BUTTERFLY GENOMICS

Genomic architecture and introgression shape a butterfly radiation

Nathaniel B. Edelman[1]*, Paul B. Frandsen[2,3], Michael Miyagi[1], Bernardo Clavijo[4], John Davey[5,20], Rebecca B. Dikow[3], Gonzalo García-Accinelli[4], Steven M. Van Belleghem[6], Nick Patterson[7,8], Daniel E. Neafsey[8,9], Richard Challis[10], Sujai Kumar[11], Gilson R. P. Moreira[12], Camilo Salazar[13], Mathieu Chouteau[14], Brian A. Counterman[15], Riccardo Papa[6,16], Mark Blaxter[10], Robert D. Reed[17], Kanchon K. Dasmahapatra[5], Marcus Kronforst[18], Mathieu Joron[19], Chris D. Jiggins[20], W. Owen McMillan[21], Federica Di Palma[4], Andrew J. Blumberg[22], John Wakeley[1], David Jaffe[8,23], James Mallet[1]*

We used 20 de novo genome assemblies to probe the speciation history and architecture of gene flow in rapidly radiating *Heliconius* butterflies. Our tests to distinguish incomplete lineage sorting from introgression indicate that gene flow has obscured several ancient phylogenetic relationships in this group over large swathes of the genome. Introgressed loci are underrepresented in low-recombination and gene-rich regions, consistent with the purging of foreign alleles more tightly linked to incompatibility loci. Here, we identify a hitherto unknown inversion that traps a color pattern switch locus. We infer that this inversion was transferred between lineages by introgression and is convergent with a similar rearrangement in another part of the genus. These multiple de novo genome sequences enable improved understanding of the importance of introgression and selective processes in adaptive radiation.

Adaptive radiations play a fundamental role in generating biodiversity. Initiated by key innovations and ecological opportunity, radiation is fueled by niche competition that promotes rapid diversification of species (*1*). Reticulate evolution may enhance radiation by introducing genetic variation, enabling rapidly emerging populations to take advantage of new ecological opportunities (*2*, *3*). Diverging from its sister genus *Eueides* ~12 million years (My) ago, *Heliconius* radiated in a burst of speciation in the last ~5 My (*4*). Introgression is well known in *Heliconius*, with widespread reticulate evolution across the genus (*5*), although this has been disputed (*6*). Nonetheless, how introgression varies across the genome is known only in one pair of sister lineages (*7*, *8*). Here, we use multiple de novo whole-genome assemblies to improve the resolution of introgression, incomplete lineage sorting (ILS), and genome architecture in deeper branches of the *Heliconius* phylogeny.

Phylogenetic analysis

We generated 20 de novo genome assemblies for species in both major *Heliconius* subclades and three additional genera of Heliconiini. We then aligned the 16 highest-quality Heliconiini assemblies to two *Heliconius* reference genomes and seven other Lepidoptera genomes, resulting in an alignment of 25 taxa (*9*). De novo assembly provides superior sequence information for low-complexity regions, allows for discovery of structural rearrangements, and improves alignment of evolutionarily distant clades (*10*). Other studies in *Heliconius* have shown a high level of phylogenetic discordance, arguably a result of rampant introgression (*4*, *5*). We attempted to reconstruct a bifurcating species tree by estimating relationships using protein-coding genes, conserved coding regions, and conserved noncoding regions. We generated phylogenies with coalescent-based and concatenation approaches using both the full Lepidoptera alignment and a restricted, Heliconiini-only subalignment. These topologies were largely congruent among analytical approaches, but weakly supported nodes were resolved inconsistently. These approaches therefore failed to resolve the phylogeny of *Heliconius* as a simple bifurcating tree (Fig. 1A and fig. S20).

