

1 **Supergene formation is associated with a major shift in genome-wide patterns of diversity in a**
2 **butterfly**

3

4 María Ángeles Rodríguez de Cara^{1*\$}, Paul Jay^{1*\$}, Mathieu Chouteau^{1,2}, Annabel Whibley^{3,4}, Barbara
5 Huber⁵, Florence Piron-Prunier³, Renato Rogner Ramos⁶, André V. L. Freitas⁶, Camilo Salazar⁷,
6 Karina Lucas Silva-Brandão⁸, Tatiana Texeira Torres⁹, Mathieu Joron^{1S}

7

8 * contributed equally

9 ¹Centre d'Ecologie Fonctionnelle et Evolutive (CEFE), Univ Montpellier, CNRS, EPHE, IRD,
10 Montpellier, France

11 ²Laboratoire Ecologie, Evolution, Interactions Des Systèmes Amazoniens (LEEISA), Université de
12 Guyane, IFREMER, CNRS, Cayenne, Guyane Française

13 ³Institut de Systématique Evolution Biodiversité (ISYEB), Museum National d'Histoire Naturelle,
14 CNRS, Sorbonne-Université, EPHE, Université des Antilles, Paris, France

15 ⁴School of Biological Sciences, University of Auckland, Auckland, New Zealand

16 ⁵Instituto de Ciencias Ecológicas y Ambientales (ICAE), Univ de los Andes, Mérida, Venezuela

17 ⁶Departamento de Biologia Animal, Instituto de Biologia, Unicamp, Campinas, São Paulo, Brazil

18 ⁷Department of Biology, Faculty of Natural Sciences, Universidad del Rosario, Carrera 24 No 63C-
19 69, Bogotá 111221, Colombia.

20 ⁸Centro de Biologia Molecular e Engenharia Genética, Universidade Estadual de Campinas. Av.
21 Candido Rondon 400. Campinas, São Paulo, Brazil

22 ⁹Department of Genetics and Evolutionary Biology, Institute of Biosciences, University of São
23 Paulo (USP), São Paulo, Brazil

24 \$ Corresponding authors: angeles.decara@gmail.com, paul.yann.jay@gmail.com,
25 mathieu.joron@cefe.cnrs.fr

26

27 **Abstract:** Selection shapes genetic diversity around target mutations, yet little is known about how
28 selection on specific loci affects the genetic trajectories of populations, including their genome-
29 wide patterns of diversity and demographic responses. Adaptive introgression provides a way to
30 assess how adaptive evolution at one locus impacts whole-genome biology. Here we study the
31 patterns of genetic variation and geographic structure in a neotropical butterfly, *Heliconius numata*,
32 and its closely related allies in the so-called melpomene-silvaniform subclade. *H. numata* is known
33 to have evolved a supergene via the introgression of an adaptive inversion about 2.2 million years
34 ago, triggering a polymorphism maintained by balancing selection. This locus controls variation in
35 wing patterns involved in mimicry associations with distinct groups of co-mimics, and butterflies
36 show disassortative mate preferences and heterozygote advantage at this locus. We contrasted
37 patterns of genetic diversity and structure 1) among extant polymorphic and monomorphic
38 populations of *H. numata*, 2) between *H. numata* and its close relatives, and 3) between ancestral
39 lineages in a phylogenetic framework. We show that *H. numata* populations which carry the
40 introgressed inversions in a balanced polymorphism show markedly distinct patterns of diversity
41 compared to all other taxa. They show the highest diversity and demographic estimates in the entire
42 clade, as well as a remarkably low level of geographic structure and isolation by distance across the
43 entire Amazon basin. By contrast, monomorphic populations of *H. numata* as well as its sister
44 species and their ancestral lineages all show the lowest effective population sizes and genetic
45 diversity in the clade, and higher levels of geographical structure across the continent. This suggests
46 that the large effective population size of polymorphic populations could be a property associated
47 with harbouring the supergene. Our results are consistent with the hypothesis that the adaptive
48 introgression of the inversion triggered a shift from directional to balancing selection and a change
49 in gene flow due to disassortative mating, causing a general increase in genetic diversity and the
50 homogenisation of genomes at the continental scale.

51

52 **Introduction:** Genetic diversity is shaped by selective processes such as stabilizing or disruptive
53 selection, and by demographic processes such as fluctuations in effective population size. Empirical
54 studies on genetic diversity within and among populations abound, fuelled by an increasing
55 availability of whole genome data, and spurred by our interest in understanding the underlying
56 causes of variation in diversity (e.g. Beichmann 2018, Muers 2009; Murray 2017; Nielsen et al.
57 2009). At the locus scale, strong directional or disruptive selection tends to reduce diversity within
58 populations (Mitchell-Olds et al. 2007), while balancing selection tends to enhance diversity
59 (Charlesworth 2006). Genome-wide factors reducing diversity include low effective population
60 sizes, generating drift, while high genetic diversity is enhanced by large population sizes and gene
61 flow. Overall, it is well recognised that demographic changes should have a genome-wide effect on
62 diversity, while positive selection is expected to play a role on the sites within and around the genes
63 involved in trait variation (Glinka et al. 2003, Muers 2009, Nielsen et al. 2009).

64

65 Variation in behaviour and life-history traits, for instance involving changes in offspring viability or
66 dispersal distance, may also affect species demography, and thus whole genome genetic diversity.
67 However, whether and how genetic variability in a population may be driven by phenotypic
68 evolution at certain traits is poorly understood, and confounding effects may affect patterns of
69 genomic diversity, such as variation in census population size or colonization history. Dissecting
70 how selection on a trait may affect genome-wide diversity can be tackled by comparing closely-
71 related populations differing at this trait coupled with knowledge of when the differences evolved.
72 Here, we took advantage of the dated introgressive origin of a chromosomal inversion associated

73 with major life-history variation to study the demographics and whole genome consequences of
74 changes in the selection regime at a major-effect locus.

