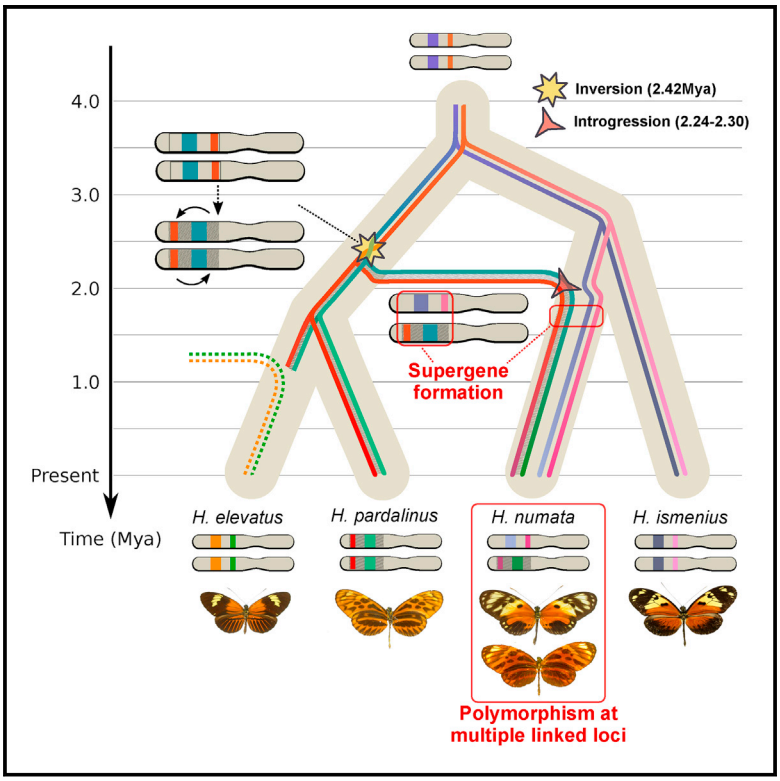
Report

Current Biology

Supergene Evolution Triggered by the Introgression of a Chromosomal Inversion

Graphical Abstract



Authors

Paul Jay, Annabel Whibley, Lise Frézal, ..., James Mallet, Kanchon K. Dasmahapatra, Mathieu Joron

Correspondence

paul.jay@cefe.cnrs.fr (P.J.), mathieu.joron@cefe.cnrs.fr (M.J.)

In Brief

Supergenes are genetic architectures underlying complex polymorphisms in many organisms. Jay et al. show that a supergene controlling mimicry polymorphism in a butterfly was formed by the introgression of a chromosomal inversion from another species. Their results emphasize the role of hybridization in the evolution of novel genetic architectures.

Highlights

- Chromosomal inversions underlie wing-pattern polymorphism in a *Heliconius* butterfly
- Inversion was introgressed from another species, then maintained as a polymorphism
- 1.3 Ma of evolution in separate taxa explains the divergence of supergene haplotypes
- Introgression may play a key role in the formation of novel genetic architectures

Jay et al., 2018, Current Biology *28*, 1–7
June 4, 2018
https://doi.org/10.1016/j.cub.2018.04.072

CellPress

Please cite this article in press as: Jay et al., Supergene Evolution Triggered by the Introgression of a Chromosomal Inversion, Current Biology (2018), https://doi.org/10.1016/j.cub.2018.04.072

Current Biology
Report

Supergene Evolution Triggered by the Introgression of a Chromosomal Inversion

Paul Jay,[1,*] Annabel Whibley,[2] Lise Frézal,[3] María Ángeles Rodríguez de Cara,[1] Reuben W. Nowell,[4] James Mallet,[5] Kanchon K. Dasmahapatra,[6] and Mathieu Joron[1,7,*]
[1]CEFE, CNRS, Université de Montpellier, Université Paul Valéry Montpellier 3, EPHE, IRD, Montpellier, France
[2]Department of Cell and Developmental Biology, John Innes Centre, Norwich NR4 7UH, UK
[3]Institut de Biologie de l'Ecole Normale Supérieure, CNRS, INSERM, ENS, Paris Sciences et Lettres, Paris, France
[4]Department of Life Sciences, Imperial College London, Silwood Park Campus, Buckhurst Road, Ascot, Berkshire SL5 7PY, UK
[5]Department of Organismic and Evolutionary Biology, Biology Laboratories, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA
[6]Department of Biology, University of York Wentworth Way, Heslington YO10 5DD, UK
[7]Lead Contact
*Correspondence: paul.jay@cefe.cnrs.fr (P.J.), mathieu.joron@cefe.cnrs.fr (M.J.)
https://doi.org/10.1016/j.cub.2018.04.072