To determine whether hybridization was a cause of the species tree uncertainty, we calculated Patterson's *D* statistics (*11*) for every triplet of the 13 *Heliconius* species using a member of the sister genus, *Eueides tales*, as the outgroup. In 201 of 286 triplets, we observed values significantly different from zero based on block-jackknifing, demonstrating strong evidence for introgression (fig. S53). However, these tests alone yield little quantitative information about admixture. We therefore used phyloNet (*12*) to infer reticulate phylogenetic networks of these species based on random samples of 100 10-kb windows across the alignment. For each sample, we coestimated all 100 regional gene trees and the overall species network in parallel (*12*). To improve alignments, we analyzed the *melpomene*-silvaniform group with respect to the *Heliconius melpomene* Hmel2.5 assembly (*13*) and the *erato-sara* group with respect to the *H. erato demophoon* v1 assembly (*9*, *14*). Most species exhibited an admixture event at some point in their history using this method; we confirmed extensive reticulation among silvaniform species and discovered major gene-flow events in the *erato-sara* clade. On the basis of these results, we propose the reticulate phylogenies shown in Fig. 1, B to C.

Correlation of local ancestry with genome architecture

We next analyzed the distribution of tree topologies across the genome, again treating each major clade separately and using its respective reference genome. The *melpomene*-silvaniform group lacked topological consensus, unsurprisingly because introgression, especially of key mimicry loci, is well known in this clade (*15*). The most common tree topology was found in only 4.3% of windows, with an additional 14 topologies appearing in 1.0 to 3.4% of windows (fig. S19 to S21). By contrast, we here focus on the *erato-sara* group, in which two topologies dominate (Fig. 2). One (Fig. 2B, Tree 2) matched our bifurcating consensus topology (Fig. 1A) and a recently published tree (*4*), whereas the other (Tree 1) differs in that it places *H. hecalesia* and *H. telesiphe* as sisters.

Regions with local topologies discordant from the species tree may have arisen through introgression or ILS. To make within-topology

