



HAL
open science

Supergene Evolution Triggered by the Introgression of a Chromosomal Inversion

Paul Jay, Annabel Whibley, Lise Frézal, María Ángeles Rodríguez de Cara, Reuben W Nowell, James Mallet, Kanchon Dasmahapatra, Mathieu Joron

► **To cite this version:**

Paul Jay, Annabel Whibley, Lise Frézal, María Ángeles Rodríguez de Cara, Reuben W Nowell, et al.. Supergene Evolution Triggered by the Introgression of a Chromosomal Inversion. *Current Biology - CB*, 2018, 28 (11), pp.1839-1845.e3. 10.1016/j.cub.2018.04.072 . hal-02324488

HAL Id: hal-02324488

<https://hal.science/hal-02324488v1>

Submitted on 9 Jan 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

1 Title : **Supergene evolution triggered by the introgression of a**
2 **chromosomal inversion.**

3 **Authors:** Paul Jay^{1*}, Annabel Whibley², Lise Frezal³, María Ángeles Rodríguez de Cara⁴,
4 Reuben W. Nowell⁵, James Mallet⁶, Kanchon K. Dasmahapatra⁷, Mathieu Joron^{4*}
5

6 **Affiliations:**

7 ¹ CEFE, Univ Montpellier, CNRS, Univ Paul Valéry Montpellier 3, EPHE, IRD, Montpellier,
8 France

9 ² Department of Cell and Developmental Biology, John Innes Centre, Norwich, NR4 7UH, UK

10 ³ IBENS, Département de Biologie, Ecole Normale Supérieure, CNRS, Inserm, PSL Research
11 University, Paris

12 ⁴ CEFE, CNRS, Univ Montpellier, Univ Paul Valéry Montpellier 3, EPHE, IRD, Montpellier,
13 France

14 ⁵ Department of Life Sciences, Imperial College London, Silwood Park Campus, Buckhurst
15 Road, Ascot, Berkshire, SL5 7PY, UK

16 ⁶ Department of Organismic and Evolutionary Biology, Biology Laboratories, Harvard
17 University, 16 Divinity Avenue, CAMBRIDGE, MA 02138, USA

18 ⁷ Department of Biology, University of York Wentworth Way, Heslington YO10 5DD

19 ***Correspondence:**

20 PJ: paul.jay@cefe.cnrs.fr

21 MJ: mathieu.joron@cefe.cnrs.fr (lead contact)

22 **Summary:**

23 Supergenes are groups of tightly linked loci whose variation is inherited as a single Mendelian
24 locus and are a common genetic architecture for complex traits under balancing selection [1–
25 8]. Supergene alleles are long-range haplotypes with numerous mutations underlying distinct
26 adaptive strategies, often maintained in linkage disequilibrium through the suppression of
27 recombination by chromosomal rearrangements [1,5,7–9]. However, the mechanism governing
28 the formation of supergenes is not well understood, and poses the paradox of establishing
29 divergent functional haplotypes in face of recombination. Here, we show that the formation of
30 the supergene alleles encoding mimicry polymorphism in the butterfly *Heliconius numata* is
31 associated with the introgression of a divergent, inverted chromosomal segment. Haplotype
32 divergence and linkage disequilibrium indicate that supergene alleles, each allowing precise
33 wing-pattern resemblance to distinct butterfly models, originate from over a million years of
34 independent chromosomal evolution in separate lineages. These “superalleles” have evolved
35 from a chromosomal inversion captured by introgression and maintained in balanced
36 polymorphism, triggering supergene inheritance. This mode of evolution involving the
37 introgression of a chromosomal rearrangement is likely to be a common feature of complex
38 structural polymorphisms associated with the coexistence of distinct adaptive syndromes. This
39 shows that the reticulation of genealogies may have a powerful influence on the evolution of
40 genetic architectures in nature.

41 **Results**

42 How new beneficial traits which require more than one novel mutation emerge in natural
43 populations is a long-standing question in biology [10–12] . Supergenes control alternative
44 adaptive strategies that require the association of multiple co-adapted characters, and have
45 evolved repeatedly in many taxa under balancing selection. Examples include floral
46 heteromorphy determining alternative pollination strategies [1] , butterfly mimicry of alternative
47 wing-pattern and behaviours of toxic models [2–4], contrasting mating tactics in several birds
48 [5,6] , and alternative social organization in ant colonies [7]. In most documented cases, the
49 maintenance of character associations is mediated by polymorphic rearrangements, such as
50 inversions, which suppress local recombination and allow the differentiated supergene alleles
51 to persist [1,5,7–9]. However, the build-up of differentiated haplotypes from initially
52 recombining loci is poorly understood [13,14]. Recombination is necessary to bring into linkage
53 mutations that arise on different haplotypes, but also acts to break down co-adapted
54 combinations. While inversions may capture epistatic alleles at adjacent loci, this requires
55 adaptive polymorphism at both loci prior to the rearrangement. Furthermore, linkage
56 disequilibrium around functional mutations under balancing selection persists only over short
57 evolutionary times [15]. The few models of supergene evolution [10,16] do not readily yield the
58 conditions for the formation of differentiated haplotypes or the evolutionary trajectory of
59 functional genetic elements within rearranged non-recombining regions after the initial
60 structural variation.

