polymorphism to broad genome-wide population genetic parameters. However, all three reviewers raise concerns about the manuscript as it stands. The general issue is that with a sample size of 1 there is only rather tentative evidence to link the increase in Ne with supergene formation. This is clearly acknowledged in the Discussion section but not really reflected in the title and abstract and general framing of the paper. One reviewer suggests additional analyses to try and infer the timing of Ne changes relative to supergene introgression, which would provide further evidence for a causal link. An alternative approach might be to use simulations to estimate the increase in Ne expected for a given level of disassortative mating and balancing selection (these parameters are presumably reasonably well known in H. numata) - comparison of empirical and theoretical expectations might help support the hypothesis.

It is worth noting that H. melpomene shows a similar difference between Amazonian and Atlantic forest populations, with Atlantic populations far less diverse (4 H. m. nanna samples were published here: Belleghem, S. M. V. et al. Patterns of Z chromosome divergence among
Heliconius species highlight the importance of historical demography. Molecular Ecology 27, 3852–3872 (2018).). This somewhat undermines the argument that this difference is due to the supergene in H. numata, but analysis of the melpomene population data would provide a contrast that may support the proposed hypothesis if a much greater Amazon/Atlantic difference is seen in H. numata.

Overall there are also a large number of smaller comments that need addressing.
Regarding the broad conclusions the supergene hypothesis either needs to be reduced in prominence through the text, or some additional analyses conducted to further support the hypothesis (the latter would be much preferable).

Reply to Editor:

Dear Editor, thank you for your summary and comments. We have indeed performed additional analyses in which we try to infer *Ne* changes as well as timing of population splits and gene flow using mutliple individuals rather than one. We also have used simulations as suggested to further support our working hypotheis. Finally, we have reformulated that title and abstract as well as some place in the discussion regarding our board conclusion. We are confident that altogether our new analysis and change in writing should lead to an improved manuscript.

Reply to reviewer

Reviewed by Christelle Fraïsse, 23 Dec 2021 18:49/

In this manuscript, de Cara and collaborators investigate the genome-wide effect of a supergene controlling wing patterns in Heliconius butterflies. Based on whole-genome resequencing, they showed that the Amazonian populations of H. numata are more diverse and less structured than all other taxa they investigated. These populations are also the only ones polymorphic for the supergene. The authors, therefore, hypothesize that disassortative mating following the onset of polymorphism through adaptive introgression is responsible for the enhanced diversity observed in H. numata Amazonian populations.

The main results that adaptive introgression can affect population demography (i.e. gene flow and effective sizes) due to change in the mating system is appealing. And I agree with the authors that few studies carefully investigated this effect, making the present work valuable. That being said, I am not entirely convinced that the authors have clearly demonstrated the connection between polymorphism at the supergene and enhanced diversity. Mainly, I think that the demographic analyses should be strengthened. Please, see the detailed comments hereafter.

R: Thank you for all the constructive comments.

1. Demographic inferences – alternatives to G-PhoCS:

The main results of this work are based on the comparison of closely-related populations differing at a trait affecting genome-wide diversity, coupled with knowledge of when the differences evolved

(L70-71). More precisely, the supergene formation should precede the change in demography. Therefore, it is essential to carefully estimate the timing of population size (Ne) and gene flow (M) changes. From this point of view, G-PhoCS does not seem to be the most appropriate method (see below), so I think a different type of demographic inferences should complement it. First, G-PhoCS assumes that Ne is constant along branches of the phylogeny, and so Ne can change only when the ancestral population diverge. Moreover, the method cannot capture bottlenecks or size expansions on individual branches. The changes in population sizes (and their timing) should be tested explicitly in the present study.

R: Changes in population sizes are now tested explicitly in a more appropriate framework with a different method, as suggested by the reviewer (see below).

Second, in the original G-PhoCS paper (Gronau et al. 2011), the authors tested the ability of their method to estimate Ne and M accurately based on simulations. They found that the method has limited power given the features of their data. This should encourage the authors to perform similar validation analyses in the present study.

The third limit of G-PhoCS is its computational burden which led the authors only to use two diploid genomes per taxa.

I understand that the advantage of G-PhoCS is to reconstruct the history of many species along a phylogeny, where other demographic inference methods are mainly applied to 2-population comparisons. Considering closely-related lineages in the inference is particularly important if

genetic exchanges are pervasive among the taxa. Yet this is not the case of H. numata, which is only connected to H. pardalinus (or its ancestor) based on results in Table S6.
For all these reasons, I think it would be helpful to strengthen the demographic results by running a different method that can explicitly model temporal changes in Ne and M and use data from more than two samples. Such methods are implemented in programs like dadi (Gutenkunst et al. 2009: 10.1371/journal.pgen.1000695), moments (Jouganous et al. 2017: 10.1534/genetics.117.200493), FastSimCoal (Excoffier et al. 2011: 10.1093/bioinformatics/btr124) or DILS (Fraïsse et al. 2021: 10.1111/1755-0998.13323). The 2-population versions of these programs could be used (Amazonian vs Atlantic H. numata), or a 3-population version if H. pardalinus is to be considered.

Alternatively, the authors could reconstruct the changes in Ne through time for each population using PSMC-like methods (e.g. Li & Durbin 2011: 10.1038/nature10231), but then, migration has to be neglected.
R: Indeed, we fully agree with all these remarks that prevent any firm conclusion. We also think that using SMC-based methods are limited in that they do not account for gene flow and are not always able to capture the true changes in effective population size (see work by Mazet et al. 2016 for instance).
Therefore, we have now complemented our results with those obtained from ∂a∂i, where we allow temporal changes in *Ne*, we included migration in our model and took into account other confounding effect such as linked selection, which we approximated through a reduction in *Ne*, and barriers to gene flow, approximated through a change in rates of migration (M). We made a comparison of *Amazonian* vs *Atlantic H. numata sample* and of *H. pardalinus* against *H. numata Atlantic*.