[1]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. [2]Department of Plant and Wildlife Sciences, Brigham Young University, Provo, UT 84602, USA. [3]Data Science Lab, Office of the Chief Information Officer, Smithsonian Institution, Washington, DC 20560, USA. [4]Earlham Institute, Norwich Research Park, Norwich NR4 7UZ, UK. [5]Bioscience Technology Facility, Department of Biology, University of York, York YO10 5DD, UK. [6]Department of Biology, University of Puerto Rico, Río Piedras Campus, San Juan, PR 00931-3360, Puerto Rico. [7]Department of Human Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. [8]Broad Institute of MIT and Harvard, Cambridge, MA, 02142 USA. [9]Harvard TH Chan School of Public Health, Boston, MA 02115, USA. [10]Wellcome Sanger Institute, Wellcome Genome Campus, Cambridge CB10 1SA, UK. [11]Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, UK. [12]Departamento de Zoologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, 91501-970 Brasil. [13]Biology Program, Faculty of Natural Sciences and Mathematics, Universidad del Rosario, Carrera 24, No. 63C-69, Bogotá D.C. 111221, Colombia. [14]Laboratoire Ecologie, Evolution, Interactions des Systèmes Amazoniens (LEEISA), USR 3456, Université De Guyane, CNRS Guyane, 275 Route de Montabo, 97334 Cayenne, French Guiana. [15]Department of Biological Sciences, Mississippi State University, Starkville, MS 39762, USA. [16]Molecular Sciences and Research Center, University of Puerto Rico, San Juan, PR 00931-3360, Puerto Rico. [17]Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA. [18]Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637, USA. [19]CEFE, CNRS, Université de Montpellier, Université Paul Valéry Montpellier 3, EPHE, IRD, 34090 Montpellier, France. [20]Department of Zoology, University of Cambridge, Cambridge CB2 3EJ, UK. [21]Smithsonian Tropical Research Institute, Apartado 0843-03092 Panamá, Panama. [22]Department of Mathematics, University of Texas, Austin, TX 78712, USA. [23]10x Genomics, Pleasanton, CA 94566, USA.
*Corresponding author. Email: nedelman@g.harvard.edu (N.B.E.); jmallet@oeb.harvard.edu (J.M.)
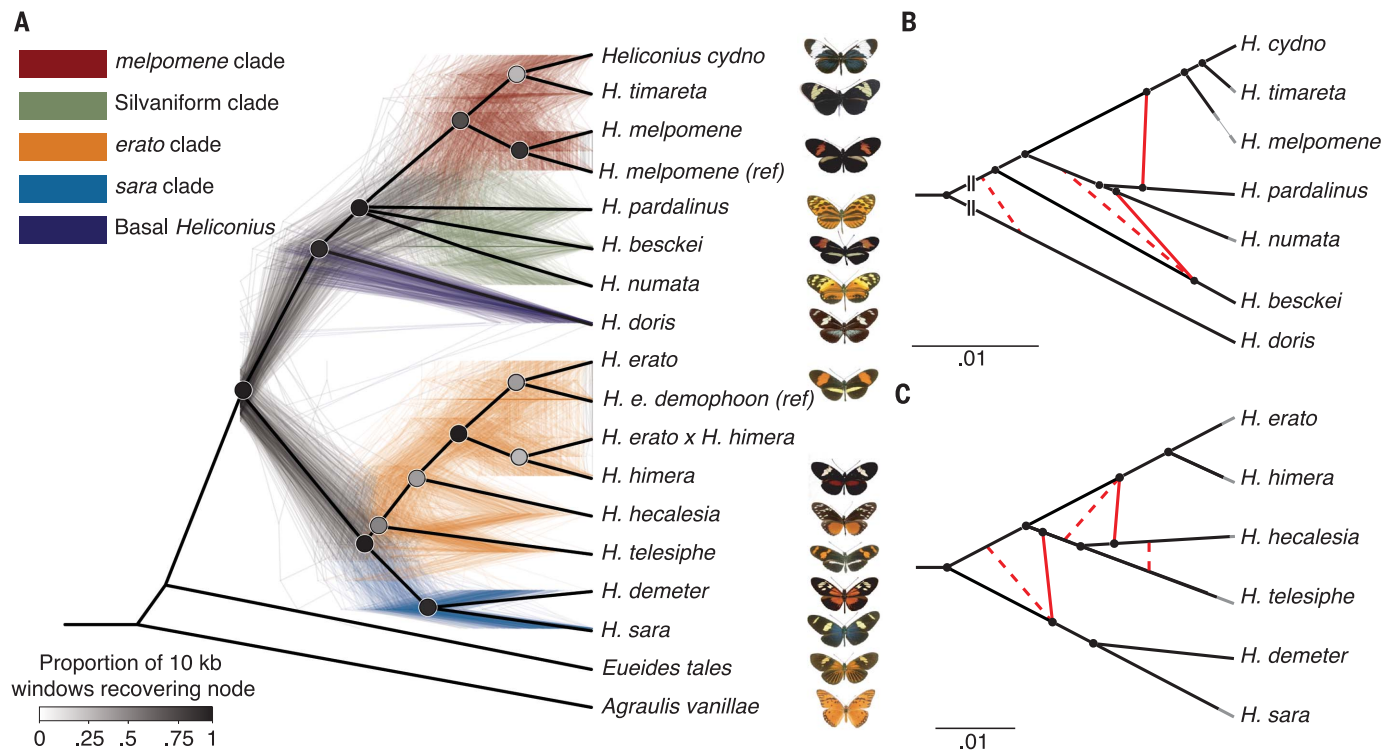
Fig. 1. Phylogeny and phylogenetic networks of *Heliconius* do not support a bifurcating tree. (**A**) All nodes resolved in a majority of species trees are shown in this cladogram (heavy black lines), whereas the poorly resolved silvaniform clade is collapsed as a polytomy (fig. S20). The 500 colored trees were sampled from 10-kb nonoverlapping windows and constructed with maximum likelihood. (**B** and **C**) High-confidence tree structure (black) and introgression events (red) are shown as solid lines. Dashed red lines indicate weakly supported introgression events. Gray branch ends are cosmetic. The *melpomene*-silvaniform clade is shown in (B) and the *erato-sara* clade in (C). Euclidean lengths of solid black lines are proportional to genetic distance along the branches. Scale bars are in units of substitutions per site. Breaks at the base in (B) indicate that the branch leading to *H. doris* has been shortened for display.