61 To understand allelic evolution in supergenes, we studied Amazonian populations of the
62 butterfly *Heliconius numata*, in which up to seven distinct wing-pattern morphs coexist (Figure
63 1A), each one matching to near perfection the colours and shapes of other toxic Lepidoptera
64 (Heliconiinae, Danainae, Pericopiinae) [12]. This balanced polymorphism is controlled by a
65 supergene locus (*P*) associated with an inversion polymorphism [12] that captures multiple

66 genetic loci controlling wing-pattern variation in butterflies and moths [4,17–20] and allows
67 multiple wing elements to be inherited as a single Mendelian character. The ancestral
68 chromosomal arrangement, called *Hn0*, is associated with the recessive supergene allele [21]
69 which controls the widely distributed morph *silvana*. All other characterized supergene alleles,
70 grouped into a family of alleles called *Hn1*, determine a diversity of mimetic morphs dominant
71 to *silvana* and associated with the 400-kb inversion P_1 (Figure 1A; Ref. [9,21]. A subset of
72 these alleles is associated with additional rearrangements (P_2) in adjacent positions [12]. The
73 emergence of the P supergene architecture is therefore associated with the introduction of
74 inversion P_1 , maintained at intermediate frequency by balancing selection and followed by
75 adjacent rearrangements. To explore the origin and evolution of the P supergene, we thus
76 tracked the history of inversion P_1 . This inversion forms a well-differentiated haplotype distinct
77 from the ancestral haplotype along its entire length (Figure 4B), and with extreme values of
78 linkage disequilibrium (LD) [12]. Inversion P_1 therefore stands as a block of up to 7000
79 differentiated SNPs along its 400 kb length, associated with supergene evolution, adaptive
80 diversification and dominance variation.

81 *Heliconius numata* belongs to the so-called silvaniform clade of ten species which diverged ca.
82 4 My ago from its sister clade (Figure 1b ; Figure 2a; Figure S1; Ref. [22]). The Heliconius, and
83 silvaniform members particularly are known to be highly connected by gene flow, and to
84 notably exchange wing pattern loci [18,23–26]. To investigate the history of inversion P_1 , we
85 surveyed the presence of this inversion in other species of the clade. PCR amplification of
86 inversion breakpoints showed that inversion P_1 was polymorphic in *H. numata* (*Hn*) across its
87 Amazonian range, and was also found fixed in all population of *H. pardalinus* (*Hp*), a non-sister
88 species deeply divergent from *H. numata* within the silvaniform clade (Figure 1B and Table
89 S3). All other taxa including the sister species of *H. numata* and that of *H. pardalinus* were
90 positive only for markers diagnostic of the ancestral gene order. Furthermore, a 4kb duplication

91 associated with P_1 in *Hn* was also found in whole genome sequence datasets for all *Hp*
92 individuals and no other taxon (Figure 1B and Table S2). Breakpoint homology and similar
93 molecular signatures in *Hp* and *Hn* are thus consistent with a single origin of this inversion.
94 This sharing of P_1 between non-sister species could be due to the differential fixation of an
95 ancient polymorphic inversion (incomplete lineage sorting, ILS), or to a secondary transfer
96 through introgression.

97 To clarify whether this sharing between *Hp* and *Hn* is a rare anomaly, specifically associated
98 with the supergene locus, or is a common feature that is also found elsewhere in the genome,
99 we estimated the excess of shared derived mutations between sympatric *Hp* and *Hn*, relative
100 to an allopatric control, *H. ismenius* (*Hi*, sister species of *Hn*), using the *fd* statistic [27]. We
101 estimated that a significant 6.2 % of the genome was shared via gene flow between *Hn* and *Hp*
102 (mean *fd* = 0.062*. Figure 3 and Figure S2C), consistent with a general signal of genome-wide
103 gene flow between *Hn* and other species within the silvaniform clade (Figure S2) and between
104 other *Heliconius* species [23]. When *fd* is estimated using *Hn* specimens homozygous for
105 inversion P_1 (*Hn1*), the supergene scaffold is associated with a strong peak of shared derived
106 mutations between *Hn* and *Hp* (mean= 0.38, 95% interval 0,34-0.41, Figure 3, blue arrow).
107 This is not observed between *Hn1* and other silvaniforms (Figure S2) nor when using *Hn*
108 specimens homozygous for the ancestral supergene arrangement (*Hn0* ; Figure S2C).

109 Between *Hn1* and *Hp*, the entire P_1 inversion shows a high level of *fd*, which drops to
110 background levels precisely at inversion breakpoints (Figure 4C). *Hn1* and *Hp* therefore share
111 a block of derived mutations associated with the inversion.