Methods (Line 226-274)
Demographic Reconstruction of population size changes, split and mixtures
The G-phocs method provided useful information across all species but i) do not allow to quantify the time scale of population size change, ii) is limited in the number of individuals it can handle and iii) displayed limited accuracy to distinguish Ne and m in a simulation study (Gronau et al. 2011). We thus constructed additional models to test the hypothesis that *H. numata* populations with inversion polymorphism display an increased effective population size due to disassortative mating To test this, we used ∂a∂i to reconstruct the demographic history of *H. numata* individuals from the Amazonian forest, quantify their historical changes in effective population size and test their divergence history from 1) *H. numata* from the Brazilian area, which do not carry the inversion; and 2) *H. pardalinus* individuals. We allowed for change in effective population size in both the ancestral populations. Theoretically, the change in effective population size in *H. numata* associated with the change in mating system should be more recent than the time of introgression of the inversion into *H. numata*. To verify this hypothesis we allowed for change in *Ne* of the daughter population at any time after the split. We tested different models of divergence with and without (asymmetric) migration and included the effect of linked selection.
Since, the conditions of historical divergence are not known, we tested a model of divergence with ongoing migration (IM) a model of divergence with ancient migration if gene-flow has stopped recently (AM) and, in the case of divergence into multiple refugia, a model of secondary contact (SC). We also included a model of Strict Isolation (SI) as a null model.
The models shared the following parameters: the ancestral populations of size N_{anc} can grow or shrink to a size N_{anc2} between T_{anc} and up until its splits at time T_{split} into two daughter populations of size N_1 and N_2. Under the SI model, no gene flow occurs between the two populations. Under AM, gene flow occurred between T_{split} and T_{am} and is followed by a period of strict isolation. Under IM, gene flow occurs at a constant rate at each generation between the two populations. Gene flow can be asymmetric, so that two independent migration rates m_{12} (from population 2 to 1) and m_{21} (from population 1 to 2) were modeled. Under the SC model, the population evolved in strict isolation

between T_{split} and until T_{sc} where a secondary contact occurs continuously up to present time. Gene flow is modeled as M = $2N_{REF}$.m. In ∂a∂i, heterogeneity in effective population size was used to account for linked selection by defining two categories of loci with varying effective population sizes (proportion 1-Q of loci with a "neutral Ne" and a proportion Q of loci with a reduced effective population size due to either selection at linked site). To quantify how linked selection affects reduced N_e, we used a Hill-Robertson scaling factor (Hrf) to relate the effective population size of loci influenced by selection (Nr = Hrf * N_e) to that of neutral loci (N_e). A hierarchical approach was used to avoid over-fitting: first we compared models assuming constant effective population size. Second, the best identified models were modified to incorporate population expansion or decline, as expected given the observed distribution of genetic diversity. Population expansion was implemented using two additional parameters for population 1 and population 2, allowing each population to either grow or decline exponentially at any time after their split from the ancestral population (controlled by parameters s1 and s2 for population 1 and 2 respectively).
Models were fitted using the diffusion theory implemented in ∂a∂i (Gutenkunst et al. 2009) and includes the effect of linked selection. ∂a∂i uses the SFS as a summary of the data. For a given demographic model, the SFS is computed using diffusion approximation and compared to the empirical SFS using AIC.
We used stringent filtering (GQ>30, 4 < mean depth < 80) and no missing data to keep high quality sites and remove potential paralogs or PCR duplicates exhibiting excessive read depth. To minimize linkage we subset our data to keep one SNP every 5kb. No MAF filter was used and singletons were kept to avoid ascertainment bias in estimates of demographic parameters. For each model, 32 independent replicate runs were performed and only models with the lowest AIC and ΔAIC were kept.

The results are now as follows
Line 361 – 377: "The model selection procedure based on AIC gave higher support for a model of secondary contact (SC) in the pairwise comparison between *H. numata* from Peru and *H. numata* from Brazil. The pairwise comparison between *H. numata* and *H. pardalinus* supported a model of divergence with continuous gene-flow (IM) (**Table S7, Figure S3**). All models supported_an expansion occuring in the ancestral population, followed by further growth of the *H. numata* carrying the inversion supergene to reach a size of several millions, which was by far the largest effective size compared to all other species. This stands in stark contrast with the results observed in the samples from Brazil (which do not harbour the inversion) (**Table 1**). Accordingly, *H. numata* populations from the Atlantic forests of Brazil appear to have been subject to a bottleneck at the start of their divergence from Amazonian populations, followed by exponential growth, suggesting a strong (and recent) founding event, leading to a comparatively smaller population size than that observed in the rest of *H. numata*. It is worth noting however that effective population size was hard to estimate in pairwise comparisons between *H. numata* from Peru and SE Brazil. Indeed, parameter uncertainty was large, and model residuals (**Figure S3**) were also large. Our results indicated that *H. pardalinus* displayed an initially large population size followed by a comparatively smaller size expansion than *H. numata* (**Table 1**). Estimates of current effective population sizes are therefore qualitatively similar to those from G-phocs."

***H. numata* Peru (Amazonian) – vs – *H. numata* Brasil (Mata Atlântica)**		***H. Numata* Peru (Amazonian) – vs – *H. pardalinus***	
Model	**AIC**	**Model**	**AIC**
SI2NG	13878	SI2N	10297
SI2N	11405	SI2NG	9917
AM2N	9838	IM2N	6211
IM2N	9762	AM2N	6183
SC2N	8963	SC2N	5923
IM2NG	7587	SC2NG	4029
AMA2NG	6713	SCA2NG	4029

IMA2NG	6179	SIA2NG	3310
AM2NG	6089	IM2NG	3131
SC2NG	5542	AMA2NG	2932
SCA2NG	***2888***	***IMA2NG***	***1294***

Table S7: model selection results comparing different model with linked selection (suffix 2N) with or without growth of the daughter populations (suffix G). The possibility for growth in the ancestral population was also included. The best models are highlighted in bold-italics.