75

76 *Heliconius* butterflies are aposematic, chemically-defended butterflies distributed over the
77 American tropics from Southern Brazil to Southern USA (Emsley 1965; Brown 1979) (Fig 1A).
78 *Heliconius* butterflies are well-known for visual resemblance among coexisting species, a
79 relationship called Müllerian mimicry which confers increased protection from bird predators
80 through the evolution of similar warning signals (Sheppard et al. 1985). Most species are locally
81 monomorphic, but their mimicry associations vary among regions, and most species display a
82 geographic mosaic of distinct mimetic “races” through their range. In contrast to most *Heliconius*
83 species, the tiger-patterned *Heliconius numata* is well-known for maintaining both mimicry
84 polymorphism within localities, with up to seven differentiated coexisting forms, and extensive
85 geographic variation in the distribution of wing phenotypes (Brown & Benson 1974; Joron et al.
86 1999). Forms of *H. numata* combine multiple wing characters conveying resemblance to distinct
87 sympatric species in the genus *Melinaea* and other local Ithomiini species (Nymphalidae:
88 Danainae). Polymorphism in *H. numata* is controlled by a supergene, i.e. a group of multiple linked
89 functional loci segregating together as a single Mendelian locus, coordinating the variation of
90 distinct elements of phenotype (Brown & Benson 1974; Joron et al. 2006). Supergene alleles are
91 characterized by rearrangements of the ancestral chromosomal structure, forming three distinct
92 chromosomal forms with zero (ancestral type, Hn0), one (Hn1) or three chromosomal
93 rearrangements (Hn123) (Fig 1B). The ancestral arrangement, Hn0, devoid of inversions, is fixed in
94 most *Heliconius* species (although an inversion in the same region evolved independently in a
95 distantly-related *Heliconius* lineage (Edelman et al. 2019)). Arrangement Hn1 contains a 400kb
96 inversion called P₁ originating from an introgression event about 2.2 My ago from *H. pardalinus*, in
97 which P₁ is fixed (Jay et al. 2018). This introgression is thought to be the founding event triggering
98 the formation of the supergene and the maintenance of polymorphism in *H. numata* (Jay et al.
99 2018). Arrangement Hn123 displays two additional inversions, P₂ and P₃, in linkage with P₁, and
100 therefore originated after the introgression of P₁ into the *H. numata* lineage (Jay et al. 2021).

101

102 *Heliconius numata* is widespread in the lowland and foothill tropical forests of the Amazon basin,
103 the Guianas, and the Brazilian Atlantic Forest (Mata Atlântica), but the frequencies of the three
104 chromosome arrangements vary across the range. Ancestral type Hn0 is fixed in the Atlantic Forest
105 populations of Brazil (forms *robigus* or *ethra*), but segregates at intermediate frequencies in all
106 other *H. numata* populations throughout the range (forms *silvana* and *laura*) (Fig 1C). Chromosome
107 type Hn1 is associated with the Andean mimetic form *bicoloratus* and is found in the Eastern
108 Andean foothills of Ecuador, Peru, and Bolivia. Chromosome type Hn123 is associated with a large
109 diversity of wing-pattern forms of intermediate allelic dominance, including *tarapotensis*, *arcuella*
110 and *aurora*, and is reported from Andean, lowland Amazonian and Guianese populations. Inversion
111 polymorphism is therefore structured across the range, with populations being fixed for the
112 ancestral chromosome (Atlantic Forest, see Text S1 & Table S1-2), or displaying a polymorphism
113 with two (Amazon-Guiana) or three (Andes) chromosomal types in coexistence (Joron et al. 2011).
114 Monomorphic populations of the Atlantic forest, devoid of rearrangements at the supergene locus,
115 might represent the ancestral state displayed by *H. numata* populations before the evolution of the
116 supergene via introgression (Fig 1C).

117

118 The wing patterns of *H. numata* are subject to selection on their resemblance to local co-mimics

119 (Chouteau et al. 2016), but the polymorphism is maintained by balancing selection on the
120 chromosome types. Balancing selection is indeed mediated by disassortative mating favouring
121 mixed-form mating (Chouteau et al. 2017) and is likely to have evolved in the response to the
122 deleterious mutational load carried by inversions, which causes heterozygous advantage in *H.*
123 *numata* (Jay et al. 2021, Faria et al. 2019, Maisonneuve et al. 2019). The introgression of P_1 and the
124 formation of a supergene were associated with a major shift in the selection regime and in the
125 mating system and may therefore have profoundly affected the population biology of the recipient
126 species, *H. numata*. We investigate here whether the adaptive introgression of a balanced inversion
127 is associated with a signature in the genetic diversity and geographic structure. We analyse changes
128 in the demographic history of the clade containing *H. numata* and closely related taxa, as well as
129 their current patterns of diversity and demography, using three well separated populations of *H.*
130 *numata* representing different states of inversion polymorphism. Our results are consistent with the
131 selection regime and mating system associated with supergene formation having enhanced gene
132 flow among populations and increased effective population size. Moreover, our findings highlight
133 that balancing selection and a shift in mating systems associated with chromosomal polymorphism
134 may reshape genomewide diversity, with crucial consequences on current patterns of genetic
135 structure and population ecology.

136

137 **Material and Methods**

138 We used here whole genome resequencing from 137 specimens of *Heliconius*, including 68 *H.*
139 *numata*. Sampling included specimens from populations in the Andean foothills (3 chromosome
140 types), from the upper Amazon (2 chromosome types), from French Guiana (2 chromosome types)
141 and from the Brazilian Atlantic Forests (1 chromosome type) (Fig 1C; Table S3). Related taxa were
142 represented by the sister species *H. ismenius*, found west of the Andes (parapatric to *H. numata*), by
143 Amazonian representatives of the lineage *H. pardalinus* (donor of the inversion), *H. elevatus*, *H.*
144 *ethilla*, *H. besckei* as well as *H. hecale*, and by *H. melpomene* and *H. cydno* as outgroups. Only
145 Andean, Amazonian and Guianese populations of *H. numata* display chromosomal polymorphism,
146 all other taxa being fixed for the standard gene arrangement (Hn0), or for the inverted arrangement
147 Hn1 (*H. pardalinus*) (Jay et al. 2018). Hereafter, *H. numata* populations from the Andes, Amazon
148 and French Guiana will be collectively referred to as “Amazonian”, and populations from the
149 Atlantic Forest as “Atlantic”. Butterfly bodies were preserved in NaCl saturated DMSO solution at
150 20°C and DNA was extracted using QIAGEN DNeasy blood and tissue kits according to the
151 manufacturer’s instructions with RNase treatment. Illumina Truseq paired-end whole genome
152 libraries were prepared and 2x100bp reads were sequenced on the Illumina HiSeq 2000 platform.
153 Reads were mapped to the *H. melpomene* Hmel2 reference genome (Davey et al., 2016) using
154 Stampy (version 1.0.28; Lunter and Goodson, 2011) with default settings except for the substitution
155 rate which was set to 0.05 to allow for the expected divergence from the reference of individuals in
156 the so-called silvaniform clade (*H. numata*, *H. pardalinus*, *H. elevatus*, *H. hecale*, *H. ismenius*, *H.*
157 *besckei* and *H. ethilla*). *H. melpomene* and *H. cydno* belonging to the so-called *melpomene* clade,
158 their genomes were mapped with a substitution rate of 0.02. Alignment file manipulations were
159 performed using SAMtools v0.1.3 (Li et al. 2009). After mapping, duplicate reads were excluded
160 using the *MarkDuplicates* tool in Picard (v1.1125; <http://broadinstitute.github.io/picard>) and local
161 indel realignment using IndelRealigner was performed with GATK (v3.5; DePristo et al. 2011).
162 Invariant and polymorphic sites were called with GATK HaplotypeCaller, with options --
163 min_base_quality_score 25 --min_mapping_quality_score 25 -stand_emit_conf 20 --heterozygosity
164 0.015.