112 Contrary to estimates from the whole genome, a local excess of *fd*, denoting an local excess of
113 derived mutation between two taxa, may be due to incomplete lineage sorting or to gene flow.
114 We estimated the divergence times of *Hn1* and *Hp* within and outside of inversion P_1 to
115 determine the cause of the local excess of *fd* at the supergene. The unique ancestor of

116 inversion P_1 in *Hp* and *Hn1* was estimated to be 2.30 My old (95% interval 1.98-2.63 Mya,
117 Figure 4D; Pink triangles in Figure 2), significantly more recent than the divergence time of the
118 rest of the genome (3,59 Mya; 95% interval 3.37-3.75 Mya; Figure 4D), which indicates that the
119 inversion was shared by gene flow among lineages well after their split. This introgression can
120 be dated to an interval between the time to the most recent common ancestor (TMRCA) of *Hp*
121 and *Hn1* inversions (*i.e.* 2.30 Mya) and the TMRCA of all *Hn1* inversions (2.24 Mya, 95 %
122 interval 1.89-2.59 Mya, Figure S3D ; Orange triangle in Figure 2), *i.e.* about 1.30 My after *Hp*-
123 *Hn* speciation. We then estimated the age of the inversion considering that its occurrence also
124 induce the 4kb duplication we detected. We identified the two sequences of the duplicated
125 region associated with the inversion in an *Hn1* BAC library and in an *Hn1* genome assembly,
126 and estimated their divergence time. We found that the duplication and most probably the
127 inversion occurred 2.41 Mya (95% interval 1.96-2.71 Mya). This indicate that inversion P_1 may
128 have spread between lineage *Hp* and *Hn* shortly after the its occurrence.

129 To determine the direction of introgression, we surveyed the position of the sister species to
130 *Hn* (*Hi*) and to *Hp* (*H. elevatus*, *He*) in phylogenies computed along the supergene scaffold and
131 in other regions of the genome. The genome as a whole and regions flanking the inversion all
132 show a similar topology to the one found by Kozak *et al.* [22], with expected sister relationships
133 of *Hi* and *Hn* and of *He* and *Hp* (Figure 2A and Figure S1). Evaluating the support for each
134 possible topology among the five informative taxa (*Hn0*, *Hn1*, *Hi*, *Hp*, *He*) using Twisst [28]
135 confirmed the consistent support for the separation of (*Hn*, *Hi*) and (*Hp*, *He*) clades despite a
136 high level of incomplete lineage sorting within each clade (Figure 4A and Figure S4). By
137 contrast, the inversion P_1 shows strong support for topologies that group *Hn1* with *Hp*, and
138 major topology changes coincide with inversion breakpoints (Figures 2B and 4A and Figure
139 S4), consistent with a single origin of the inversion. Within the inversion, the highest support
140 consistently goes to *Hn1* grouping within (*Hp*, *He*) and away from (*Hn*, *Hi*) (Figure 4C, topology

141 2), indicating an introgression from *Hp* to *Hn*. This conclusion is robust to the species used as
142 sister groups to *Hn* or *Hp* (Figure S4D-G). Alternative topologies (3 and 4) are also found in
143 relatively high proportions in the interval ~650-850kb, presumably owing to high levels of
144 incomplete lineage sorting at the clade level in this region, or ancient gene flow among other
145 species of the clade. Supporting these interpretations, topology analysis with taxa unaffected
146 by *Hn1-Hp* introgression (for instance using *Hn0* and replacing *Hp* with a closely related
147 species, *H. hecale*) still showed the same pattern of unresolved phylogenetic signal in this
148 interval between the two major branches of the clade (*Hp-He-H. hecale* vs. *Hn-Hi*) (Figure
149 S4H-I). This suggests that the mixed phylogenetic signal found in this interval is independent of
150 the introgression. Overall, our results show that the inversion P_1 most likely occurred in *Hp*
151 2.41 Mya and was introgressed in *Hn* between 2.24 and 2.30 Mya, where it remained
152 polymorphic, forming the P supergene.

153 **Discussion**

154 Sustained differentiation between *P* alleles over the entire length of the inversion in *H. numata*
155 is therefore explained by the 1.3 My of independent evolution of an inverted haplotype within
156 *H. pardalinus*. This differentiation was maintained and accentuated after introgression by the
157 suppression of recombination. Our results show that, as previously hypothesised [5,13,14],
158 complex balanced polymorphism such as those controlled by supergene may evolve via the
159 differentiation of rearranged haplotypes in separate lineages, followed by adaptive
160 introgression in a host population where differentiated haplotypes are preserved through
161 suppression of recombination, and maintained by balancing selection. This provides the first
162 empirical evidence for a mechanism to explain the formation of supergene, and offers a
163 parsimonious solution to the paradox of the evolution of divergent haplotypes in face of
164 recombination. This mechanism may be widespread and may explain how other supergenes

165 have evolved, from the social organisation supergene in ant [7] to the coloration and behaviour
166 supergene of the white-throated sparrow [5].

167 Supergene formation through adaptive introgression requires an initial selective advantage to
168 the inversion in the recipient population, and balancing selection maintaining the
169 polymorphism. In *H. numata*, the introgressed arrangement is associated with a successful
170 melanic phenotype (*bicoloratus*) mimicking abundant local species in the foothills of the Andes
171 and enjoying a 7-fold increased protection relative to ancestral arrangements [29]. This
172 introgression likely constitutes an ecological and altitudinal expansion to premontane Andean
173 foothills where the melanic wing mimicry ring dominates, and an empirical example for the
174 theoretical role of inversions as “adaptive cassettes” triggering eco-geographical expansions in
175 an introgressed lineage [30]. Despite their role in reproductive isolation [31], inversions may be
176 prone to adaptive introgression through combined selection on linked mutations [32]. This is
177 supported by the rapid introgression of inversion P_1 after it was formed.