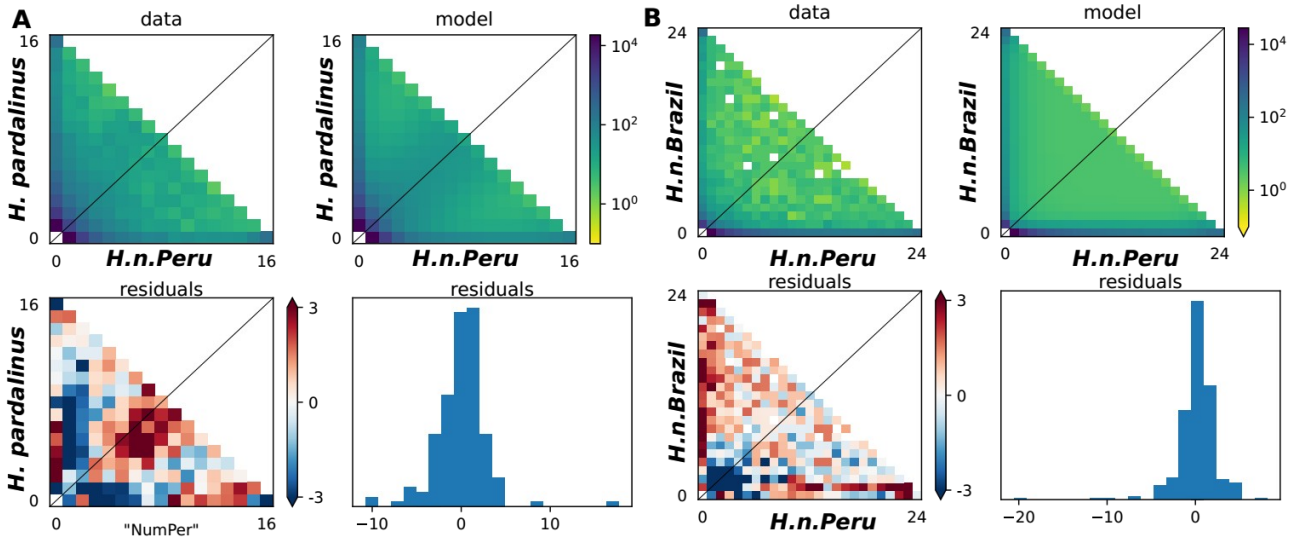


Figure S3 | Differences between observed and predicted jSFS along with model residuals from ∂a∂i for the best models. A) Comparison between H. *numata* from the Amazonian rainforest in Peru and *H. pardalinus.* The best model is one of isolation with migration with linked selection and population size change (IMA2NG). B) Comparison between *H. numata* from Mata Atlantica in Brazil versus *H. numata* from the Amazonian forest in Peru. The best model is one of secondary contact with linked selection and population size change (SCA2NG). Migration was highly asymmetric. Site frequency spectrum for Peru in panel A) was down-sampled to match the number of individuals in *H. pardalinus.*

H. Numata* Perou - *H. pardalinus		***H. numata* Perou- *H. numata* Brazil**	
Best model	**IMA2NG**	Best. Model	**SCA2NG**
Log Lik.	-634.48	Log. Lik	-1430.36
Theta	16,757.39	Theta	3481.23
Nref	6,829,000	Nref	1,036,000
NeAnc	538,000 [0 – 1,281,000]	NeAnc	12,912,000[702,000 – 32,850,000]
NeParda	9,209,000 [3,163,000 – 15,254,000]	NeBrasil	83,000 [74,000 – 93,000]
			60,373,000
NePeru	650,000 [625,000 – 676,000]	NePeru	[6,969,000 – 113,775,000]
NeB1Parda	25,233,000 [7,282,000 – 48,477,000]	NeB1Brasil	292,000 [147,000 – 465,000]
	11,746,000		373,573,000
NeB2Peru	[6,511,000 – 17,368,000]	NeB2Peru	[13,310,000 – 1,190,748,000]
m12	1.99e-08 [0 – 8.59e-08]	m12	4.71E-08 [0 – 1.25e-07]
m21	1.37e-07 [9.84e-8 – 1.79e-7]	m21	3.42E-09 [0 – 2.19e-07]
T_{ANC}	3,237,000 [3,039,000 – 6,169,000]	T_{ANC}	501,000 [254,000 – 748,000]
T_{SPLIT}	1,892,000 [1,674,000 – 2,111,000]	T_{SPLIT}	110,000 [69,000 – 151,000]
	108,000		
Tp1	[662,000 – 1,500,000]	*Tp1*	11,000 [6,000 – 17,000]
Tp2	964,000 [779,000 – 1,150,000]	*Tp2*	55,000 [0 -157,000]
		T_{SC}	41,000 [7,000 – 75,000]

H_{RF}	0.10 [0.03 – 0.17]	H_{RF}	0.47 [0.43 – 0.50]
Q	0.37 [0 – 0.99]	Q	0.47 [0.39 – 0.56]

Table 1: Estimates of demographic parameters for each best model in each comparison. Biological parameters assumied a mutation rates of 2.9e-09 µ/bp/generation. NeAnc, = Effective population sizes for the ancestral population, and descending population 1 and 2. NeParda, NePeru and NeBrazil corresponds to effective population size of *H. pardalinus H. numata* in Peru and *H. numata* in Brazil respectively. NeB1Parda and NeB2Peru, NeB1Brasil provides effective population size for the population *H. pardalinus H. numata* in Peru and *H. numata* in Brazil respectively after their expansion. Hrf = Hill-Robertson factor reflecting the reduction in effective population size due to linked selection. Q = proportion of the genome with a reduced effective population size. m12 = migration from population 2 into population 1. m21 = migration from population 1 into population 2 (scaled according to $2N*ref*m_{ij}$). T_{ANC} = Time of population size change in the ancestral population, Tp1 = Time of size change in population 1 and Tp2 = Time of size change in population 2. T_{ANC}, $T_{SPLIT}$and Tsc, Tp1, Tp2 are provided assuming for generations per year.

2. Demographic inferences - other comments

- As far as I understand, what is expected in terms of Ne is an increase in the polymorphic populations (Amazonian H. numata), which agrees with the inferences (Figure 3). However, connecting the “polymorphic / monomorphic” status with genetic diversity and effective size is not so obvious. Indeed, G-PhoCS infers a demographic expansion in H. elevatus, while this species is monomorphic. Moreover, the diversity difference between the two H. numata populations is partly explained by the decrease in the size of the Atlantic populations (Figure 3).

R: To better investigate the effects of being polymorphic at the supergene we made a forward simulation model and tested the hypothesis that disassortative mating associated with the supergene increased the effective population size through an increase of genetic diversity. We then compared the results to a scenario of assortative mating (expected under local adaptation or speciation) and a null model of random-mating. This is detailed in our reply to the reviewer 2 below