SUMMARY

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

RESULTS

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

To understand allelic evolution in supergenes, we studied Amazonian populations of the butterfly *Heliconius numata*, in which up to seven distinct wing-pattern morphs coexist (Figure 1A), each one matching to near perfection the colors and shapes of other toxic Lepidoptera (Heliconiinae, Danainae, Pericopiinae) [12]. This balanced polymorphism is controlled by a supergene locus (*P*) associated with an inversion polymorphism [12] that captures multiple genetic loci controlling wing-pattern variation in butterflies and moths [4, 18–21] and allows multiple wing elements to be inherited as a single Mendelian character. The ancestral chromosomal arrangement, called *Hn0*, is associated with the recessive supergene allele [17], which controls the widely distributed morph *silvana.* All other characterized supergene alleles, grouped into a family of alleles called *Hn1*, determine a diversity of mimetic morphs dominant to *silvana* and associated with the 400-kb inversion P_1 (Figure 1A; [9, 17]). A subset of these alleles is associated with additional

Please cite this article in press as: Jay et al., Supergene Evolution Triggered by the Introgression of a Chromosomal Inversion, Current Biology (2018), https://doi.org/10.1016/j.cub.2018.04.072

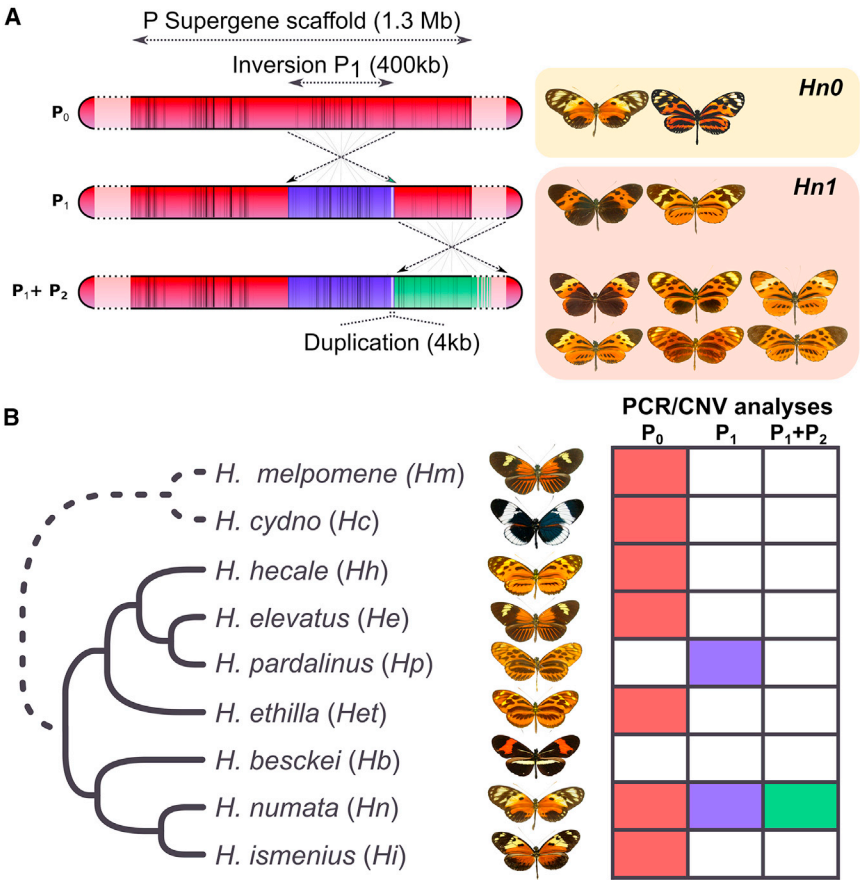


Figure 1. Distribution of Supergene Inversions in the Silvaniform Clade of *Heliconius*

(A) Structure of the *H. numata* (*Hn*) mimicry supergene *P* characterized by polymorphic inversions and some of the morphs associated with each arrangement. *P* allows *Hn* to produce highly distinguishable morphs in the same location. The first derived inversion (P_1, blue), is common to all rearranged alleles (*Hn1*) and distinguishes them from the ancestral, recessive *P* alleles (mimetic forms *silvana* or *laura*, *Hn0*). The *P* dominant allele (Andean mimetic form *bicoloratus* and *peeblesi*) is controlled by a rearrangement including only the chromosomal inversion P_1. A further rearrangement (P_2, green) linked to the first inversion is associated with a large diversity of derived, intermediate dominant mimicry alleles [9, 17]. A 4-kb duplication was also detected only in individuals showing the inversion P_1.