178 Inversion P_1 linked with the adjacent rearrangement P_2 , is also associated with other well-
179 protected mimetic forms [9,29], and most *H. numata* phenotypes associated with the inversion
180 are unmatched in *H. pardalinus*, indicating that introgression was followed by further adaptive
181 diversification to local mimicry niches. Balancing selection, mediated by negative assortative
182 mating among inversion genotypes, prevents the fixation of the inversion, as reflected by a
183 deficit of homozygotes for the introgressed haplotype in the wild [33]. Supergene evolution is
184 therefore consistent with the introgressed inversion having a strong advantage under mimicry
185 selection but being maintained in a polymorphism with ancestral haplotypes by negative
186 frequency-dependence.

187 Beyond suggesting a mechanism for supergene evolution, these findings demonstrate how
188 introgression, when involving structural variants, can trigger the emergence of novel genetic
189 architectures. This scenario may underlie the evolution of many complex polymorphisms under

190 balancing selection in a wide variety of organisms, such as MHC loci in vertebrates [34], self-
191 incompatibility loci in plants [35], mating types in fungi [36] or, much more generally, sex
192 chromosomes. Our results therefore shed new light on the importance of introgression as a
193 mechanism shaping the architecture of genomes and assisting the evolution of complex
194 adaptive strategies.

195 **Acknowledgements**

196 The authors thank Mathieu Chouteau, Violaine Llaurens, Marianne Elias, Stéphanie Gallusser,
197 César Ramírez, Benigno Calderón, Moisés Abanto, Lisa de Silva, Armando Silva, for help
198 during fieldwork, Gerardo Lamas for help with research permits in Peru, Florence Piron-
199 Prunier, Agnès Bulski, Guillaume Achaz, and Mark Blaxter, for help with lab and analytical
200 work. Analyses were conducted with the support of bioinformatic platforms Genotoul
201 (Toulouse) and MBB (Montpellier). This research was conducted under SERFOR research
202 permits from the Peruvian Ministry of Agriculture, and was supported by ANR Grant HYBEVOL
203 (ANR-12-JSV7-0005) and European Research Council Grant MimEvol (StG-243179) to MJ.

204 **Author contributions**

205 P.J., A.W. and M.J., designed the study and wrote the paper. A.W., J.M., K.K.D. and M.J.
206 generated the genomic data. P.J., A.W. and M.A.R.C. performed the genomic analyses. A.W.,
207 L.F., R.W.N. and M.J. performed marker analyses. All authors contributed to editing the
208 manuscript.

209 **Declaration of interests**

210 The authors declare no competing interests.

211

1. Li, J., Cocker, J.M., Wright, J., Webster, M.A., McMullan, M., Dyer, S., Swarbreck, D., Caccamo, M., Oosterhout, C. van, and Gilmartin, P.M. (2016). Genetic architecture and evolution of the S locus supergene in *Primula vulgaris*. *Nat. Plants* 2, 16188.
2. Timmermans, M.J.T.N., Baxter, S.W., Clark, R., Heckel, D.G., Vogel, H., Collins, S., Papanicolaou, A., Fukova, I., Joron, M., Thompson, M.J., *et al.* (2014). Comparative genomics of the mimicry switch in *Papilio dardanus*. *Proc R Soc B* 281, 20140465.
3. Kunte, K., Zhang, W., Tenger-Trolander, A., Palmer, D.H., Martin, A., Reed, R.D., Mullen, S.P., and Kronforst, M.R. (2014). doublesex is a mimicry supergene. *Nature* 507, 229–232.
4. Joron, M., Papa, R., Beltrán, M., Chamberlain, N., Mavárez, J., Baxter, S., Abanto, M., Bermingham, E., Humphray, S.J., Rogers, J., *et al.* (2006). A Conserved Supergene Locus Controls Colour Pattern Diversity in *Heliconius* Butterflies. *PLOS Biol.* 4, e303.
5. Tuttle, E.M., Bergland, A.O., Korody, M.L., Brewer, M.S., Newhouse, D.J., Minx, P., Stager, M., Betuel, A., Cheviron, Z.A., Warren, W.C., *et al.* (2016). Divergence and Functional Degradation of a Sex Chromosome-like Supergene. *Curr. Biol.*
6. Küpper, C., Stocks, M., Risse, J.E., dos Remedios, N., Farrell, L.L., McRae, S.B., Morgan, T.C., Karlionova, N., Pinchuk, P., Verkuil, Y.I., *et al.* (2016). A supergene determines highly divergent male reproductive morphs in the ruff. *Nat. Genet.* 48, 79–83.
7. Wang, J., Wurm, Y., Nipitwattanaphon, M., Riba-Grognuz, O., Huang, Y.-C., Shoemaker, D., and Keller, L. (2013). A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* 493, 664–668.
8. Lamichhaney, S., Fan, G., Widemo, F., Gunnarsson, U., Thalmann, D.S., Hoepfner, M.P., Kerje, S., Gustafson, U., Shi, C., Zhang, H., *et al.* (2016). Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nat. Genet.* 48, 84–88.
9. Joron, M., Frezal, L., Jones, R.T., Chamberlain, N.L., Lee, S.F., Haag, C.R., Whibley, A., Becuwe, M., Baxter, S.W., Ferguson, L., *et al.* (2011). Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* 477, 203–206.
10. Charlesworth, D., and Charlesworth, B. (1975). Theoretical genetics of Batesian mimicry II. Evolution of supergenes. *J. Theor. Biol.* 55, 305–324.
11. Fisher, R.A. (1930). *The Genetical Theory Of Natural Selection* (At The Clarendon Press) Available at: <http://archive.org/details/geneticaltheoryo031631mbp> [Accessed December 4, 2017].
12. Franks, D.W., and Sherratt, T.N. (2007). The evolution of multicomponent mimicry. *J. Theor. Biol.* 244, 631–639.
13. Laurens, V., Whibley, A., and Joron, M. (2017). Genetic architecture and balancing selection: the life and death of differentiated variants. *Mol. Ecol.* 26, 2430–2448.
14. Schwander, T., Libbrecht, R., and Keller, L. (2014). Supergenes and Complex