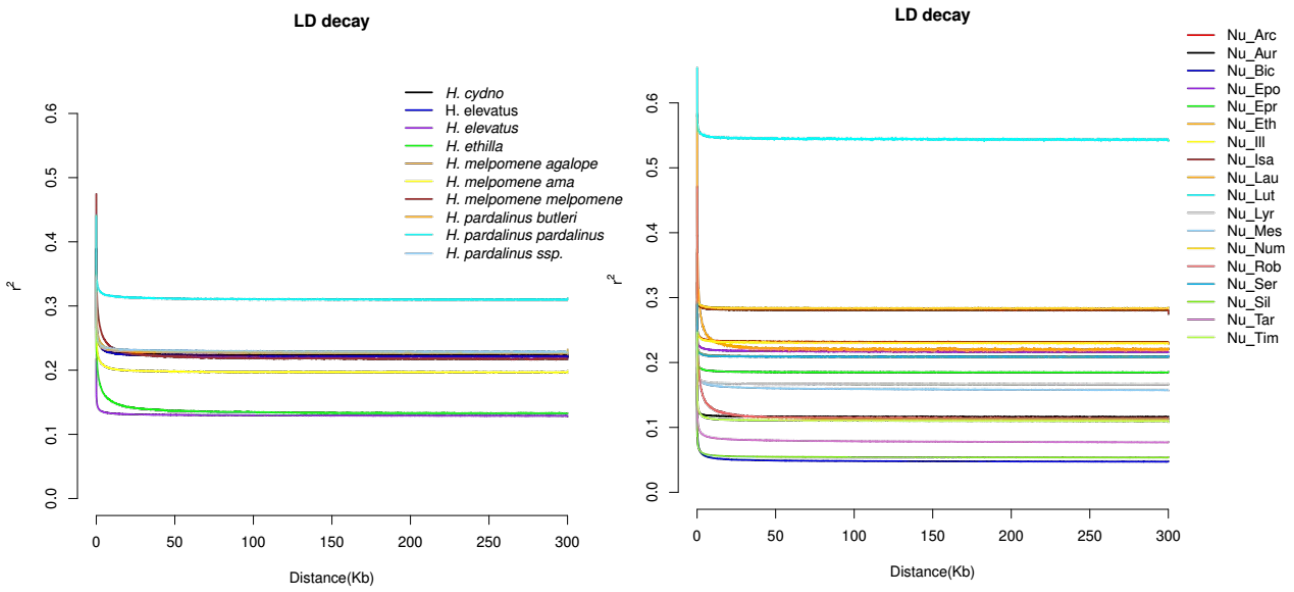
We have also now estimated the size of the Atlantic populations using dadi to complement our previous analysis (see above).

- From the PCA (Figure 2), three genetic clusters make up H. numata (Atlantic, Amazonia, French Guiana) with similar variance explained along each axis. I wondered whether this could inflate the effective size of the Amazonian H. numata populations inferred with G-PhoCS. Given the population structure (even if weak), it may be worth applying the demographic analyses without the French Guiana samples.

R: Indeed, samples from H. Guyana tends to be slightly divergent based on the PCA and were removed from all our demographic inferences to reduce the bias when estimating Ne.

- G-PhoCS assumes no intralocus recombination but free interlocus recombination. Do the filters given on L205 (“4092 genomic regions, each 1kb in length and spaced at approximately 30kb intervals") comply with these assumptions?

R: Based on our computation of the rate of LD decay which decreased rapidly at very short distance, we assumed that 30kb should be far greater than the LD decay to ensure that each genomic regions should be approximately independent and should therefore recombine freely.



- On L200, it is stated that the "inferences are conditioned on a given population phylogeny". I wondered how robust is the phylogeny used in G-PhoCS (Figure 3)?
R: Although Heliconius butterflies are well-known for exchanging allele via gene flow, the species phylogeny of the major Heliconius clades has been particularly well studied, is "robust" and consensual in the Heliconius community. (e.g. Kozak et al. 2015). We state this now in the method.

- Could you please justify what the criteria of Freedman (L213) is based on? It is not obvious why the migration rate threshold of "0.03 with posterior probability larger than 0.5" is meaningful here.
R: Freedman et al. (2013) set on this "lax" criterion to make sure that all gene flow supported by data was taken into account. While this criterion is rather ad-hoc, it matches the results of the other criterion for significance, 0 not included in the 95% credible interval, with one exception (gene flow from the ancestor of *H.pardalinus* and *H. elevatus* to *H.numata*) which is due to the large uncertainty in that migration band.

3. Supergene polymorphism:

- I think it would be helpful to be more precise regarding what is expected for the different types of configurations. Are differences in genetic diversity expected: i) between polymorphic populations with two vs three configurations?; ii) between monomorphic populations with Hn0 (Brazil) vs Hn1 (Venezuela) configurations? R: We state these expectations at the end of the introduction now as follows: *Line 126-133: "We investigate here whether the adaptive introgression of a balanced inversion is associated with a signature in the genetic diversity and geographic structure. In particular we predict that genetic diversity should be higher in H. numata than in closely related taxa. Similarly nucleotide diversity should be higher in all polymorphic population carrying either one segment (Hn1) or two (Hn1,Hn123) compared to the population that is monomorphic and carry only the non-inverted segment (Hn0) in the Brazilian Atlantic Forest."*
- An interesting point not discussed by the authors is the presence of a monomorphic H. numata population (carrying the inversion Hn1) in Tachira (Venezuela). **This sample (n=1) is not presented in Figure 3, so we cannot compare its diversity (pi) with that of the other H. numata populations. If the authors decide to go with PSMC as an alternative demographic method, they can even include the Tachira sample in the analysis to estimate its temporal Ne changes**.
R: We have added the genetic diversity of this sample as well as two additional samples with n = 1 individual. The Tachira sample shows genetic diversity similar to *H. pardalinus.*

4. Confoundingfactors:

- Given the phylogenetic proximity of the taxa considered, I suspect that they should share similar geographic barriers or have a similar dispersal rate. Still, this may not be true. And if differences

exist between taxa, caution is required to interpret the isolation by distance patterns shown in Figure 2C. Typically, if H. numata has higher dispersal capacities than the other taxa and if a geographic barrier exists in the disjunct species range (South East of Brazil, Figure S2), **then Figure 2C could be explained without the need of invoking the supergene effects. Could the authors comment on that?**
R: Indeed. To better understand if the supergene had an effect on levels of genetic diversity and differentiation compared to systems with local adaptation and assortative mating (as in the other species for instance), we performed different sets of forward simulations that are detailed in our reply to Reviewer 2.
A limit of our figure 2C is that we lack short distance and intermediate distance data points for most of our outgroup species. It does not seem obvious from all current literature and available knowledge that all the species have different dispersal ability, however the species differs sometimes in their ecology and habitat preference (e.g. *H. melpomene* is found preferentially along rivers, *H. pardalinus* prefers dry area) These may set constraints into dispersal and effective gene flow.

- In Figure 2C, it seems that two Fst values are not depicted based on Table S4: between H. pardalinus butleri vs sergestus (Fst=0.30574, 30 km) and between H. pardalinus sergestus vs Ssp (Fst=0.30155, 430 km). These two values are outliers (i.e. strong differentiation at small distances). Could the authors explain why they were removed from the figure? Do they correspond to subspecies? At least, an explanation has to be indicated in the legend of Figure 2C.
R: Indeed, these are divergent subpsecies of *Heliconius pardalinus* with elevated levels of sequence divergence and well documented reproductive isolation (see for instance Rosser et al. 2022 https://doi.org/10.1111/mec.16272). They are not necessarily relevant for our purposes.