locus-by-locus inferences, we developed a statistical test to distinguish between ILS and introgression based on the distribution of internal branch lengths among windows for a given three-taxon subtree, conditional on its topology. We call this method "quantifying introgression via branch lengths" (QuIBL). In the absence of introgression, we expect internal branch lengths of triplet topologies discordant with the species tree (due to ILS) to be exponentially distributed. However, if introgression has occurred, then their distribution should have that same exponential component but also include an additional component with a non-zero mode corresponding to the time between the introgression event and the most recent common ancestor of all three species (*9*). Like other tree-based methods, QuIBL is potentially sensitive to the assumption that each tree is inferred from loci with limited internal recombination (fig. S75). We therefore chose small (5-kb) windows to reduce the probability of intralocus recombination breakpoints.

For every triplet in the *erato-sara* clade, we calculated the likelihood that the distribution of internal branch lengths is consistent with introgression or with ILS only. We formally distinguished between these two models using a Bayesian information criterion (BIC) test with a strict cutoff of ΔBIC > 10. Consistent with our results from *D* statistics, we found that 13 of 20 triplets have evidence for introgression (table S13). For example, using QuIBL on the triplet *H. erato–H. hecalesia–H. telesiphe*, we infer that 76% of discordant loci, or 38% of all loci genome-wide, are introgressed. Averaging over all triplets, we infer that 71% (67% with BIC filtering) of loci with discordant gene trees have a history of introgression, or 20% (19% with BIC filtering) of all triplet loci, indicating a broad signal of introgression throughout the clade [Eq. 7.7, table S13; see (*9*) for additional discussion].

In hybrid populations, individuals have genomic regions that originate from different species and may be incompatible with the recipient genome or with their environment (*16*). Linked selection causes harmless or even beneficial introgressed loci to be removed along with these deleterious loci if they are tightly linked; this effect depends on the strength of selection and the local recombination rate (*17, 18*). We therefore expect introgressed loci to be enriched in regions where selection is likely to be weak, such as gene deserts or regions of high recombination, where harmless introgressed loci more readily recombine away from linked incompatibility loci.

In *Heliconius*, even distant species such as *H. erato* and *H. melpomene* have the same number of broadly collinear chromosomes (*13*), facilitating direct comparisons among species. Furthermore, each chromosome in *Heliconius* has approximately one crossover per meiosis in males (there is no crossing over in female *Heliconius*) (*14, 19*). Chromosomes vary in length, and chromosome size is inversely proportional to recombination rate per base pair (*8, 13*). We found a strong correlation between the fraction of windows in each chromosome that show a given topology and physical chromosome length (Fig. 3A). Such relationships exist for all eight trees in Fig. 2B (*9*), but we focus here on the two most common trees: Tree 1 has a strongly negative correlation with chromosome size ($r^2 = 0.883$, $t = 11.7$, 18 df, $p < 0.0001$), whereas Tree 2 (concordant with our inferred species tree) has a positive correlation ($r^2 = 0.726$, $t = 6.9$, 18 df, $p < 0.0001$). Results from QuIBL indicate that 94% of windows that recover a Tree 1 triplet topology are consistent with introgression (fig. S70 and table S13). The Z (sex) chromosome 21 is strongly