165

166 F_{ST} , d_{XY} and π , were calculated in overlapping windows of 25 kb based on linkage disequilibrium
167 decay (*Heliconius* Genome Consortium 2012) using custom scripts provided by Simon H. Martin
168 (<https://github.com/simonhmartin>), and the genome-wide average was calculated using our own
169 scripts (available from <https://github.com/angelesdecarra>). Distance in km between sampling sites
170 was measured along a straight line, not taking into account potential physical barriers. The slopes of
171 F_{ST} versus distance was calculated using the R package *lsmeans* (Lenth 2016); the slope difference
172 among species or between populations within species was estimated with an ANOVA and its
173 significance evaluated with function pairs of this package (Text S1 and see example script on
174 github.com/angelesdecarra).

175

176 Admixture (Alexander et al. 2009) analyses were run on a subset of the 68 *H. numata* genomes,
177 keeping only 15 individuals from Peru to have a more balanced representation of individuals across
178 the geographic distribution. Filters were applied to keep biallelic sites with minimum mean depth of
179 8, maximum mean depth of 200 and at most 50% genotypes missing. We only kept 1 SNP per
180 kilobase to remove linked variants, and we obtained the optimal number of clusters using cross-
181 validation for values of K from 1 to 10 (Alexander et al. 2009). Principal component analyses
182 (PCA) were performed with the same filters as for admixture, using the same *H. numata* genomes
183 as for the admixture analyses, using smartpca (Patterson et al. 2006).

184

185 In order to estimate demographic parameters independently of the effect of selection on diversity,
186 we performed stringent filtering on the dataset. We removed all predicted genes and their 10,000
187 base-pair flanking regions, before performing G-PhoCS (Gronau et al. 2011) analyses as detailed
188 below. Repetitive regions were masked using RepeatMasker and Tandem Repeat Finder (Benson
189 1999). GC islands detected with CpGcluster.pl with parameters 50 and 1E-5 (Hackenberg et al.,
190 2006) were also masked. Scaffolds carrying the supergene rearrangements (Hmel215006 to
191 Hmel215028) were excluded, as were scaffolds from the sex chromosome (Z), since those are
192 expected to show unusual patterns of diversity due to selection and different effective population
193 sizes.

194

195 We analysed the demographic history of *H. cydno*, *H. numata*, *H. ismenius*, *H. pardalinus* and *H.*
196 *elevatus* with G-PhoCS, which allows for the joint inference of divergence time, effective
197 population sizes and gene flow. In order to detect differences in demography correlating with the
198 presence of the supergene in *H. numata*, we conducted analyses separating the Atlantic population
199 of *H. numata* from Amazonian populations. G-PhoCS is an inference method based on a full
200 coalescent isolation-with-migration model. Inferences are conditioned on a given population
201 phylogeny with migration bands that describe allowed scenarios of post-divergence gene flow. The
202 model assumes distinct migration rate parameters associated with each pair of populations, and
203 allows for asymmetric gene flow. Given the computational burden of G-PhoCS, we selected two
204 individuals per taxon or population, retaining those with the highest sequencing depth (see Table
205 S3). The input dataset consisted of 4092 genomic regions, each 1kb in length and spaced at
206 approximately 30kb intervals and with genotypes in at least one of the two samples of each taxon
207 We used as priors for coalescence times (τ) and genetic diversity (θ), Gamma functions with $\alpha=1$
208 and $\beta=100$, and for migration bands $\alpha=0.002$ and $\beta=0.00001$. These priors were chosen to allow
209 good convergence while also ensuring non informativity. In order to calculate the highest posterior
210 density interval, we used the library HDInterval in R, and to integrate such posterior densities we

211 used the library sfsmisc in R. We rescaled the results using a mutation rate of 1.9E-9 (Martin et al.
212 2016) and 4 generations per year (i.e., $g=0.25$). Migration bands were considered significant
213 following the criteria of Freedman et al. (2012): if the 95% HPD interval did not include 0 or if the
214 total migration was larger than 0.03 with posterior probability larger than 0.5.

215

216

217

218 **Results**

219 Using cross validation error as a measure of the optimal number of clusters with Admixture, we
220 found that $K=2$ was the optimal cluster number describing within-species genetic variation in *H.*
221 *numata* (Fig 2A). One cluster corresponds to the Atlantic population, forming a well-differentiated
222 genetic entity compared to all other *H. numata* populations. All Amazonian populations of *H.*
223 *numata* showed a remarkable uniformity, with the exception of a few individuals sharing some
224 variation with SE Brazil. This pattern is consistent with the population structure inferred using
225 microsatellite markers (Fig S1). Population structure revealed by PCA is in line with the admixture
226 analysis (Fig 2B). Individuals from the Atlantic population of *H. numata* clustered together to one
227 side of the first PCA axis, whereas all other individuals from all other populations clustered to the
228 other side. The second axis of the PCA separates individuals from French Guiana from the other
229 samples of the Amazon. This clustering was not found with Admixture (i.e. with $K=3$), suggesting
230 that the divergence between Amazonian populations is very reduced. In accordance, pairwise
231 genome-wide estimates of differentiation (F_{ST}) between *H. numata* populations showed elevated
232 values when comparing the Atlantic population to other populations, but very small values when
233 comparing pairs of Amazonian populations, even at a large distance (Fig 2C, Table S4). For
234 instance, the population from La Merced in Peru shows an $F_{ST}=0.032$ with the population from
235 French Guiana at a distance of 3019km, but an $F_{ST}=0.311$ (an order of magnitude higher) with the
236 Atlantic population at a similar distance. Isolation by distance among Amazonian populations of *H.*
237 *numata*, estimated using the proxy F_{ST}/km , shows a very different pattern to other species, with a
238 highly significantly shallower increase in F_{ST} with distance in *H. numata* compared to all other taxa
239 (Fig 2C, Table S4). By contrast, differentiation as a function of distance between Atlantic and
240 Amazonian populations of *H. numata* is close to what is observed in other species, and not
241 significantly different (see Supp. Text S1).

242

243 Analyses of genetic diversity show that all populations of *H. numata*, except those from the Atlantic
244 Forest, have a similar high genetic diversity (Fig 3A). By comparison, closely related *Heliconius*
245 taxa show significantly lower genetic diversity (Fig 3A). These patterns are similar to those
246 obtained using G-PhoCS to analyse the demographic histories in a phylogenetic context, where
247 Amazonian populations of *H. numata* show higher population sizes compared to the Atlantic
248 population (Fig 3B, Table S5). G-PhoCS analyses also show a demographic history in which gene
249 flow plays a crucial role (Table S6). For instance, our analyses show strong significant gene flow
250 right at the beginning of the divergence between *H. ismenius* and the other silvaniforms, as well as
251 in the divergence between *H. pardalinus* and *H. elevatus*. The effective population sizes inferred
252 from Atlantic genomes are one order of magnitude lower than that obtained using *H. numata*
253 populations from other localities (Fig 3A and Table S5). In our cladogram, the increase in *H.*
254 *numata* population size is restricted to the Amazonian branch, excluding Atlantic populations.