5. Minor comments
- L41: “they show the highest [...] demographic estimates”. Please, reformulate and specify what are the demographic estimates.
R: This is replaced by:
“They show the highest genetic diversity and effective population size estimates in the entire clade”

- L190: “Scaffolds carrying the supergene rearrangements (Hmel215006 to Hmel215028) were excluded”. Does this correspond to the whole chromosome 15 or only to the scaffolds of chromosome 15 that carry the supergene?
R: Only those that carry the supergene.

- L275: “which contrasts with the low diversity found in the most closely related taxa such as H. ismenius or H. besckei”. I think that H. besckei does not appear anywhere in the results.
R: Indeed, it was not included in this study, so the reference to this species was removed.

- L282: “the distribution of parameters across lineages”. The wording is a bit unclear; please reformulate.
R: This is changed, we integrate our new results in this part of the discussion: “Although low-diversity lineages could have lost diversity due to recent events such as strong bottlenecks, estimates of effective population size from dadi indicated that the Amazonian populations of *H. numata* underwent a dramatic increase in effective population size posterior to their split with Central American (*H. ismenius*) and Atlantic populations, , in agreement with G-Phocs.”

- L349-351: this sentence sounds redundant with L340-343. R: This sentence has been removed.

- Figure 1C: it was unclear whether the colour code (orange, pink, grey) refers to the number of chromosomal configurations. If yes, the Tachira sample should be depicted in grey.

R: These refers to the absence of inversion (Hn0 in grey) or the presence of the inversion P1 (in orange) or the three inversions (P1,P2+P3) with P2 in green and P3 in red.

- Figure 2B: maybe, indicate “Brazil (Atlantic)” instead of “Atlantic” in the legend. R: Done.

- Figure 3A: a space is missing between “H.” and the rest of the name in the legend. Moreover, there is a typo in “Numata French Guiana PR” (I think “PR” should be “FG”).
R: Indeed. This is corrected.

- Figure 3B: there is an extra “s” in “Population names indicate[s]”, and there is a typo in “[showing] that Amazonian”. R: Corrected.

- Figure S1: an “s” is missing in “20 STRUCTURE run[]”, and “reps” should be replaced by “replicates”. R: Corrected.

- Table S4: dxy was clculated, but it was not used in the manuscript. Could the authors comment on that? R: The Dxy was not used and therefore it was removed.

- Table S5: it seems that some values do not fit with the ones reported in Figure 3B. R:Indeed, there was several mismatches that have been corrected now in Figure 3B that have been corrected based on Table S5.

- Table S6: it seems that the “probability that the estimated total migration was greater than 0.03” column is absent.
R: We apologise. This was a wrong table. It has been replaced by a full table including the probability of migration being greater than 0.03

- Text S1: the “Analyses of the slope of Fst versus distance as measured in km” results are hard to follow. It would be clearer if the expectations for each test were stated before the results. R: yes. We have changed this, moreover, it was changed into (Fst/1-Fst) and log(km) in a two dimensional habitat as formulated initially by Rousset under a 2d model (Rousset, 1997).
The maint text in the method is as follows:
Line 178-185: *“Following Rousset (1997), in a 2-dimensional habitat, under a model of isolation by distance (IBD) differentiation, measured as $F_{ST}/(1-F_{ST})$, should increase as a function of the logarithm of the distance. Therefore, we tested for the existence and intensity of an IBD signal among species and between populations of H. numata using a linear model. If IBD is stronger in species not polymorphic for the inversion we should observed significantly steeper slopes in these species. To test this, we measured IBD (I) within populations of each species separately, (ii) for all H. numata within the Amazonian forest but without the Atlantic and (iii) for all H. numata including the Atlantic….”*
We remind these expectations at the start of text S1.
- Text S1: the “Testing for an effect of the inversion on population differentiation” results are hard to follow. Maybe add a sentence that clearly says if results with vs without chromosome 15 were the same or not.
R: Indeed, this is changed as follows: *“Therefore the slope of Fst vs distance without chromosome 15 was significantly higher when including the Atlantic population. The result were the same with or without chromosome 15 and suggests that including or not the inversion in our computation did not have a major influence. Overall the slope of the signal of IBD was lower when comparing only Amazonian populations than when the Atlantic population was considered.”*

##########
Reviewed by anonymous reviewer, 22 Dec 2021 12:24

This is a short article describing the genetic diversity of populations of Helionius numata, with comparisons with that of other Heliconius butterfly species.
The main finding seems that the H. numata species is divided into two populations:
A large population in Amazonia, which includes the samples taken from the Andean foothills (the vast majority of samples in the study) and samples from a locality in French Guiana; this population has very high genetic diversity and low isolation by distance.
A population in the Mata Atlântica region, represented by four samples, which has much lower genetic diversity.
The authors compare the genetic diversity of these two populations with that of populations of other Heliconius species, and include a model of the evolution of these species using the programme G-PhoCS. These analyses suggest that the Amazonian population does indeed have higher genetic diversity than the ancestral state.

I have two major reservations regarding this article.
First, the sampling strategy is extremely unbalanced: the genetic diversity of the Amazonian population is estimated with many samples from numerous places, whereas the diversity of the samples are estimated from only four samples from a single locality. The authors give excellent evidence that the genetic diversity of Amazonian populations is high. However, without a larger sampling breadth for the Mata Atlântica region, it is impossible to know whether the sampled population is representative of the entire region. **As far as we know, the Atlantic forests of Brazil are comparatively more highly fragmented than the Amazonian rainforest - a recent bottleneck of the sampled population cannot be ruled out.**
R: Indeed, our few samples may not represent the breadth of genetic diversity of the region. Therefore, we have added a few additional WGS samples (n total for Atlantic = 12). Moreover; as pointed out by the reviewer, a bottleneck may explain the low genetic diversity observed. Therefore, we tested this hypothesis explicitly in ∂a∂i by allowing for change in population size of the daughter populations through time after the split of the ancestral *H. numata* population into the Amazonian and Atlantic population (see details in reply to R1 above).
As suggested by the reviewer, our results indeed suggest a bottleneck in this population (see reply to R1 above on the dadi results section)

We also discuss the need for further extensive sampling of the region:
Line 501-504: *"Reduced effective population size is supported by our data. One major caveat associated to our inference remains the small number of individuals (n = 12) from the Atlantic forest. Genetic diversity might be underestimated, notably if populations have a history of fragmentation in this area."*