Phenotypes. *Curr. Biol.* 24, R288–R294.

15. Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet.* 2, e64.
16. Yeaman, S. (2013). Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proc. Natl. Acad. Sci.* 110, E1743–E1751.
17. Huber, B., Whibley, A., Poul, Y.L., Navarro, N., Martin, A., Baxter, S., Shah, A., Gilles, B., Wirth, T., McMillan, W.O., *et al.* (2015). Conservatism and novelty in the genetic architecture of adaptation in *Heliconius* butterflies. *Heredity* 114, 515–524.
18. Nadeau, N.J., Pardo-Diaz, C., Whibley, A., Supple, M.A., Saenko, S.V., Wallbank, R.W.R., Wu, G.C., Maroja, L., Ferguson, L., Hanly, J.J., *et al.* (2016). The gene cortex controls mimicry and crypsis in butterflies and moths. *Nature* 534, 106–110.
19. van't Hof, A.E., Campagne, P., Rigden, D.J., Yung, C.J., Lingley, J., Quail, M.A., Hall, N., Darby, A.C., and Saccheri, I.J. (2016). The industrial melanism mutation in British peppered moths is a transposable element. *Nature* 534, 102–105.
20. Van Belleghem, S.M., Rastas, P., Papanicolaou, A., Martin, S.H., Arias, C.F., Supple, M.A., Hanly, J.J., Mallet, J., Lewis, J.J., and Hines, H.M. (2017). Complex modular architecture around a simple toolkit of wing pattern genes. *Nat. Ecol. Evol.* 1, 0052.
21. Le Poul, Y., Whibley, A., Chouteau, M., Prunier, F., Llaurens, V., and Joron, M. (2014). Evolution of dominance mechanisms at a butterfly mimicry supergene. *Nat. Commun.* 5, 5644.
22. Kozak, K.M., Wahlberg, N., Neild, A., Dasmahapatra, K.K., Mallet, J., and Jiggins, C.D. (2015). Multilocus Species Trees Show the Recent Adaptive Radiation of the Mimetic *Heliconius* Butterflies. *Syst. Biol.*
23. The *Heliconius* Genome Consortium (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* 487, 94–98.
24. Enciso-Romero, J., Pardo-Díaz, C., Martin, S.H., Arias, C.F., Linares, M., McMillan, W.O., Jiggins, C.D., and Salazar, C. (2017). Evolution of novel mimicry rings facilitated by adaptive introgression in tropical butterflies. *Mol. Ecol.* 26, 5160–5172.
25. Wallbank, R.W.R., Baxter, S.W., Pardo-Diaz, C., Hanly, J.J., Martin, S.H., Mallet, J., Dasmahapatra, K.K., Salazar, C., Joron, M., Nadeau, N., *et al.* (2016). Evolutionary Novelty in a Butterfly Wing Pattern through Enhancer Shuffling. *PLOS Biol.* 14, e1002353.
26. Zhang, W., Dasmahapatra, K.K., Mallet, J., Moreira, G.R.P., and Kronforst, M.R. (2016). Genome-wide introgression among distantly related *Heliconius* butterfly species. *Genome Biol.* 17, 25.
27. Martin, S.H., Davey, J.W., and Jiggins, C.D. (2015). Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci. *Mol. Biol. Evol.* 32, 244–257.
28. Martin, S.H., and Van Belleghem, S.M. (2017). Exploring evolutionary relationships across the genome using topology weighting. *Genetics* 206, 429–438.