(B) Presence/absence of the two major rearrangements in species closely related to *H. numata* (silvaniform clade), tested by PCR of breakpoint-diagnostic markers and independently by duplication-diagnostic CNV assays. All species are fixed for the ancestral arrangement (red), except *H. pardalinus* (*Hp*), fixed for P_1, and *H. numata* showing polymorphism for P_1 and P_2. Silvaniform members are represented with a solid line on the species tree, while outgroup species are represented with a dashed line.

See also Table S1, S2, and S3.

rearrangements (P_2) in adjacent positions [12]. The emergence of the *P* supergene architecture is therefore associated with the introduction of inversion P_1, maintained at intermediate frequency by balancing selection and followed by adjacent rearrangements. To explore the origin and evolution of the *P* supergene, we thus tracked the history of inversion P_1. This inversion forms a well-differentiated haplotype that is distinct from the ancestral haplotype along its entire length (Figure 4B) and with extreme values of linkage disequilibrium (LD) [12]. Inversion P_1 therefore stands as a block of up to 7,000 differentiated SNPs along its 400-kb length, associated with supergene evolution, adaptive diversification, and dominance variation.

Heliconius numata belongs to the so-called silvaniform clade of ten species, which diverged ca. 4 million years ago (mya) from its sister clade (Figure 1B; Figure 2A; Figure S1; [23]). The Heliconius, and silvaniform members particularly, are knows to be highly connected by gene flow and to notably exchange wing pattern loci [19, 24–27]. To investigate the history of inversion P_1, we surveyed the presence of this inversion in other species of the clade. PCR amplification of inversion breakpoints showed that inversion P_1 was polymorphic in *H. numata* (*Hn*) across its Amazonian range and was also found fixed in all populations of *H. pardalinus* (*Hp*), a non-sister species deeply divergent from *H. numata* within the silvaniform clade (Figure 1B and Table S3). All other taxa, including the sister species of *H. numata* and that of *H. pardalinus*, were positive only for markers diagnostic of the ancestral gene order. Furthermore, a 4-kb duplication associated with P_1 in *Hn* was also found in whole-genome-sequence datasets for all *Hp* individuals and no other taxon (Figure 1B and Table S2). Breakpoint homology and similar molecular signatures in *Hp* and *Hn* are thus consistent with a single origin of this inversion. This sharing of P_1 between non-sister species could be due to the differential fixation of an ancient polymorphic inversion (incomplete lineage sorting, ILS) or to a secondary transfer through introgression.

To clarify whether this sharing between *Hp* and *Hn* is a rare anomaly specifically associated with the supergene locus or a common feature that is also found elsewhere in the genome, we estimated the excess of shared derived mutations between sympatric *Hp* and *Hn* relative to an allopatric control, *H. ismenius* (*Hi*, sister species of *Hn*), using the *fd* statistic [28]. We estimated that a significant 6.2% of the genome was shared via gene flow between *Hn* and *Hp* (mean *fd* = 0.062*, Figure 3 and Figure S2C), consistent with a general signal of genome-wide gene flow between *Hn* and other species within the silvaniform clade (Figure S2) and between other *Heliconius* species [24]. When *fd* is estimated using *Hn* specimens homozygous for inversion P_1 (*Hn1*), the supergene scaffold is associated with a strong peak of shared derived mutations between *Hn* and *Hp* (mean = 0.38, 95% interval 0.34–0.41, Figure 3, blue arrow). This is not observed between *Hn1* and other silvaniforms (Figure S2), nor when using *Hn* specimens homozygous for the ancestral supergene arrangement (*Hn0* ; Figure S2C). Between *Hn1* and *Hp*, the entire P_1 inversion shows a high level of *fd*, which drops to

Figure 2. Whole-Genome and Inversion Phylogenies of *H. numata* and Related Species

(A) Whole genome phylogeny, showing two well-separated branches grouping *H. pardalinus* and *H. elevatus* on the one hand and *H. numata* and *H. ismenius* on the other hand, consistent with previous studies (e.g., [23], see Figure S1 for the phylogeny with all taxa).

(B) Undated inversion P_1 phylogeny. All *Hn* individuals displaying the inversion P_1 (*Hn1*) group with *Hp*, while *Hn* individuals displaying the ancestral arrangement (*Hn0*) remain with sister species *Hi*. *He* groups closer to the outgroup (*Hc*) reflecting introgression with *H. melpomene*, a species closely related to *Hc* (Figure S1; [24]). For clarity, only species that are informative to introgression history are represented here. The inversion is a 400-kb segment displaying much phylogenetic heterogeneity among the other taxa, reflecting a complex history of gene flow and incomplete lineage sorting (see Figure S1 for phylogenies including all taxa). See also Table S1.

background levels precisely at inversion breakpoints (Figure 4C). *Hn1* and *Hp* therefore share a block of derived mutations associated with the inversion.