The second major reservation is the interpretation of the genetic diversity as being caused by disassortative mating as a result of dominance-selection regime of the mimicry polymorphism. **The authors do not show that the genetic diversity observed is higher than what would be expected under a very large population size and random mating.** Because of that, the main claim (as plausible as it is) seems too forceful in the way it is expressed in the Title, Abstract, Introduction and Discussion. Saying that, we do agree that the authors should discuss the possible importance of disassortative mating in the Discussion section.
R: Indeed. To gain further insights into the processes that can generate high generate diversity we now have explicitly tested the hypothesis that genetic diversity observed in *H. numata* harbouring the supergene is higher under a model with disassortative mating than expected under random mating. We also tested the expected levels of genetic diversity under assortative mating (the common case in locally adapted populations of *Heliconius* species).
To do so we performed forward simulations as detailed below:

Method (Line 276-315):

Forward Simulations

In order to better understand the nature of the processes that generate higher genetic diversity in *H. numata* compared to closely related taxa, we used simulations to test the hypothesis that disassortative mating generates an increase in levels of genetic diversity at a genome-wide scale. We hypothesized that such level of genetic diversity is higher than expected under i) random mating (a model similar to panmixia) or ii) assortative mating, as commonly observed in other *Heliconius* species. To test this hypothesis we run forward simulation under disassortative, assortative and random matting using slim v3.6 (Messer et al. 2013).

We simulated a stepping stone model with 10 demes, each composed of 1,000 diploid individuals and connected by a (symmetric) migration parameter (m). Each individual received neutral and deleterious (ratio 16:6) mutations at a rate μ = 1e-8 μ/bp/generation (rescaled to μ = 1e-6 for faster simulation of a larger population). We simulated an individual with a pair of 1Mb chromosome, including a single locus with 5 alleles with perfect dominance (allele 1 > allele 2 > allele 3 > allele 4 > allele 5) given 5 possible alternative phenotypes (referred hereafter as "morph"). Each allele was fully linked (no recombination) with a given deleterious recessive mutation, generating overdominance at this loci so that polymorphism is always maintained. Local adaptation was introduced in the model through a single parameter defining randomly which morphs were favored in each population. In each population, either 2 or 3 morphs benefited from a fitness advantage compared to the others. The fitness reduction varied between 0 (= fitness of zero for migrants in a demes) and 1 (no reduction of fitness). We tested 3 possible values for this parameter (0, 0.25 and 0.5).

Finally, disassortative mating was controlled by a mate choice parameter defining whether a morph would reproduce with another morph. The strength of the parameter varied between 0 (= complete disassortative mating) and 1 (= no mating weight). We tested 3 possible values for this parameter (0, 0.25 and 0.5).

We run the model for 80,000 generations to reach demographic equilibrium and assessed levels of synonymous diversity (π_s). We tested all combinations of the 3 values for levels of disassortative mating and local adaptation and ran 10 replicates per combination in order to estimate the variance around π_s.

Similarly, we run a model with strict assortative mating, controlled by a parameter defining whether similar morphs reproduced together. The strength of the parameter varied between 0 (complete assortative mating where a given individual mate only with an identical morph) and 1 (where individual mate randomly with regards to the morph). We tested 3 possible values for this parameter (0, 0.25 and 0.5). As for disassortative mating, all combinations of assortative mating and local adaptation values were tested. For each model we tested 3 values for the migration rate, m = 1e-4, 1e-6 and 1e-8, resulting in a total of 54 comparisons.

For graphical display in Figure 4, the values of assortative/disassortative mating were rescaled on a scale between (0 and 1) with 0 indicating no disassortative mating but complete assortative mating and 1 complete disassortative mating (or no assoartive mating). A value of 0.5 was equivalent to random mating.

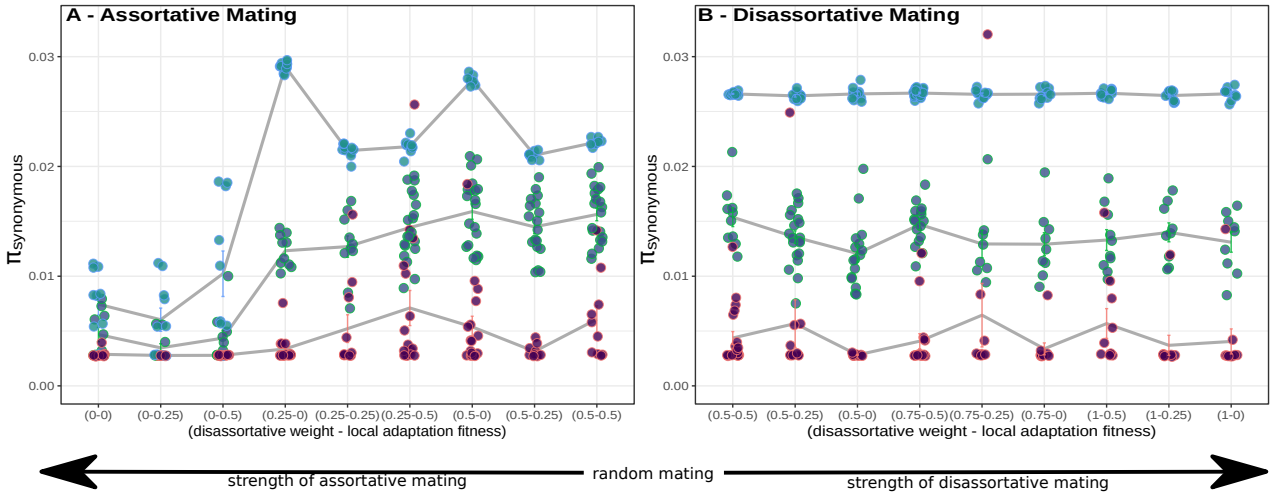
Results (Line 380 -- 388)

Forward simulations

Forward simulations under different levels of local adaptation (controlled by the strength of divergent selection), disassortative mating and migration are displayed in **Figure 4B**. The same results under a model of assortative mating involving different levels of selection and migration are

displayed in **Figure 4A**. Overall, synonymous genome-wide nucleotide diversity (π_S) was higher in 73 % of the models including disassortative mating (average π_S = 0.0145) when compared to their equivalent under assortative mating (average π_S = 0.011), a weak but significant difference ($p < 0.01$, see **Figure S5**). In summary, modest differences were observed among models with different strengths of divergent selection or disassortative mating, the most influential variable being the rate of migration (**Figure 4)**.