255

256

257 Discussion

258 Our results suggest that populations displaying inversion polymorphism in the *P* supergene in *H.*
259 *numata* also display distinctive population demography and gene flow. Differences in demographic
260 and differentiation regimes associated with structural variation at this locus are revealed when
261 comparing polymorphic populations of *H. numata* to closely-related monomorphic taxa, such as (1)
262 peripheral populations of *H. numata*, (2) sister taxa, and (3) inferred ancestral lineages. This
263 suggests that the existence of a mimicry supergene controlling polymorphism in *H. numata* is
264 associated, in time and in space, with major differences in population biology. We hypothesize this
265 to be due to a change in the balancing selection regime due to heterozygote advantage (Jay et al.
266 2021) and in the associated evolution of disassortative mating (Chouteau et al. 2017) following the
267 onset of inversion polymorphism, causing direct effects on ecological parameters such as gene flow,
268 immigration success and effective population size.

269
270 Our analyses show large-scale variation in genetic diversity among closely related taxa in this clade
271 of *Heliconius* butterflies. Within *H. numata*, the genetic diversity of polymorphic Amazonian
272 populations is one to two orders of magnitude higher than the diversity found in populations from
273 the Atlantic Forest. Generally, Amazonian populations of *H. numata* harbour the highest genetic
274 diversity in the entire *melpomene*/silvaniform clade, which contrasts with the low diversity found in
275 the most closely related taxa such as *H. ismenius* or *H. besckei*. Inferring historical demography
276 during the diversification of the *H. numata* lineage reveals that the large effective population size in
277 that species is only associated with the branch representing polymorphic, Amazonian *H. numata*
278 populations, while internal branches all show very low diversity estimates. This suggests that
279 ancestral monomorphic populations of *H. numata* were similar in their diversity parameters to
280 current sister species *H. ismenius* populations, or to current peripheral Atlantic *H. numata*
281 populations. Although low-diversity lineages could have lost diversity due to recent events such as
282 strong bottlenecks, the distribution of parameters across lineages rather suggests that the
283 Amazonian populations of *H. numata* underwent a dramatic increase in effective population size
284 posterior to their split with Central American (*H. ismenius*) and Atlantic populations. The
285 Amazonian branch of the *H. numata* radiation is characterized by the long-term maintenance of
286 inversion polymorphism, triggered by the introgression of a chromosomal inversion about 2.2 Ma
287 ago. Therefore, the major shift in demography between Amazonian and Atlantic populations indeed
288 appears associated with the occurrence of inversion polymorphism, even though the lack of
289 replication of this event impedes firmly establishing causality here.

290
291 Another striking result is the low genetic structure displayed by *H. numata* across the Amazon, with
292 all Amazonian and Guianese populations forming a single genetic cluster. Only Atlantic
293 populations stand out and display high differentiation with other *H. numata* from the rest of the
294 range. French Guiana and Peruvian populations, separated by over 3000 km across the Amazon,
295 are remarkably genetically similar compared to pairs of populations at comparable distances in
296 other species, and show similar differentiation as pairs of *H. numata* populations taken at short
297 distances. *H. numata* populations from the Amazon show significantly lower isolation by distance
298 than all other taxa, as measured by the change in F_{ST} across distance (F_{ST}/km) (Fig. 2C), with a very
299 distinctive, flat slope of isolation by distance. The only exception is found when comparing
300 Amazonian populations with Atlantic populations of Brazil, displaying a level of differentiation in
301 line with that of pairs of populations at similar distances within other taxa.

302

303 Effective population size is affected by census size, mating system, and the force and type of
304 selection acting on traits (Charlesworth 2009). Selection is often viewed as a force only affecting
305 the genetic variation around specific, functional loci in the genome, but it may also affect whole
306 genome diversity, for instance when its action is sufficient to modify local demography or mating
307 patterns. In *H. numata*, morphs and therefore inversion genotypes show disassortative mate
308 preferences, i.e., they preferentially mate with individuals carrying different chromosome types
309 (Chouteau et al. 2017). Disassortative mating enhances heterozygosity and the mating success of
310 individuals expressing rare alleles (negative frequency dependence) (Knoppien 1985; Hedrick et al.
311 2018). Consequently, immigrants expressing rare, recessive alleles have a mating advantage in *H.*
312 *numata*. Their recessive effect on wing pattern lets them escape negative selection caused by their
313 inadequate mimicry patterns. Disassortative mating associated with the supergene should therefore
314 bring an advantage to immigrant genomes in LD with recessive supergene alleles, enhancing
315 genome-wide gene flow. This effective migration regime is quite different to that observed in other
316 mimetic taxa such as *Heliconius melpomene* or *H. erato*, in which mimicry variation is controlled
317 by multiple loci with diverse dominance patterns. In those taxa, hybrid offspring display
318 recombinant patterns breaking down mimicry, even after multiple generations of backcrossing, and
319 pure forms mate assortatively with respect to wing pattern (McMillan et al. 1997, Mallet et al. 1998,
320 Jiggins et al. 2001); both processes select against mimetic variants migrating from adjacent areas
321 with distinct warning patterns. In *H. numata*, the evolution of a polymorphic mimicry supergene
322 and disassortative mate preferences could therefore explain the relative lack, compared to other
323 *Heliconius* taxa, of differentiation among polymorphic populations, even across large distances.
324 Furthermore, enhanced gene flow could also cause an increase in effective population size estimates
325 (Slatkin 1987), putatively explaining why polymorphic populations of *H. numata* harbour the
326 highest genetic diversity, and display the highest N_e estimates in the entire *melpomene*-silvaniform
327 clade of *Heliconius*.

328
329 Alternative processes may of course contribute to the observed patterns. Amazonian and Atlantic
330 populations may differ in other aspects that could also result in differences in genetic diversity.
331 Habitat availability and structure may be different, possibly entailing differences in the maintenance
332 of diversity. The Atlantic Forest is vast in area, but may represent a smaller biome compared to the
333 Amazon, and is isolated from the bulk of the range of *H. numata*, which could result in a population
334 ecology displaying characteristics of peripheral populations with smaller effective population sizes
335 (Eckert et al. 2008). The other *Heliconius* species in the clade have much in common with *H.*
336 *numata* in terms of habitat and general ecology, yet their niche and life-history specificities and
337 their phylogenetic histories may result in consistent differences with the polymorphic *H. numata*
338 populations. All those specificities may contribute to the observed pattern in which polymorphic
339 Amazonian populations of *H. numata* display high effective population size and a lack of
340 geographic structure in genome-wide genetic variation. Yet this pattern of variation correlates
341 parsimoniously with the evolution of a supergene causing disassortative mating in certain *H.*
342 *numata* populations, which provides an elegant mechanism explaining their differences with extant
343 and ancestral closely-related lineages. However, we cannot rule out a role for conjectural
344 differences in ecology and geography with all other taxa.