Fig. 2. Local evolutionary history in the erato-sara clade is heterogeneous across the genome. (**A**) Each bar represents a chromosome, in terms of the *H. erato* reference (*14*). Colored bands represent tree topologies of each 50-kb window; colors correspond to the topologies in (**B**), with black regions showing missing data. (**B**) The eight most common trees. The value in the top left corner is the percentage of all 50-kb windows that recover that topology. (**C**) Each histogram corresponds to the topology of the same color in (**B**) and shows the distribution of the number of consecutive 50-kb windows with that topology. Arrows indicate long blocks in inversions.
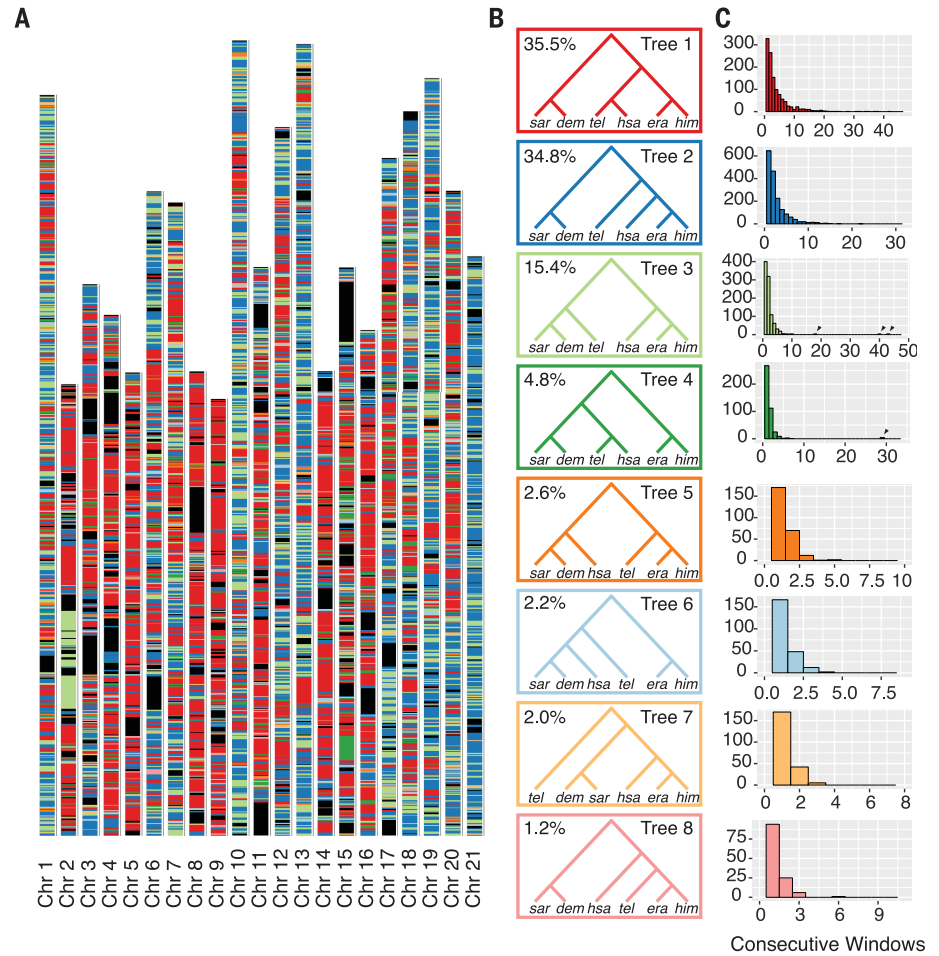
Fig. 3. Chromosomal architecture is strongly correlated with local topology. Tree 1 is shown in red and Tree 2 is shown in blue, as in Fig. 2. (**A**) Tree 1 shows a negative relationship with chromosome size, whereas Tree 2 shows a positive relationship. Lines are linear regressions with chromosome 21 excluded. Numbers along the top indicate chromosome number. (**B**) Each chromosome was