29. Chouteau, M., Arias, M., and Joron, M. (2016). Warning signals are under positive frequency-dependent selection in nature. *Proc. Natl. Acad. Sci. U. S. A.* *113*, 2164–2169.
30. Kirkpatrick, M., and Barrett, B. (2015). Chromosome inversions, adaptive cassettes and the evolution of species' ranges. *Mol. Ecol.* *24*, 2046–2055.
31. Hoffmann, A.A., and Rieseberg, L.H. (2008). Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation? *Annu. Rev. Ecol. Evol. Syst.* *39*, 21–42.
32. Kirkpatrick, M., and Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics* *173*, 419–434.
33. Chouteau, M., Llaurens, V., Piron-Prunier, F., and Joron, M. (2017). Polymorphism at a mimicry supergene maintained by opposing frequency-dependent selection pressures. *Proc. Natl. Acad. Sci.* *114*, 8325–8329.
34. Grossen, C., Keller, L., Biebach, I., International Goat Genome Consortium, and Croll, D. (2014). Introgression from domestic goat generated variation at the major histocompatibility complex of Alpine ibex. *PLoS Genet.* *10*, e1004438.
35. Castric, V., Bechsgaard, J., Schierup, M.H., and Vekemans, X. (2008). Repeated adaptive introgression at a gene under multiallelic balancing selection. *PLoS Genet.* *4*, e1000168.
36. Corcoran, P., Anderson, J.L., Jacobson, D.J., Sun, Y., Ni, P., Lascoux, M., and Johannesson, H. (2016). Introgression maintains the genetic integrity of the mating-type determining chromosome of the fungus *Neurospora tetrasperma*. *Genome Res.* *26*, 486–498.
37. Lunter, G., and Goodson, M. (2011). Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Res.* *21*, 936–939.
38. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and others (2009). The sequence alignment/map format and SAMtools. *Bioinformatics* *25*, 2078–2079.
39. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., Del Angel, G., Rivas, M.A., Hanna, M., *et al.* (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* *43*, 491–498.
40. Cingolani, P., Patel, V.M., Coon, M., Nguyen, T., Land, S.J., Ruden, D.M., and Lu, X. (2012). Using *Drosophila melanogaster* as a Model for Genotoxic Chemical Mutational Studies with a New Program, SnpSift. *Front. Genet.* *3*, 35.
41. Abyzov, A., Urban, A.E., Snyder, M., and Gerstein, M. (2011). CNVnator: An approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* *21*, 974–984.
42. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990). Basic local alignment search tool. *J. Mol. Biol.* *215*, 403–410.

43. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* *32*, 1792–1797.
44. Lee, T.-H., Guo, H., Wang, X., Kim, C., and Paterson, A.H. (2014). SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics* *15*, 162.
45. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* *30*, 1312–1313.
46. Browning, S.R., and Browning, B.L. (2007). Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering. *Am. J. Hum. Genet.* *81*, 1084–1097.
47. Lartillot, N., Lepage, T., and Blanquart, S. (2009). PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* *25*, 2286–2288.
48. Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol. Biol. Evol.* *33*, 1635–1638.

213 **Figure 1 | Distribution of supergene inversions in the silvaniform clade of *Heliconius*.**

214 **A** Structure of the *H. numata* (*Hn*) mimicry supergene *P* characterised by polymorphic
215 inversions and some of the morphs associated with each arrangement. *P* allows *Hn* to produce
216 highly distinguishable morphs in the same location. The first derived inversion (P_1 , blue), is
217 common to all rearranged alleles (*Hn1*), and distinguishes them from the ancestral, recessive
218 *P* alleles (mimetic forms *silvana* or *laura*, *Hn0*). The *P* dominant allele (Andean mimetic form
219 *bicoloratus* and *peeblesi*) is controlled by a rearrangement including only the chromosomal
220 inversion P_1 . A further rearrangement (P_2 , green) linked to the first inversion is associated with
221 a large diversity of derived, intermediate dominant mimicry alleles [9,21]. A 4kb duplication was
222 also detected only in individuals showing the inversion P_1 . **B** Presence/absence of the two
223 major rearrangements in species closely-related to *H. numata* (silvaniform clade), tested by
224 PCR of breakpoint-diagnostic markers, and independently by duplication-diagnostic CNV
225 assays. All species are fixed for the ancestral arrangement (red), except *H. pardalinus* (*Hp*),
226 fixed for P_1 , and *H. numata* showing polymorphism for P_1 and P_2 . Silvaniform members are
227 represented with a solid line on the species tree, while outgroup species are represented with
228 a dashed line. See also Table S1, S2 and S3.

229

230 **Figure 2 | Whole genome and inversion phylogenies of *H. numata* and related species**

231 **A** Whole genome phylogeny, showing two well-separated branches grouping *H. pardalinus*
232 and *H. elevatus* on the one hand and *H. numata* and *H. ismenius* on the other hand,
233 consistently with previous studies (*i.e.* Ref. [22]. See Figure S1 for the phylogeny with all taxa).
234 **B** Undated inversion P_1 phylogeny. All *Hn* individuals displaying the inversion P_1 (*Hn1*) group
235 with *Hp*, while *Hn* individuals displaying the ancestral arrangement (*Hn0*) remain with sister
236 species *Hi*. *He* groups closer to the outgroup (*Hc*) reflecting introgression with *H. melpomene*,

237 a species closely related to *Hc* (Figure S1; Ref. [23]). For clarity, only species informative to
238 introgression history are represented here. The inversion is a 400kb segment displaying much
239 phylogenetic heterogeneity among the other taxa, reflecting a complex history of gene flow and
240 incomplete lineage sorting (see Figure S1 for phylogenies including all taxa).