345
346 In conclusion, our results show a remarkable contrast in the demography and differentiation of
347 populations within the Amazonian range of *H. numata* compared to closely related taxa and
348 ancestral lineages, as well as with other taxa in the *melpomene*/silvaniform clade. Although those

349 populations may differ in many uncharacterized ways from all other taxa, one known and consistent
350 difference is the maintenance of inversion polymorphism associated with a specific mating system
351 and selection regime in Amazonian *H. numata*. This distinctiveness of the only widely polymorphic
352 populations in the clade is consistent with the hypothesis that the evolution of a supergene
353 maintained by balancing selection represents a major transition in this lineage, triggering changes in
354 genome-wide patterns of diversity and population ecology over the last 2 million years since its
355 formation. If this hypothesis is correct, the evolution of a locus under balancing selection may
356 therefore feed-back on population ecology and diversification, and consequently on speciation.
357 More work on the determinants of variation in effective population sizes in the *Heliconius* genus is
358 needed to determine the precise impact of the supergene on demography of *H. numata*. We believe
359 that our results emphasize a potential link between genomic architecture, selection and demography,
360 and should inspire future theoretical and modelling studies. Finally, the eco-evolutionary feedbacks
361 between changes in genomic architecture and the ecological parameters of populations are well-
362 known when considering self-incompatibility loci in plants, but may be more common than
363 previously thought. Indeed, our result suggests that balancing selection maintaining structural
364 polymorphisms affecting life-history traits may have a profound influence on species ecology.

365

366 **Contributions:**

367 MARdC, PJ and MJ designed the study and wrote the manuscript. BH, AVLF, TTT, RRR, KLSB
368 provided the Atlantic samples. CS provided the Colombian samples. MARdC and PJ performed
369 genomic analyses with input from AW. MARdC, PJ, MJ, FPP and MC collected the Peruvian and
370 Ecuadorian samples. MC performed microsatellite analyses and organized fieldworks and butterfly
371 rearing. All authors contributed to editing the manuscript.

372 **Aknowledgements:**

373 This work was funded by grants HYBEVOL (ANR-12-JSV7-0005) and Supergene (ANR-18-
374 CE02-0019-01) from the Agence Nationale de la Recherche and European Research Council Grant
375 MimEvol (StG-243179). We acknowledge the Genotoul and the Montpellier Bioinformatics
376 Biodiversity (MBB) platforms for providing us with calculation time. We thank Dr. Vitor Becker, at
377 the Serra Bonita Reserve (Bahia), Alexandre Soares, at the MN/UFRJ (Rio de Janeiro) and Dr.
378 Marcelo Duarte at the MZ/USP (Sao Paulo) for their contribution to the collection of butterflies in
379 Brazil. Field collections in Colombia were conducted under permit no. 530 issued by the Autoridad
380 Nacional de Licencias Ambientales (ANLA). We are grateful to Marianne Elias, Violaine Llaurens,
381 Quentin Rougemont for comments and discussions. AVLF acknowledges support from Fundação
382 de Amparo à Pesquisa do Estado de São Paulo – (FAPESP) (Biota-Fapesp grants 2011/50225-3,
383 2013/50297-0) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)
384 (421248/2017-3 and 304291/2020-0). KLSB acknowledges the financial support of FAPESP
385 Process # 2012/16266-7. Brazilian specimens are registered under SISGEN (A701768).

386

387 **Data availability:**

388 The raw sequence data were deposited in NCBI SRA and accession numbers are indicated in
389 Supplementary table 3.

390

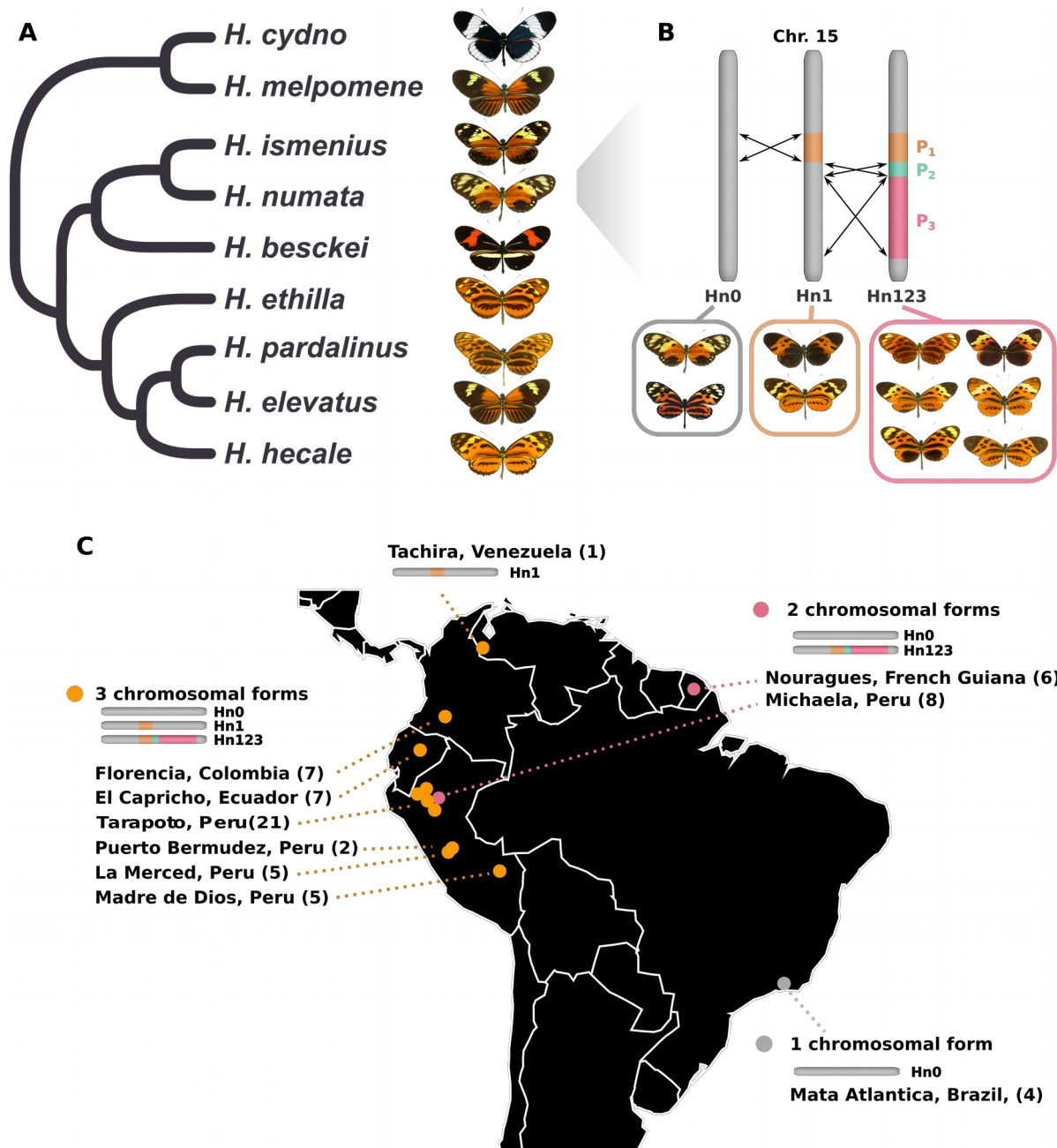
391 **References**

- 392 1. Alexander DH, Novembre J, Lange K (2009). Fast model-based estimation of ancestry in
393 unrelated individuals. *Genome Research* **19**:1655-1664.