divided into 10 equally sized bins, and the occupancy of each topology in each bin was calculated as the number of windows that recovered the topology in the bin divided by the number of windows that recovered the topology in the chromosome. (**C**) Windows are binned by recombination rate, and boxes show the fraction of each tree in each bin for each chromosome separately. Numbers above boxes are the numbers of windows in each bin. (**D**) Boxes showing the relationship of tree topology with coding density. Asterisk denotes significance at the 5% level (paired *t* test, $p < 0.025$). In all boxplots, the central line is the median, box edges are first and third quartile, and whiskers extend to the largest value no farther than 1.5 × (interquartile range).
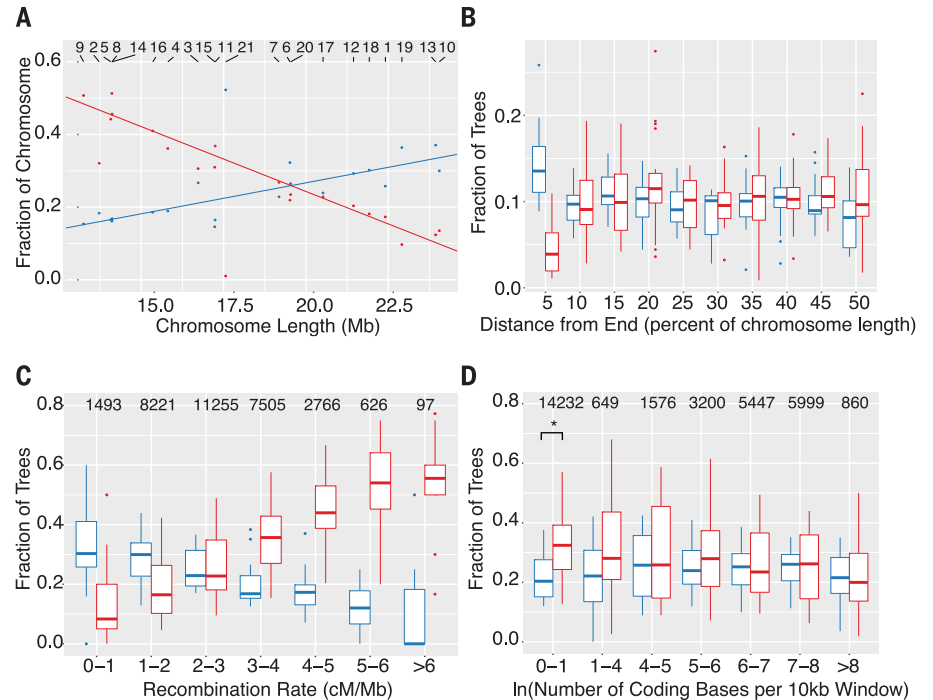
enriched for Tree 2, suggesting that it may harbor more incompatibility loci than autosomes. Interspecific hybrid females in *Heliconius* are often sterile, conforming to Haldane's rule, and sex chromosomes have been implicated as being particularly important in generating incompatibilities (*8, 20–24*).