241

242 **Figure 3 | Excess of shared derived mutations between *Hp* and *Hn1*.**

243 *fd* statistic computed in non-overlapping 20 kb sliding windows and plotted along the whole
244 genome. The ABBA-BABA framework and related statistics assess here the excess of shared
245 derived mutations between *Hp* and *Hn1*, relative to a control (*Hi*) not connected by gene flow to
246 the others. Outgroup *Hc* allows the mutations to be polarized as “ancestral”(A) or “derived”(B).
247 A mean *fd* = 0 is expected if *Hp* is not connected by gene flow to *Hn1*. Unmapped contigs are
248 grouped within an “A” chromosome. The supergene scaffold (HE667780) is indicated by a blue
249 arrow. Standard error was assessed with block jackknifing (600 kb block size). See also Figure
250 S2.

251

252 **Figure 4 | Phylogenetic and divergence variation at the supergene scaffold**

253 **A** Weightings (Twisst [28]) for all fifteen possible phylogenetical topologies between *H. numata*
254 with inversion (*Hn1*), *H. numata* without inversion (*Hn0*), *H. ismenius* (*Hi*), *H. pardalinus* (*Hp*)
255 and *H. elevatus* (*He*), with loess smoothing (level = 0.05). Topology 1 is the species topology.
256 Strong topology change occurs around inversion breakpoints. Within the inversion, the best
257 supported topologies (2, 3 & 4) group *Hn1* close to *Hp*. See Figure S4 for Twisst analyses with
258 other taxa. **B** F_{ST} scan between *Hn1* and *Hn0*. Inversion P_1 shows a generally high F_{ST} value
259 contrary to the rest of the genome. P_2 rearrangement (1028-1330 kb) shows lower but
260 nonetheless elevated F_{ST} values. **C** *fd* statistic (ABBA-BABA) computed in 10 kb sliding
261 windows (increment = 500bp) with $P_1 = Hn0$, $P_2 = Hn1$, $P_3 = Hp$, $O = Hc$. Outside the inversion, a

262 *fd* value close to 0 is observed, as expected under a no gene flow scenario. At P₁ inversion
263 breakpoints, the *fd* values strongly increase and remains high in the whole inversion. **D**
264 Variation in divergence time between *Hn1* and *Hp*, computed in 10 kb non-overlapping sliding
265 windows. The divergence time inside the inversion is significantly lower than in the rest of the
266 genome. See also Figure S3.

267 **Contact for reagent and resource sharing**

268 Further information and requests of resources should be directed to and will be fulfilled by the
269 Lead Contact, Mathieu Joron (mathieu.joron@cefe.cnrs.fr)

270 **STAR Methods**

271 **Experimental model and subject details**

272 92 specimens (male or female without distinction) of *H. numata*, *H. ismenius*, *H. elevatus*, *H.*
273 *pardalinus*, *H. hecale*, *H. ethilla*, *H. besckei*, *H. melpomene* and *H. cydno* were collected in the
274 wild in Peru, Ecuador, Colombia, French Guiana, Panama and Mexico (Table S1)

275 **Methods details**

276 ***Dna extraction and sequencing.***

277 Butterfly' bodies were conserved in NaCl saturated DMSO solution at -20°C and DNA was
278 extracted using Qiagen DNeasy blood and tissue kits according to the manufacturers'
279 instructions and with RNase treatment. Illumina Truseq paired-end whole genome libraries
280 were prepared and 2x100bp reads were sequenced on the Illumina HiSeq 2000 platform.
281 Reads were mapped to the *H. melpomene* Hmel1 reference genome [23] using Stampy
282 v1.0.23 [37] with default settings except for setting the substitution rate to 0.05 to allow for
283 expected divergence from the reference. Alignment file manipulations used SAMtools v0.1.19
284 [38]. After mapping, duplicate reads were excluded using the *MarkDuplicates* tool in Picard
285 (v1.107; <http://broadinstitute.github.io/picard>) and local indel realignment using IndelRealigner
286 was performed with GATK v2.1.5 [39]. Invariant and polymorphic sites were called with GATK
287 UnifiedGenotyper. Filtering was performed on individual samples using GATK VariantFiltration
288 to remove sites with depth <10 or greater than 4 times the median coverage of the sample, or
289 sites with low mapping quality (using the expression "MQ < 40.0 || MQ0>= 4 && ((MQ0
290 /(1.0*DP))>0.1)". SnpSift filter [40] was used to exclude sites with QUAL or GQ less than or
291 equal to 30. After filtering, variant call files were merged using GATK CombineVariants.