Figure 4: Differences in synonymous nucleotide diversity (π_S) emerged at a genome-wide scale under divergent selection and mating region. Results from forward simulations of 10 populations



undergoing local adaptation and different mating strategy are presented. Shown are levels of synonymous diversity obtained under assortative (A) versus disassortative mating (B) under different rates of migration and different local adaptation fitness. Each combination of parameters in brackets display the (dis)-assortative mating weight and the fitness value for local adaptation respectively. A left value of 0 in the bracket means complete assortative mating and 0.5 means no assortative mating or disassortative mating. Value of 1 means complete disassortative mating. A right value of 0 in the bracket means a fitness of 0 for non locally adapted individuals in a demes, A value of 0.5 means a reduced fitness of 0.5 relative to the maximum value.

The discussion has been changed according to these results at several places of the manuscript. See directly in text.

Minor comments:

The Methods and Results are written well, but lack preciseness in places.

R: We have attempted to increase precision everywhere when possible (see all minor corrections in text).

in line 222 and line 226, the authors should mention the number of individuals in each population (n=XX)

R: Done. (N = 12)

For the FST analysis (in line 234, Figure 2C and the relevant part of the Methods section), it is unclear what the authors are doing: which populations are compared with which populations?

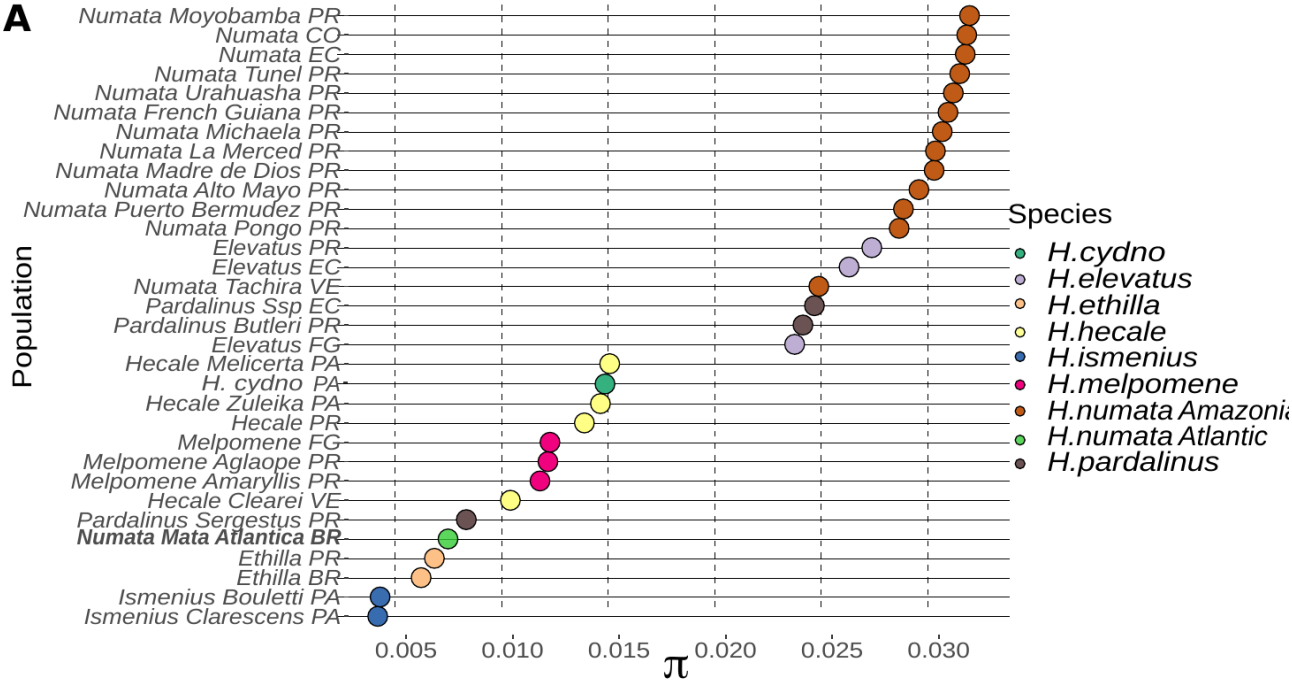
R: See the new Figure 2C, its legends as well as methods section.

Line 179 – 185:Following Rousset (1997), in a 2-dimensional habitat, under a model of isolation by distance (IBD) differentiation, measured as $F_{ST}/(1-F_{ST})$, should increase as a function of the logarithm of the distance. Therefore, we tested for the existence and intensity of an IBD signal

among species and between populations of H. numata using a linear model. If IBD is stronger in species not polymorphic for the inversion we should observed significantly steeper slopes in these species. To test this, we measured IBD (I) within populations of each species separately, (ii) for all H. numata within the Amazonian forest but without the Atlantic and (iii) for all H. numata including the Atlantic.

Results (line 333:)
Isolation by distance among Amazonian populations of H. numata, estimated using the proxy F_{ST}/1-Fst ~ ln(km) was significant (R^2 = 0.407, p = 1.46e-06, slope = 4.95). Comparison among other species also revealed a significant IBD (R^2 = 0.22, p = 0.007, slope = 5.9) although the very different sampling design probably contribute to the difference of variance explained. An analysis of the slope revealed a lower rate of increase in F_{ST} with distance in H. numata compared to all other taxa (Fig 2C, Table S4, Supp Text S1). By contrast, IBD between Atlantic and Amazonian populations of H. numata is close to what is observed in other species, and not significantly different(see Supp. Text S1).

In Figure 3A, it is not very clear which population is the numata one from Mata Atlântica. This is a general trend across the figures - the labels and biological categories are difficult to track across figures.
R: See the modification which should better highlight the population.
We also have modified Figure 3B to add colors to each species, in order to match Fig 3A.
In Fig 2A-B colors are matching between admixture and the PCA so do colors of admixture in Fig S2.



##########

Reviewer #3

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.
R: Thank you

Major comments:

-

The result section is extremely short (34 lines!).
R: Indeed. The new version has now more than doubled in length!

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.
R: The chromosome 15 was actually removed from Fst and pi calculations. The Dxy was removed since it was not used.