To test whether the pattern that we observed among chromosomes is related to differences in recombination, we investigated the relationship between recombination rate and tree topology within chromosomes. The recombination rate declines at the ends of chromosomes (fig. S85), and the species tree (Tree 2) is more abundant in those regions
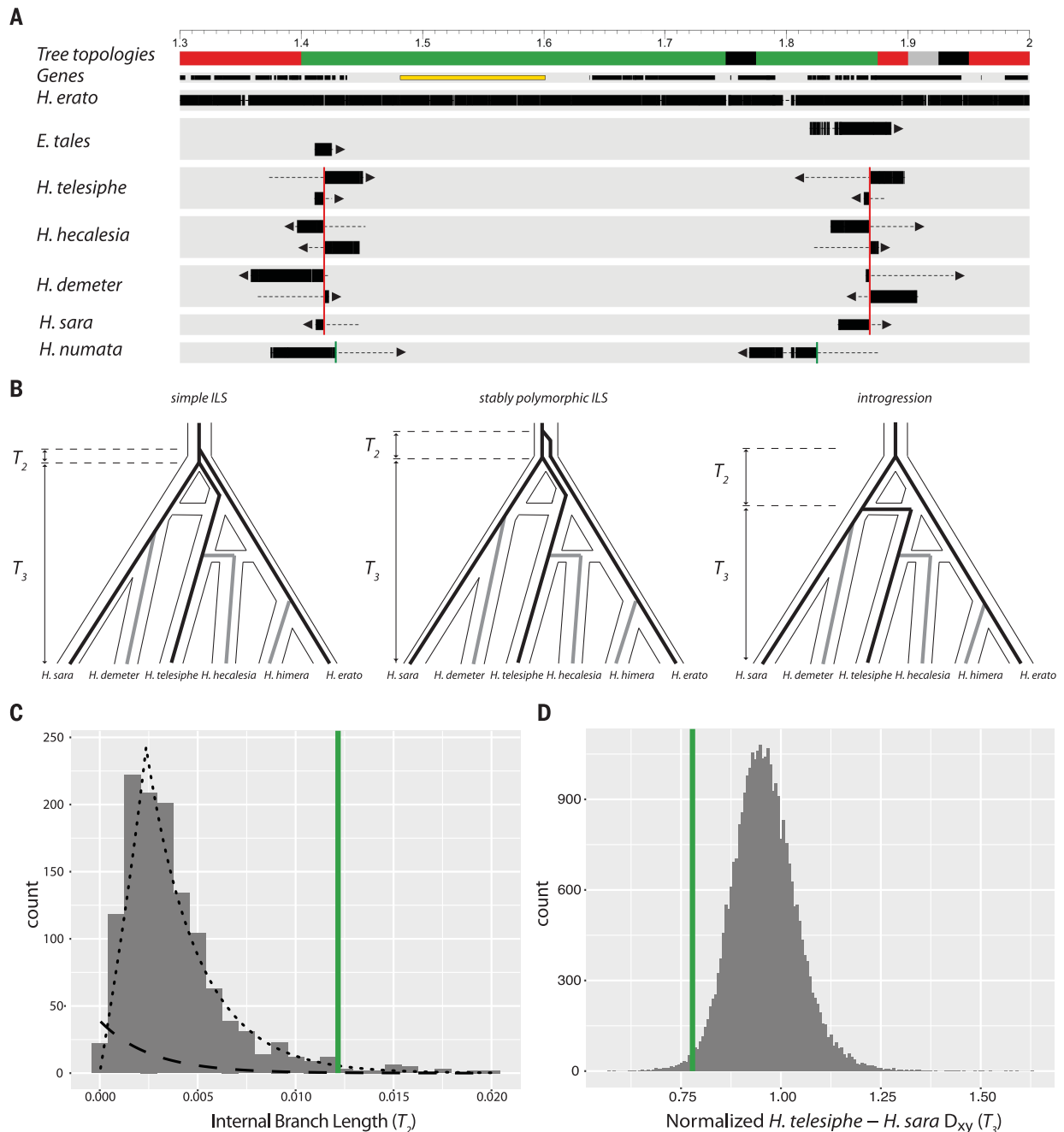
Fig. 4. Parallel evolution of a major inversion at the cortex supergene locus. (**A**) Map of 1.7-Mb region on chromosome 15. Coordinates are in terms of Hmel 2.5 and ticks are in Mb. Tree topology colors correspond to those in Fig. 2. Genes are shown as black rectangles; *cortex* is highlighted in yellow. Each line shows the mapping of a single contig. Aligned sections of each contig are shown as thick bars, whereas unaligned sections are shown as dotted lines. Arrows indicate the strand of the alignment. The *H. erato* group breakpoints are shown with red vertical lines and the *H. numata* breakpoints are shown with green vertical lines. (**B**) Evolutionary hypotheses consistent with the topology observed in this inversion in the context of the previously estimated phylogenetic network. The three species used in the triplet gene tree method, *H. erato*, *H. telesiphe*, and *H. sara*, are shown as black lines; lineages not included are shown as gray lines. (**C**) Histogram of internal branch lengths (T_2) in windows with the topology *H. erato* (*H. telesiphe*, *H. sara*). The inferred ILS distribution is shown as a dashed line, and the inferred introgression distribution is shown as a dotted line. The average internal branch length in the inversion is shown as a green vertical line. (**D**) Histogram of normalized D_{XY} (T_3) between *H. telesiphe* and *H. sara*. Mean normalized D_{XY} in the inversion is shown as a green vertical line.