- 394 2. Beichmann AC, Huerta-Sanchez E, Lohmueller KE (2018). Using Genomic Data to Infer
395 Historic Population Dynamics of Nonmodel Organisms. *Annual Review of Ecology,*
396 *Evolution, and Systematics* **49**:433–56
- 397 3. Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic*
398 *Acids Research* **27**:573-580.
- 399 4. Brown KS (1979). Ecologia Geográfica e Evolução nas Florestas Neotropicais. – Univ.
400 Estadual de Campinas, Campinas, Brazil.
- 401 5. Brown KS, Benson WW (1974). Adaptive polymorphism associated with multiple müllerian
402 mimicry in *Heliconius numata* (Lepid.: Nymph.). *Biotropica* **6**:205–228
- 403 6. Brown KS, Mielke OHH. 1972. The Heliconians of Brazil (Lepidoptera: Nymphalidae). Part
404 II. Introduction and general comments, with a supplementary revision of the tribe.
405 *Zoologica, New York*, **57**:1–40.
- 406 7. Charlesworth B (2009) Fundamental concepts in genetics: effective population size and
407 patterns of molecular evolution and variation. *Nature Reviews Genetics* **10**:195-205.
- 408 8. Chouteau M, Arias M, Joron M (2016). Warning signals are under positive frequency-
409 dependent selection in nature. *Proceedings of the National Academy of Sciences of the USA*
410 **113**:2164–2169.
- 411 9. Chouteau M, Llaurens V, Piron-Prunier F, Joron M. (2017). Polymorphism at a mimicry
412 supergene maintained by opposing frequency-dependent selection pressures. *Proceedings of*
413 *the National Academy of Sciences of the USA* **114**: 8325–8329.
- 414 10. Davey JW, Chouteau M, Barker SL, Maroja L, Baxter SW, Simpson F, et al. (2016). Major
415 Improvements to the *Heliconius melpomene* Genome Assembly Used to Confirm 10
416 Chromosome Fusion Events in 6 Million Years of Butterfly Evolution. *G3* **6**:695–708.
417 doi:10.1534/g3.115.023655
- 418 11. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del
419 Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY,
420 Cibulskis K, Gabriel SB, Altshuler D, Daly MJ (2011). A framework for variation discovery
421 and genotyping using next-generation DNA sequencing data. *Nature Genetics* **43**:491–498.
- 422 12. Eckert CG, Samis KE, Loughheed SC (2008). Genetic variation across species' geographical
423 ranges: the central–marginal hypothesis and beyond. *Molecular Ecology* **17**:1170–1188.
- 424 13. Edelman NB, Frandsen PB, Miyagi M, Clavijo B, Davey J, Dikow RB, García-Accinelli G,
425 Van Belleghem SM, Patterson N, Neafsey DE, Challis R, Kumar S, Moreira GRP, Salazar
426 C, Chouteau M, Counterman BA, Papa R, Blaxter M, Reed RD, Dasmahapatra KK,
427 Kronforst M, Joron M, Jiggins CD, McMillan WO, Di Palma F, Blumberg AJ, Wakeley J,
428 Jaffe D, Mallet J (2019). Genomic architecture and introgression shape a butterfly radiation.
429 *Science* **366**:594-599.
- 430 14. Emsley MG 1965. Speciation in *Heliconius* (Lep., Nymphalidae): morphology and
431 geographic distribution. *Zoologica, New York* **50**:191–254.
- 432 15. Faria R, Johannesson K, Butlin RK, Westram AM (2019). Evolving inversions. *Trends in*
433 *Ecology & Evolution* **34**:239-248.
- 434 16. Freedman AH, Gronau I, Schweizer RM, Ortega-Del Vecchyo D, Han E, et al. (2012)
435 Genome Sequencing Highlights the Dynamic Early History of Dogs. *PLoS Genetics*
436 **10**:e1004016.

- 437 17. Glinka S, Ometto L, Mousset S, Stephan W, De Lorenzo D (2003) Demography and natural
438 selection have shaped genetic variation in *Drosophila melanogaster*: a multi-locus approach.
439 *Genetics* **165**:1269-1278.
- 440 18. Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A (2011). Bayesian inference of ancient
441 human demography from individual genome sequences. *Nature Genetics* **43**:1031-1034.
- 442 19. Hackenberg M, Previti C, Luque-Escamilla PL, Carpena P, Martínez-Aroza J, Oliver JL.
443 (2006) CpGcluster: a distance-based algorithm for CpG-island detection. *BMC*
444 *Bioinformatics* **7**:446.
- 445 20. Hedrick PW, Tuttle EM, Gonser RA (2018) Negative-Assortative Mating in the White-
446 Throated Sparrow. *Journal of Heredity* **109**:223-231.
- 447 21. *Heliconius* Genome Consortium (2012). Butterfly genome reveals promiscuous exchange of
448 mimicry adaptations among species. *Nature* **487**: 94–8.
- 449 22. Jay P, Whibley A, Frézal L, Rodríguez de Cara MÁ, Nowell RW, Mallet J, Dasmahapatra
450 KK, Joron M. (2018). Supergene evolution triggered by the introgression of a chromosomal
451 inversion. *Current Biology* **28**:1839-1845.
- 452 23. Jay P, Chouteau M, Whibley A, Bastide H, Parrinello H, Llaurens V, Joron M. (2021).
453 Mutation load at a mimicry supergene sheds new light on the evolution of inversion
454 polymorphisms. *Nature Genetics* **53**:288-293.
- 455 24. Jiggins C, Naisbit R, Coe R, Mallet J 2001. Reproductive isolation caused by colour pattern
456 mimicry. *Nature* **411**:302–305.
- 457 25. Joron M, Wynne IR, Lamas G, Mallet J (1999) Variable selection and the coexistence of
458 multiple mimetic forms of the butterfly *Heliconius numata*. *Evol Ecol* **13**: 721–754.
- 459 26. Joron M, Papa R, Beltran M, Chamberlain N, Mavarez J, et al. (2006) A conserved
460 supergene locus controls colour pattern diversity in *Heliconius* butterflies. *PLoS Biology*
461 **4**:e303
- 462 27. Joron M, Frezal L, Jones RT, Chamberlain NL, Lee SF, Haag CR, Whibley A, Becuwe M,
463 Baxter SW, Ferguson L, Wilkinson PA, Salazar C, Davidson C, Clark R, Quail MA,
464 Beasley H, Glithero R, Lloyd C, Sims S, Jones MC, Rogers J, Jiggins CD, French-Constant
465 RH (2011). Chromosomal rearrangements maintain a polymorphic supergene controlling
466 butterfly mimicry. *Nature* **477**:203–206.
- 467 28. Knoppien P (1985) Rare male mating advantage: a review. *Biological Reviews* **60**:81-117.
- 468 29. Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology and*
469 *Evolution* **17**:183-189.
- 470 30. Lenth RV (2016). Least-Squares Means: The R Package lsmeans. *Journal of Statistical*
471 *Software* **69**:1-33.
- 472 31. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin
473 R; 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map
474 format and SAMtools. *Bioinformatics* **25**:2078-2079.
- 475 32. Lunter G, Goodson M (2011). Stampy: a statistical algorithm for sensitive and fast mapping
476 of Illumina sequence reads. *Genome Research* **21**:936-939.
- 477 33. Maisonneuve L., Chouteau M, Joron M, Llaurens V (2021). Evolution and genetic
478 architecture of disassortative mating at a locus under heterozygote advantage. *Evolution*
479 **75**:149-165.