Contrary to estimates from the whole genome, a local excess of *fd*—denoting a local excess of shared derived mutations between two taxa—may be due to incomplete lineage sorting or to gene flow. To determine the cause of the local excess of *fd* at the supergene, we estimated the divergence times of *Hn1* and *Hp* within and outside of inversion P_1. The unique ancestor of inversion P_1 in *Hp* and *Hn1* was estimated to be 2.3 million years (Ma) old (95% interval 1.98–2.63 mya, Figure 4D; gray triangles in Figure 2), significantly more recent than the divergence time of the rest of the genome (3.59 mya; 95% interval 3.37–3.75 mya; Figure 4D), which indicates that the inversion was shared by gene flow among lineages well after their split. This introgression can be dated to an interval between the time to the most recent common ancestor (TMRCA) of *Hp* and *Hn1* inversions (i.e., 2.30 mya) and the TMRCA of all *Hn1* inversions (2.24 mya, 95% interval 1.89–2.59 mya, Figure S3D; black triangle in Figure 2), i.e., about 1.30 Ma after *Hp-Hn* speciation. We then estimated the age of the inversion considering that its occurrence also induces the 4-kb duplication we detected. We identified the two sequences of the duplicated region associated with the inversion in an *Hn1* BAC library and in an *Hn1* genome assembly and estimated their divergence time. We found that the duplication and most probably the inversion occurred 2.41 mya (95% interval 1.96–2.71 mya). This indicates that inversion P_1 may have spread between lineages *Hp* and *Hn* shortly after the inversion event.

To determine the direction of introgression, we surveyed the position of the sister species to *Hn* (*Hi*) and to *Hp* (*H. elevatus, He*) in phylogenies computed along the supergene scaffold and in other regions of the genome. The genome as a whole and regions flanking the inversion all show a similar topology to the one found by Kozak et al. [23], with expected sister

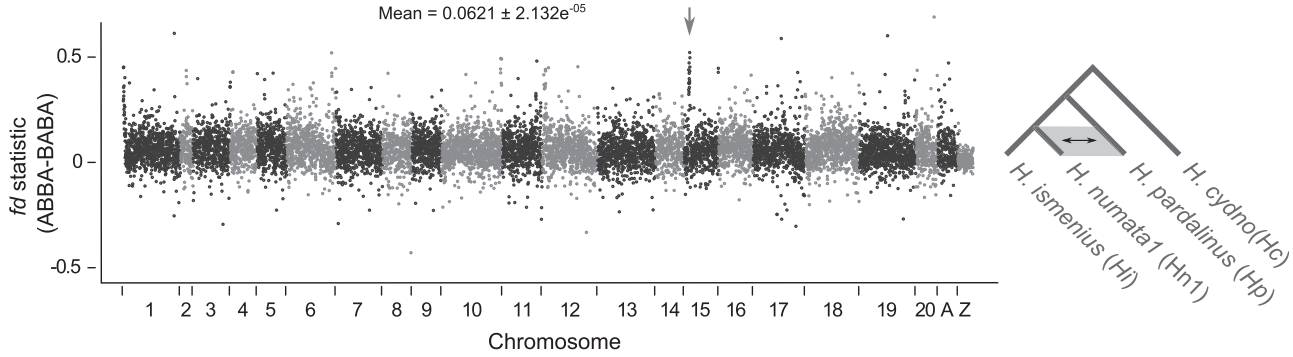


Figure 3. Excess of Shared Derived Mutations between *Hp* and *Hn1*

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

See also Figure S2 and Table S1.

Please cite this article in press as: Jay et al., Supergene Evolution Triggered by the Introgression of a Chromosomal Inversion, Current Biology (2018), https://doi.org/10.1016/j.cub.2018.04.072