292

293 **PCR analysis and genotyping**

294 Inversion breakpoints were genotyped by PCR amplification of genomic DNA using Thermo
295 Scientific® Phusion High-Fidelity DNA Polymerase. Primer sequences and PCR conditions
296 used are: for P₁, CCATTMTGCCAATTTMGTCT (forward) and TCMGGACTATCTTTGTATGC
297 (reverse), elongation time 2'30"; for P₂, CCATTMTGCCAATTTMGTCT (forward) and
298 GGTTACGGATGTCTTTAATG (reverse), elongation time 2'30"; for P₀,
299 AGTTTTTAAGCTGTTTCTCC (forward) and GTTAGTGCCCTGCCAAACAC (reverse),
300 elongation time 3'30"

301 **Duplication Analysis**

302 Copy number analysis of the supergene scaffold was performed on resequence alignments
303 after duplicate removal and local realignment using CNVnator v0.3 [41] with default settings
304 and a bin size of 100bp.

305 The 4kb sequence detected as duplicated was blasted [42] against the Hn1 BAC clone library
306 from Ref. [12] and against a *H. numata* genome, generated by the Heliconius consortium using
307 a combination of SMRT long read (Pacific Biosciences) and Illumina short read (Discover
308 assembly), and available on LebBase (<http://ensembl.lepbase.org>). Three BAC clones (38g4,
309 24i10 and 30F8) and two scaffolds (scaffold13474 and scaffold16807) showed high blast
310 values (e-value=0). Their entire sequences were mapped on the *H. melpomene* reference
311 genome with BLAST [42]. They correspond to two regions close to the two breakpoints of
312 inversion P₁. The sequences resulting from the duplications were extracted from the BAC
313 clones and the scaffolds and aligned with MUSCLE [43].

314

315 **ABBA-BABA analysis**

316 ABBA-BABA analyses were conducted with the scripts provided by Ref. [27]. The *fd* statistic
317 was computed in 20 kb non-overlapping windows for the whole genome (min. genotyped
318 position=1000) and 10 kb sliding windows with a 500bp step, (min. genotyped position=500)
319 for the supergene scaffold (HE667780).

320

321 ***Phylogenetic analyses***

322 To determine the direction of introgression, we used the fact that the introgressed species
323 should appear phylogenetically closer than expected to the donor species, but also closer to
324 the sister species of the donor. Thus, considering a species topology like (A,B),(C,(D,E)), a
325 sequence showing a (A,C)(D,(E,B)) topology probably arose by the way of an introgression
326 from E to B, whereas a sequence showing a ((B,E),A)(D,C) topology probably arose via
327 introgression from B to E. To search for such patterns, we computed a whole genome
328 phylogeny and several phylogenies at different locations within and outside the inversion.

329 The whole genome phylogeny was obtained with SNPhylo [44], with 100 bootstraps and *H.*
330 *cydno* as the outgroup. RaxML [45] was used to determine local phylogenies, with GTRCAT
331 model and 100 bootstrap. Nevertheless, we found that individuals from the different species
332 were frequently mixed and the species topology was highly variable, complicating the
333 interpretation of topology changes at the inversion location. We thus used Twisst [28] to
334 unravel the changes in topology and assess phylogenetic discordance along the supergene
335 scaffold . We used Beagle [46] to phase the haplotypes of the supergene scaffold, with 10000
336 bp size and 1000 bp overlapping sliding windows. Maximum likelihood trees were generated
337 with the `phymI_sliding_window.py` script with the GTR model and a 50 SNP sliding window
338 (<https://github.com/simonhmartin/twisst>).

339

340 ***Divergence time analyses***

341 To discriminate between introgression and ancestral polymorphism hypotheses, Bayesian
342 inferences of the divergence time between *H. pardalinus* and *H. numata* were made with
343 Phylobayes [47]. Analyses were performed on 10 kb non-overlapping sliding windows, using all
344 individuals of the two species and including individuals of all other species in our dataset to
345 obtain better resolution. Date estimates were calculated relative to the divergence of *H. cydno*
346 with the silvaniform clade, estimated by Ref. [22] to be approximately 3,84 Mya., using a log-
347 normal autocorrelated relaxed clock. Each chain ran for at least 30000 states, with 10000 burn-
348 in states. Chain convergence was checked with Tracer (<http://beast.bio.ed.ac.uk/Tracer>).
349 Resultant trees and time estimates were analysed with ete3 python library [48].
350 Divergence of the duplication-associated sequences was done in the same way. Whole
351 genome resequence data from all species except *Hn1* and *Hp* were used, as well as
352 sequences from the three BAC clones and the *H. numata* genome. *Hn1* and *Hp* specimens
353 were not used, as they tend to artificially increase the mutation rate inferred by Phylobayes
354

355 **Quantification and statistical analysis**

356 Standard error of *fd* mean at whole genome level was assessed with 1000 blocks Jackknife,
357 using 600 kb block. In a similar way, 1000 bootstrap were used to assess the 95 % confidence
358 interval of *fd* mean on the inversion P₁. 95 % confidence interval of divergence times were
359 directly obtained from the posterior distribution inferred by Phylobayes [47].
360

361 **DATA and software availability**

362 The datasets generated or analyzed during this study are available from NCBI SRA ().
363 Accession numbers are indicated in Table S1.
364

365