- 480 34. Mallet J, McMillan W, Jiggins C (1998). Estimating the mating behavior of a pair of
481 hybridizing *Heliconius* species in the wild. *Evolution* **52**:503–510.
- 482 35. Martin SH, Möst M, Palmer WJ, Salazar C, McMillan WO, Jiggins FM, Jiggins CD (2016).
483 Natural Selection and Genetic Diversity in the Butterfly *Heliconius melpomene*. *Genetics*
484 **203**:525-541.
- 485 36. Mitchell-Olds T, Willis JH, Goldstein DB (2007). Which evolutionary processes influence
486 natural genetic variation for phenotypic traits? *Nature Reviews Genetics* **8**:845–856.
- 487 37. Muers, M (2009) Separating demography from selection, *Nature Reviews Genetics* **10**:280–
488 281.
- 489 38. Murray GGR, Soares AER, Novak BJ, Schaefer NK, Cahill JA, Baker AJ, Demboski JR,
490 Doll A, Da Fonseca RR, Fulton TL, Gilbert MTP, Heintzman PD, Letts B, McIntosh G,
491 O'Connell BL, Peck M, Pipes ML, Rice ES, Santos KM, Sohrweide AG, Vohr SH, Corbett-
492 Detig RB, Green RE, Shapiro B (2017). Natural selection shaped the rise and fall of
493 passenger pigeon genomic diversity. *Science* **358**:951-954.
- 494 39. McMillan W, Jiggins C, Mallet J (1997). What initiates speciation in passion-vine
495 butterflies? *Proceedings of the National Academy of Sciences of the USA* **94**:8628–8633.
- 496 40. Nadeau NJ, Pardo-Diaz C, Whibley A, Supple M A, Saenko SV, Wallbank RWR *et al.*
497 (2016). The gene *cortex* controls mimicry and crypsis in butterflies and moths. *Nature*,
498 **534**:106–110.
- 499 41. Nielsen, R., Hubisz, M.J., Hellmann, I., Torgerson, D., Andres, A.M., Albrechtsen, A.,
500 Gutenkunst R, Adams MD, Cargill M, Boyko A, Indap A, Bustamante CD, and Clark AG
501 (2009). Darwinian and demographic forces affecting human protein coding genes. *Genome*
502 *Research* **19**:838–849.
- 503 42. Patterson N, Price AL, Reich D (2006) Population Structure and Eigenanalysis. *PLoS*
504 *Genetics* **2**: e190.
- 505 43. Rosser N, Phillimore AB, Huertas B, Willmott KR, Mallet J (2012) Testing historical
506 explanations for gradients in species richness in heliconiine butterflies of tropical America.
507 *Biological Journal of the Linnean Society* **105**:479–497.
- 508 44. Schiffels S, Durbin R (2014) Inferring human population size and separation history from
509 multiple genome sequences. *Nature Genetics* **46**:919-925.
- 510 45. Saenko SV, Chouteau M, Piron-Prunier F, Blugeon C, Joron M, Llaurens V (2019)
511 Unravelling the genes forming the wing pattern supergene in the polymorphic butterfly
512 *Heliconius numata*. *EvoDevo* **10**:1-12.
- 513 46. Sheppard PM, Turner JRG, Brown KS, Benson WW, Singer MC (1985) Genetics and the
514 evolution of Mullerian mimicry in *Heliconius* butterflies. *Philosophical Transactions of*
515 *the Royal Society of London, B Biological Sciences* **308**: 433–610
- 516 47. Slatkin M (1987) Gene flow and the geographic structure of natural populations. *Science*
517 **236**:787-792
518