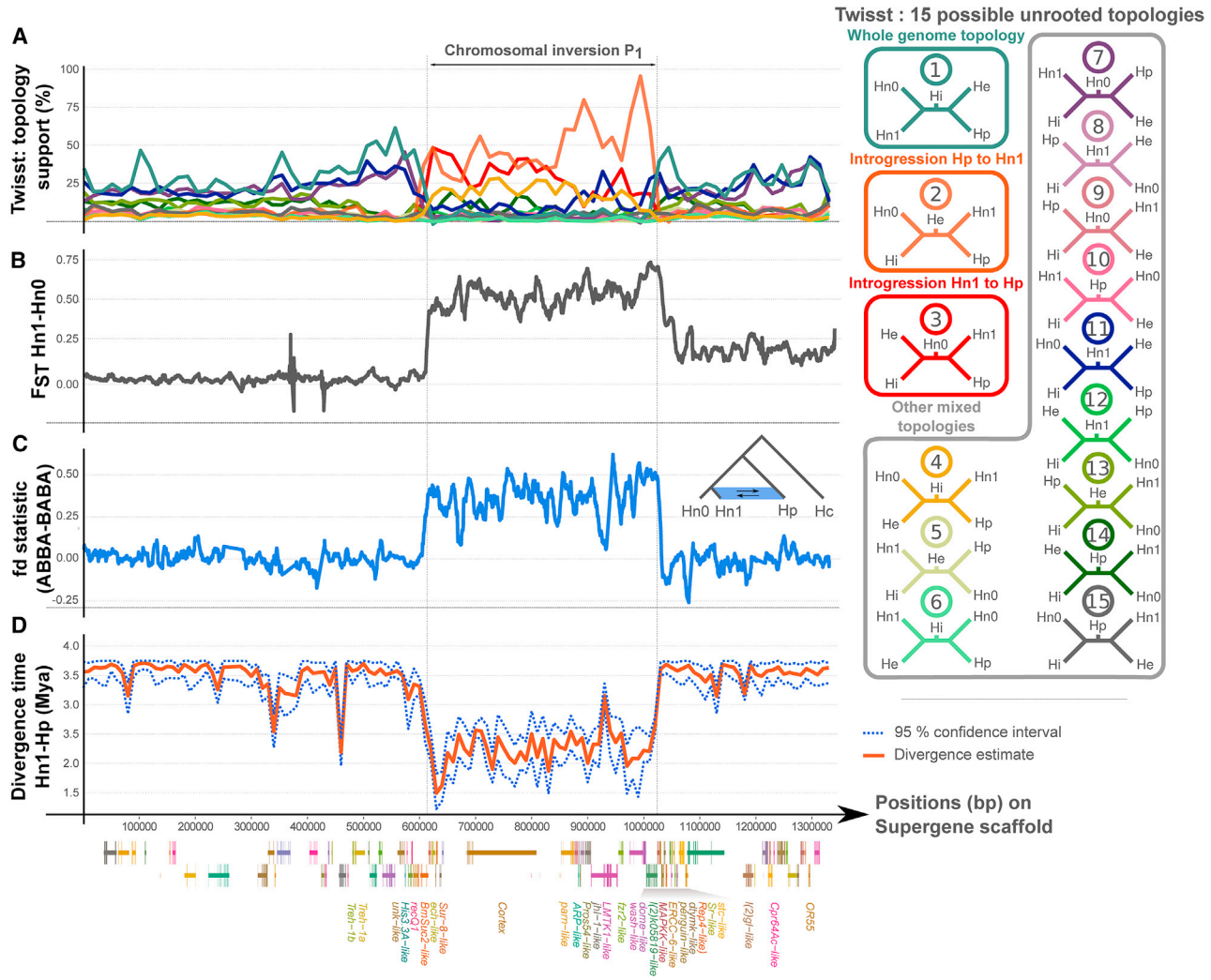


Figure 4. Phylogenetic and Divergence Variation at the Supergene Scaffold

(A) Weightings (Twisst [22]) for all fifteen possible phylogenetical topologies involving *H. numata* with inversion (*Hn1*), *H. numata* without inversion (*Hn0*), *H. ismenius* (*Hi*), *H. pardalinus* (*Hp*), and *H. elevatus* (*He*), with loess smoothing (level = 0.05). Topology 1 is the species topology. Strong topology change occurs at inversion breakpoints. Within the inversion, the best supported topologies (2, 3, and 4) group *Hn1* close to *Hp*. See Figure S4 for Twisst analyses with other taxa.

(B) F_{ST} scan between *Hn1* and *Hn0*. Inversion P_1 shows a generally high *Fst* value contrary to the rest of the genome. P_2 rearrangement (1,028–1,330 kb) shows lower but nonetheless elevated F_{ST} values.

(C) *fd* statistic (ABBA-BABA) computed in 10-kb sliding windows (increment = 500bp) with P1 = *Hn0*, *P2* = *Hn1*, *P3* = *Hp*, *O* = *Hc*. Outside the inversion, an *fd* value close to 0 is observed, as expected under a no-gene-flow scenario. At P_1 inversion breakpoints, *fd* values strongly increase and remain high across the entire inversion.

(D) Variation in divergence time between *Hn1* and *Hp*, computed in 10-kb non-overlapping sliding windows. Divergence time inside the inversion is significantly lower than in the rest of the genome.