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.
R: We agree this is a problem and choosing another set of individuals would have been useful. The chosen individuals attempted to strike a balance between quality (as measured by sequencing depth) and morphs in the areas, as our dataset had an over-representation of some morphs. In an attempt to comply with the comment raised by R1, we have chosen to complement our analyses by using ∂a∂i, in order to not only estimate Ne, but also its change through time in a more realistic model that accounts for gene flow and confounding factors such as linked selection. Further, this enabled us to explicitly compare possible models of divergence. Since our ∂a∂i analyses are based on the site frequency spectrum, with more individuals per sample we hope that they do not suffer from the aforementioned problem (see reply to R1 for all details about our new modelling). We kept the G-PhoCs results but mainly used ∂a∂i to draw our new conclusion. We also remind the limits of G-PhoCs in the methods as follows:

Line 227- 229:
"The G-phocs method provided useful information across all species but i) do not allow to quantify the time scale of population size change, i) is limited in the number of individuals it can handle and iii) displayed limited accuracy to distinguish Ne and m in a simulation study (Gronau et al. 2011)."

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.
R: indeed, a sample with only one individual is not good for Fst. They were removed for this. We now have included them in Pi computations. We keep them for PCA and admixture as well.
Figure 3A with additional samples (n = 1) is provided in reply to R2 just above.

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.
R: We have added migration arrows among major groups in a supplementary figure.
Moreover we invite you to look at our new results based on dadi in the reply to R1.
-
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.
R: In fact there actually is a significant IBD (R2 = 0.407, p = 1.6e-6) vs (R^2 = *0.22, p = 0.007)* for interspecies comparison). The difference is that the slope is actually very weak, which suggests a strong difference in dispersal and or effective population sizes.
We don't know any examples. Also, see our new slim simulation which shows that random mating could also generate similar patterns.

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.
R: Indeed, this is mostly related to differences in the methods and their sensitivity to the number of individuals. The Fst averaged over all comparison between Guiana and other Amazonian samples was 0.0285 whereas it is 0.015 when considering comparison among the Amazonian population, so there are some differences. Clustering analyses such as Structure and Admixture are known to be sensitive to (ii) unbalanced samples, (ii) the number of markers and (iii) presence of hierarchical structure and (iv) isolation by distance. Several of these factors may have affected our power to distinguish the French Guiana (FG), especially the strong structure of the Brazilian (Mata Atlantica) population which may hide a more subtle population genetic structure. Therefore, we excluded the Brazilian samples and took a random set of the Peruvian *H. numata* samples. In this case admixture analysis separates Peru from FG (the same was true when using Columbia/Ecuador in place of Peru). We present this graph in Supplementary Figure and present these results shortly.

Line 324 – 345:
*Individuals from the Atlantic populations of H. numata clustered together to one side of the first PCA axis, whereas all other individuals from all other populations clustered to the other side. The second axis of the PCA separates individuals from French Guiana from the other samples of the upper Amazon. This separation was not found with Admixture (i.e. with K=3) from the complete dataset. To better investigate the existence of a hierarchical population structure, we excluded individuals from the Atlantic populations and compared individuals from French Guiana to a randomly sampled set of Peruvian individuals. In this case we found a clear separation in two groups corresponding to French Guiana and Peru (**Fig S2A**). The same pattern was observed when replacing Peru by Colombia or Ecuador (**Fig S2B,C**).. In accordance, pairwise genome-wide estimates of differentiation (F_{ST}) between H. numata populations showed elevated values when*

comparing the Atlantic population to other populations, low values between French Guiana and other Amazonian population, and were the lower when comparing pairs of Amazonian populations outside of French Guiana. (Fig 2C, Table S4). For instance, the population from La Merced in Peru shows an F_{ST}=0.032 with the population from French Guiana at a distance of 3019km, but an F_{ST}=0.311 (an order of magnitude higher) with the Atlantic population at a similar distance. The comparison between La Merced and Ecuador was even lower (Fst = 0.0159).

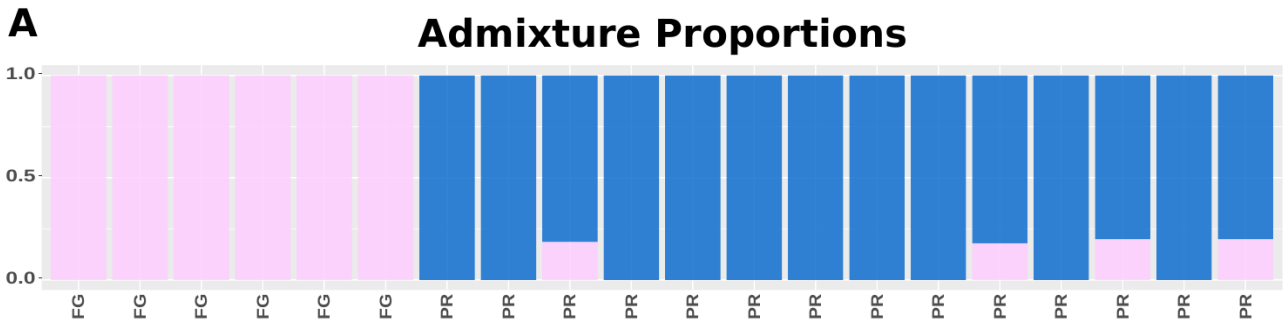


Figure S3A: admixture proportions infered between French Guiana (FG) and Peru (PR). Color code match those provided in Figure 2A and B.

L123-124: Reference?
R: (Jay et al. 2018)

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.
R: Indeed. We rephrase this as follows,
Line 124 – 127:
" The introgression of P_1 and the formation of a supergene were associated with a major shift in the selection regime (Jay et al. 2018). The mating system was also changed during or after introgression. These events may therefore have profoundly affected the population biology of the recipient species, H. numata.."

L130-132. I do not understand this sentence.
R: This is changed from:
"Our results are consistent with the selection regime and mating system associated with supergene formation having enhanced gene flow among populations and increased effective population size."
into Line 134 – 137:
"Our results suggest that following supergene formation, a change in the selection regime and mating system may have facilitated gene flow among morphs and had key consequences in current patterns of genetic structure."

L176-178: How were the 15 numata individuals from Peru chosen?
R: They were chosen to have a better representation of the morphs in the Peruvian populations (this is now written).

L180: how were the SNP chosen?
R:using the thinning function in vcftools **(**Danecek & al. 2011**)**

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

R: Corrected.

L206: is there a particular reason for this choice of 30kb? Does it correspond to a LD decay threshold?
R:This extends far beyond the rate of LD Decay, to ensure independence of the region. (see plot above in reply to R1)

L208: I am not sure exactly what migration bands are supposed to be.
R: Migration bands describe gene flow after divergence in a demographic model. We have rephrased that sentence to priors in the migration rates.

L271-273: I am not sure what the authors referring too here. Based on Figure 3A, pi varies by an order of magnitude.
R: Indeed, we have corrected this (the average pi of H. numata Amazonia is actually four times higher than in H. numata from Brazil).