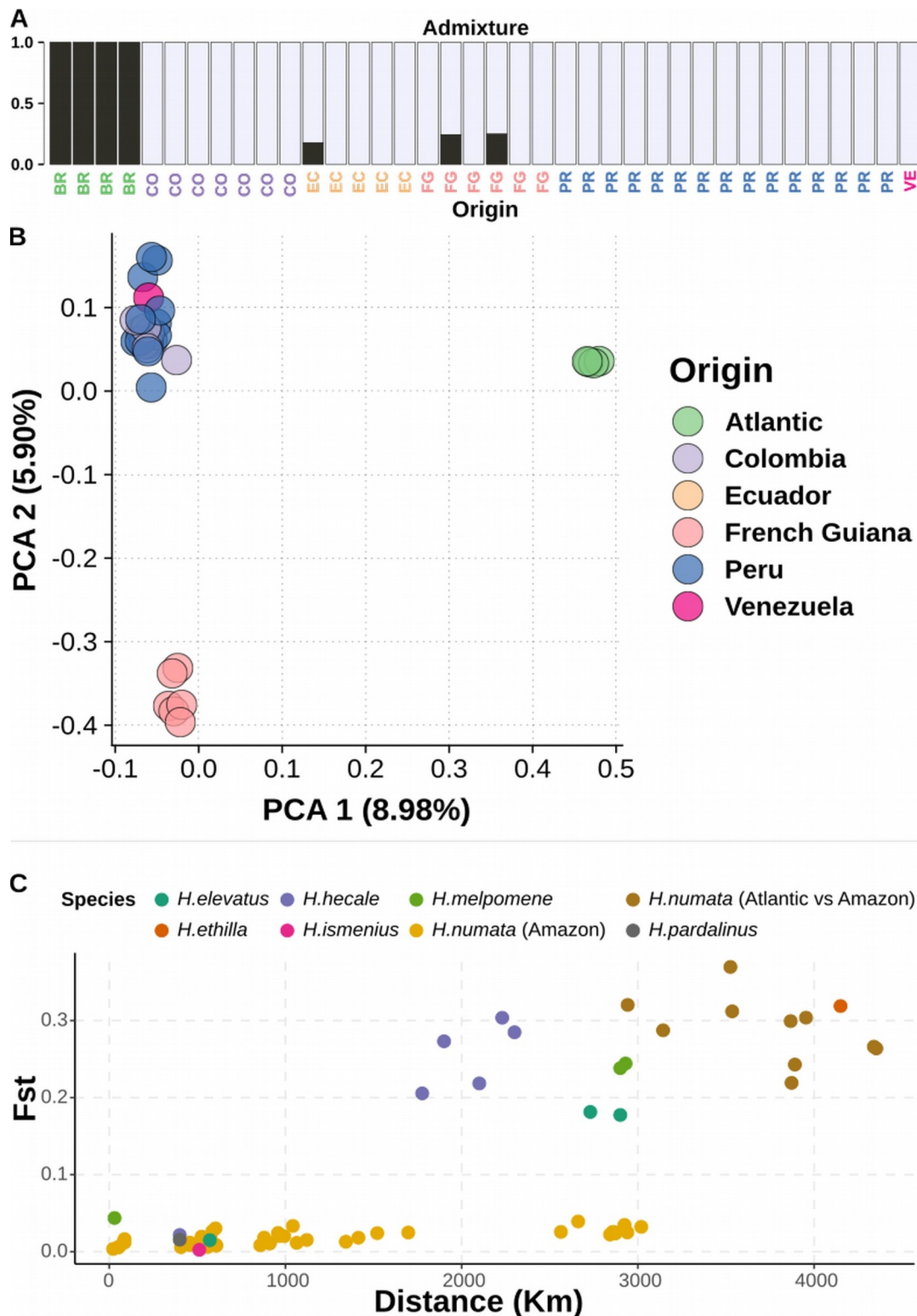


520 **Figure 1 | Genetic and population structure at the P supergene.**

521 **A.** Schematic phylogeny of the sampled species. It includes all members of the silvaniform clade
522 and two outgroups, *H. melpomene* and *H. cydno*.

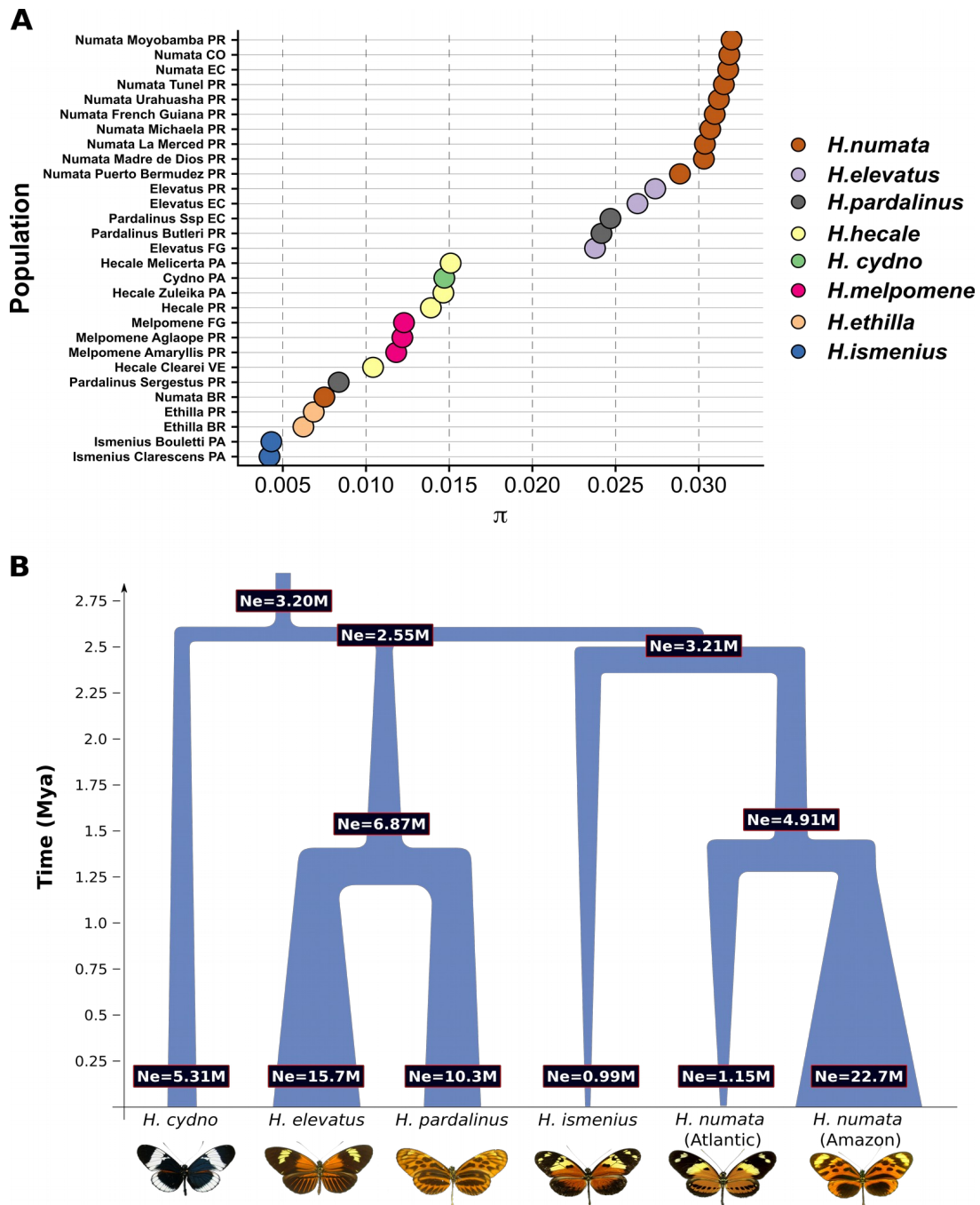
523 **B.** Schematic description of the genetic structure of the P supergene. Three chromosomal
524 arrangements coexist in *H. numata* and are associated with different morphs.

525 **C.** Origin of *H. numata* specimens used for analyses and distribution of chromosome arrangements
526 across the neotropics. Numbers in brackets indicate sampled specimens in each locality (the
527 Tarapoto population lumps several neighbouring subsamples on the map)



528 **Figure 2 | *H. numata* is characterised by low population structure.**

529 **A.** Admixture plot for *H. numata*. The optimal cluster number for *H. numata* is two, and it splits *H.*
530 *numata* into two categories, whereas they come from Atlantic forest or the Amazon. BR=Brazil
531 (Atlantic), PR=Peru, VE=Venezuela, CO=Colombia, EC=Ecuador, FG=French Guiana. **B.**
532 Principal component analysis computed on whole genome SNP. **C.** Relationship between genetic
533 differentiation (Fst) and geographical distance. Fst is measured between morphs/populations of the
534 same species. *H. numata* populations from the Amazon show low isolation by distance when
535 compared to related species.



537 **Figure 3 | Variation in present and past effective population size in *Heliconius* species**

538 **A.** Variation in π in several *Heliconius* populations, showing higher genetic diversity in *H. numata*
 539 populations from the Amazon than other taxa. Population names indicates their origin as in Figure 2
 540 (e.g. PR=Peru), with the addition of PA=Panama. The *H. numata* population with a lowest diversity
 541 is the one from the Atlantic forest (Brazil). **B.** Schematic representation of Gphocs results
 542 (presented in table S5-6). Gene flow was modelled but not represented graphically for clarity.
 543 showing that Amazonian populations of *H. numata*, which have the P supergene, show a dramatic
 544 increase in population size posterior to their split with the Atlantic populations of Brazil, which lack
 545 the supergene.

546 **List of Supplementary Materials:**

547 Table S1-6

548 Fig S1-2

549 Text S1

550