See also Figure S3 and Table S1.

relationships of *Hi* and *Hn* and of *He* and *Hp* (Figure 2A and Figure S1). Evaluating the support for each possible topology among the five informative taxa (*Hn0, Hn1, Hi, Hp, He*) using Twisst [22] confirmed the consistent support for the separation of (*Hn*, *Hi*) and (*Hp*, *He*) clades despite a high level of incomplete lineage sorting within each clade (Figure 4A and Figure S4). By contrast, the inversion P_1 shows strong support for topologies that group *Hn1* with *Hp*, and major topology changes coincide with inversion breakpoints (Figures 2B and 4A and Figure S4), consistent with a single origin of the inversion. Within the inversion, the highest support consistently goes to *Hn1* grouping within (*Hp*, *He*) and away from (*Hn*, *Hi*) (Figure 4C, topology 2), indicating an introgression from *Hp* to *Hn*. This conclusion is robust to the species used as sister groups to *Hn* or *Hp* (Figures S4D–S4G). Alternative topologies (3 and 4) are also found in relatively high proportions in the interval ~650–850 kb, presumably owing to high levels of incomplete lineage sorting at the clade level in this region or to ancient gene flow among other species of the clade. Supporting these interpretations, topology analysis with taxa unaffected by *Hn1-Hp* introgression (for instance, using *Hn0* and replacing *Hp* with a closely related species, *H. hecale*) still showed the same pattern of unresolved

phylogenetic signal in this interval between the two major branches of the clade (*Hp-He-H. hecale* versus *Hn-Hi*) (Figures S4H and S4I). This suggests that the mixed phylogenetic signal found in this interval is independent of the introgression. Overall, our results show that the inversion P_1 most likely occurred in *Hp* 2.41 mya and was introgressed in *Hn* between 2.24 and 2.30 mya, where it remained polymorphic, forming the *P* supergene.

DISCUSSION

Sustained differentiation between *P* alleles over the entire length of the inversion in *H. numata* is therefore explained by the 1.3 Ma of independent evolution of an inverted haplotype within *H. pardalinus.* This differentiation was maintained and accentuated after introgression by the suppression of recombination. Our results show that, as previously hypothesized [5, 13, 14], complex balanced polymorphisms such as those controlled by supergenes may evolve via the differentiation of rearranged haplotypes in separate lineages, followed by adaptive introgression in a host population where differentiated haplotypes are preserved through suppression of recombination, and maintained by balancing selection. This provides the first empirical evidence for a mechanism to explain the formation of supergene and offers a parsimonious solution to the paradox of the evolution of divergent haplotypes in face of recombination. This mechanism may be widespread and may explain how other supergenes have evolved, from the social organization supergene in ants [7] to the coloration and behavior supergene of the white-throated sparrow [5].

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

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

Beyond suggesting a mechanism for supergene evolution, these findings demonstrate how introgression, when involving structural variants, can trigger the emergence of novel genetic architectures. This scenario may underlie the evolution of many complex polymorphisms under balancing selection in a wide variety of organisms, such as MHC loci in vertebrates [34], self-incompatibility loci in plants [35], mating types in fungi [36], or—much more generally—sex chromosomes. Our results therefore shed new light on the importance of introgression as a mechanism shaping the architecture of genomes and assisting the evolution of complex adaptive strategies.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Dna extraction and sequencing
  - PCR analysis and genotyping
  - Duplication Analysis
  - ABBA-BABA analysis
  - Phylogenetic analyses
  - Divergence time analyses
- QUANTIFICATION AND STATISTICAL ANALYSIS
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and three tables and can be found with this article online at https://doi.org/10.1016/j.cub.2018.04.072.

ACKNOWLEDGMENTS

The authors thank Mathieu Chouteau, Violaine Llaurens, Marianne Elias, Stéphanie Gallusser, César Ramírez, Benigno Calderón, Moisés Abanto, Lisa de Silva, and Armando Silva for help during fieldwork; Gerardo Lamas for help with research permits in Peru; and Florence Piron-Prunier, Agnès Bulski, Guillaume Achaz, and Mark Blaxter for help with lab and analytical work. Analyses were conducted with the support of bioinformatic platforms Genotoul (Toulouse) and MBB (Montpellier). This research was conducted under research permits 076-2007-INRENA-IFFS-DCB, 288-2009-AG-DGFFS-DGEFFS and 0148-2011-AG-DGFFS-DGEFFS from the Peruvian Ministry of Agriculture and was supported by ANR Grant HYBEVOL (ANR-12-JSV7-0005) and European Research Council Grant MimEvol (StG-243179) to M.J.

AUTHOR CONTRIBUTIONS

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

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 13, 2018
Revised: March 29, 2018
Accepted: April 18, 2018
Published: May 24, 2018

REFERENCES

1. Li, J., Cocker, J.M., Wright, J., Webster, M.A., McMullan, M., Dyer, S., Swarbreck, D., Caccamo, M., Oosterhout, C.V., 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. Biol. Sci. *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. *26*, 344–350.
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., Hoeppner, 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 (Clarendon Press). Available at: https://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. Llaurens, 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. USA *110*, E1743–E1751.
17. 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.
18. Huber, B., Whibley, A., Poul, Y.L., Navarro, N., Martin, A., Baxter, S., Shah, A., Gilles, B., Wirth, T., McMillan, W.O., and Joron, M. (2015). Conservatism and novelty in the genetic architecture of adaptation in Heliconius butterflies. Heredity (Edinb) *114*, 515–524.
19. 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.
20. 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.
21. 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.
22. Martin, S.H., and Van Belleghem, S.M. (2017). Exploring evolutionary relationships across the genome using topology weighting. Genetics *206*, 429–438.
23. Kozak, K.M., Wahlberg, N., Neild, A.F., 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. *64*, 505–524.
24. Heliconius Genome Consortium (2012). Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature *487*, 94–98.
25. 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.
26. 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.
27. 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.
28. 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.
29. Chouteau, M., Arias, M., and Joron, M. (2016). Warning signals are under positive frequency-dependent selection in nature. Proc. Natl. Acad. Sci. USA *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. USA *114*, 8325–8329.
34. Grossen, C., Keller, L., Biebach, I., and Croll, D.; International Goat Genome Consortium (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., and Durbin, R.; 1000 Genome Project Data Processing

Subgroup (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.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological Samples		
Butterflies of the genus *Heliconius*	This study	N/A
Critical Commercial Assays		
Dneasy blood & tissue kit	QIAGEN	Cat#69504
Deposited Data		
Raw Illumina sequences	This study	NCBI SRA # PRJEB12740, PRJEB1749, PRJEB2743, PRJEB8011, PRJNA308754, PRJNA471310
Raw Pacific Biosciences sequences	This study	N/A
H. numata genome	LepBase	http://ensembl.lepbase.org/Heliconius_numata_helico3/Info/Index
H. melpomene reference genome	[24]	http://ensembl.lepbase.org/Heliconius_melpomene/Info/Index
Oligonucleotides		
24i10_F: CCATTMTGCCAATTTMGTCT	This study	N/A
24i10_R: TCMGGACTATCTTTGTATGC	This study	N/A
38G4_F: CCATTMTGCCAATTTMGTCT	This study	N/A
38G4_R: GGTTACGGATGTCTTTAATG	This study	N/A
31B4_F: AGTTTTTAAGCTGTTTCTCC	This study	N/A
31B4_R: GTTAGTGCCCTGCCAAACAC	This study	N/A
Software and Algorithms		
Twisst	[22]	N/A
Phylobayes 4.1	[37]	N/A
RaxML v8.2	[38]	N/A
SNPhylo	[39]	N/A
GATK v2.1.5	[40]	N/A
CNVnator v0.3	[41]	N/A
Stampy v1.0.23	[42]	N/A
Picard v1.107	http://broadinstitute.github.io/picard	N/A
SAMtools v0.1.19	[43]	N/A
MUSCLE	[44]	N/A
BEAGLE	[45]	N/A
BLAST	[46]	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

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

EXPERIMENTAL MODEL AND SUBJECT DETAILS

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

METHOD DETAILS

Dna extraction and sequencing

Butterfly' bodies were conserved in NaCl saturated DMSO solution at −20°C and DNA was extracted using QIAGEN DNeasy blood and tissue kits according to the manufacturers' instructions and with RNase treatment. Illumina Truseq paired-end whole genome

libraries were prepared and 2x100bp reads were sequenced on the Illumina HiSeq 2000 platform. Reads were mapped to the *H. melpomene* Hmel1 reference genome [24] using Stampy v1.0.23 [37] with default settings except for setting the substitution rate to 0.05 to allow for expected divergence from the reference. Alignment file manipulations used SAMtools v0.1.19 [38]. After mapping, duplicate reads were excluded using the *MarkDuplicates* tool in Picard (v1.107; http://broadinstitute.github.io/picard) and local indel realignment using IndelRealigner was performed with GATK v2.1.5 [39]. Invariant and polymorphic sites were called with GATK UnifiedGenotyper. Filtering was performed on individual samples using GATK VariantFiltration to remove sites with depth < 10 or greater than 4 times the median coverage of the sample, or sites with low mapping quality (using the expression "MQ < 40.0 || MQ0>= 4 && ((MQ0 /(1.0*DP))>0.1)." SnpSift filter [40] was used to exclude sites with QUAL or GQ less than or equal to 30. After filtering, variant call files were merged using GATK CombineVariants.

PCR analysis and genotyping

Inversion breakpoints were genotyped by PCR amplification of genomic DNA using Thermo Scientific® Phusion High-Fidelity DNA Polymerase. Primer sequences and PCR conditions used are: for P_1, CCATTMTGCCAATTTMGTCT (forward) and TCMGGACT ATCTTTGTATGC (reverse), elongation time 2′30"; for P_2, CCATTMTGCCAATTTMGTCT (forward) and GGTTACGGATGTCTTTAATG (reverse), elongation time 2′30"; for P_0, AGTTTTTAAGCTGTTTCTCC (forward) and GTTAGTGCCCTGCCAAACAC (reverse), elongation time 3′30"

Duplication Analysis

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

The 4kb sequence detected as duplicated was blasted [42] against the Hn1 BAC clone library from Ref [12]. and against a *H. numata* genome, generated by the Heliconius consortium using a combination of SMRT long read (Pacific Biosciences) and Illumina short read (Discovar assembly), and available on LebBase (http://ensembl.lepbase.org/index.html). Three BAC clones (38 g4, 24i10 and 30F8) and two scaffolds (scaffold13474 and scaffold16807) showed high blast values (e-value = 0). Their entire sequences were mapped on the *H. melpomene* reference genome with BLAST [42]. They correspond to two regions close to the two breakpoints of inversion P_1. The sequences resulting from the duplications were extracted from the BAC clones and the scaffolds and aligned with MUSCLE [43].

ABBA-BABA analysis

ABBA-BABA analyses were conducted with the scripts provided by Ref [28]. The *fd* statistic was computed in 20 kb non-overlapping windows for the whole genome (min. genotyped position = 1000) and 10 kb sliding windows with a 500bp step, (min. genotyped position = 500) for the supergene scaffold (HE667780).

Phylogenetic analyses

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

The whole genome phylogeny was obtained with SNPhylo [44], with 100 bootstraps and *H. cydno* as the outgroup. RaxML [45] was used to determine local phylogenies, with GTRCAT model and 100 bootstrap. Nevertheless, we found that individuals from the different species were frequently mixed and the species topology was highly variable, complicating the interpretation of topology changes at the inversion location. We thus used Twisst [22] to unravel the changes in topology and assess phylogenetic discordance along the supergene scaffold . We used Beagle [46] to phase the haplotypes of the supergene scaffold, with 10000 bp size and 1000 bp overlapping sliding windows. Maximum likelihood trees were generated with the phyml_sliding_window.py script with the GTR model and a 50 SNP sliding window (https://github.com/simonhmartin/twisst).

Divergence time analyses

To discriminate between introgression and ancestral polymorphism hypotheses, Bayesian inferences of the divergence time between *H. pardalinus* and *H. numata* were made with Phylobayes [47]. Analyses were performed on 10 kb non-overlapping sliding windows, using all individuals of the two species and including individuals of all other species in our dataset to obtain better resolution. Date estimates were calculated relative to the divergence of *H. cydno* with the silvaniform clade, estimated by Ref [23]. to be approximately 3,84 mya., using a log-normal autocorrelated relaxed clock. Each chain ran for at least 30000 states, with 10000 burn-in states. Chain convergence was checked with Tracer (http://beast.community/tracer). Resultant trees and time estimates were analyzed with ete3 python library [48].

Divergence of the duplication-associated sequences was done in the same way. Whole genome resequence data from all species except *Hn1* and *Hp* were used, as well as sequences from the three BAC clones and the *H. numata* genome. *Hn1* and *Hp* specimens were not used, as they tend to artificially increase the mutation rate inferred by Phylobayes

QUANTIFICATION AND STATISTICAL ANALYSIS

Standard error of *fd* mean at whole genome level was assessed with 1000 blocks Jackknife, using 600 kb block. In a similar way, 1000 bootstrap were used to assess the 95% confidence interval of fd mean on the inversion P_1. 95% confidence interval of divergence times were directly obtained from the posterior distribution inferred by Phylobayes [47].

DATA AND SOFTWARE AVAILABILITY

The datasets generated or analyzed during this study are available from NCBI SRA (PRJEB12740, PRJEB1749, PRJEB2743, PRJEB8011, PRJNA308754, PRJNA471310). Individual sample accession numbers are indicated in Table S1.