(Fig. 3B). In addition, when windows were grouped by local recombination rate calculated from population genetic data (*9, 14*), we observed a strong relationship with the recovered topology (Fig. 3C). Finally, we observed a minor enrichment of Tree 1 in regions of very low gene density, but this effect was weak (Fig. 3D) compared with that of recombination. Taken together, these results show that tighter linkage on longer chromosomes and in lower recombination regions within chromosomes leads to removal of more introgressed variation in those regions. This very strong correlation is consistent with a highly polygenic architecture of incompatibilities between species.

Introgression of a convergent inversion

The topology block size distribution in the *erato* clade generally decayed exponentially (Fig. 2C), but two unusually long blocks contained minor topologies: one on chromosome 2 (Tree 3, composed of three sub-blocks) and the other on chromosome 15 (Tree 4). Our study of the ~3-Mb topology block on chromosome 2 confirms an earlier finding of an inversion in *H. erato* (*13*), and we show here that its rare topology is most likely explained by ILS, including a long period of ancestral polymorphism (fig. S95).

The topology block on chromosome 15 is of particular interest because it spans *cortex*, a genetic hotspot of wing color pattern diversity in Lepidoptera (*25, 26*). We hypothesized that this block could be an inversion, as in *H. numata*, where the P_1 "supergene" inversion polymorphism around *cortex* controls color pattern switching among mimicry morphs (*27*). This block recovers *H. telesiphe* and *H. hecalesia* as a monophyletic subclade, which together are sisters to the *sara* clade (Fig. 2B, Tree 4). We searched our de novo assemblies for contigs that mapped across topology transitions. Taking *H. melpomene* as the standard arrangement, we found clear inversion breakpoints in *H. telesiphe*, *H. hecalesia*, *H. sara*, and *H. demeter*. Conversely, *H. erato*, *H. himera*, and *E. tales* all contain contigs that map in their entirety across the breakpoints (Fig. 4A), implying that they have the ancestral *H. melpomene* arrangement.

This chromosome 15 inversion covers almost exactly the same region as the 400-kb P_1 inversion in *H. numata* (*25, 27, 28*). However, de novo contigs from our *H. numata* assembly show that the breakpoints of P_1 are close to but not identical to those of the inversion in the *erato* clade (Fig. 4A). Furthermore, in topologies for *H. numata*, *H. telesiphe*, *H. erato*, and *E. tales* across chromosome 15, not a single window recovered *H. numata* and *H. telesiphe* as a monophyletic subclade, as would be expected if the *erato* group inversion were homologous to P_1 in *H. numata*.

We used QuIBL with the triplet *H. erato*–*H. telesiphe*–*H. sara* to elucidate the evolutionary history of this inversion. A small internal branch would suggest ILS, whereas a large internal branch would be more consistent with introgression (Fig. 4B). The average internal branch length in the inversion was much longer than the genome-wide average, corresponding to a 79% probability of introgression (Fig. 4C). If the inversion were polymorphic in the ancestral population for some time, then we could also recover a similarly long internal branch (Fig. 4B, center). We distinguished between this longer-term polymorphic scenario and introgression by comparing the genetic distance (D_{XY}) between *H. telesiphe* and *H. sara*, represented by T_3 in Fig. 4B. Normalized D_{XY} (as in fig. S95) within the inversion is ~25% less than in the rest of the genome. Given that this is a large genomic block, introgression is therefore the most parsimonious explanation for the evolutionary history of the inversion (Fig. 4D) (*29*).

Discussion

Species involved in rapid radiations are prone to hybridization because of frequent geographical overlap with closely related taxa. In both the *melpomene* and *erato* clades of *Heliconius*, introgression has overwritten the original bifurcation history of several species across large swathes of the genome, a pattern also observed in *Anopheles* mosquitos (*30*). This observation is also consistent with genomic analysis of other rapid radiations characterized by widespread hybridization and introgression, including Darwin's finches (*2*) and African cichlids (*31*). In other radiations, the role of introgression is less clear: in *Tamias* chipmunks, widespread introgression of mitochondrial DNA was identified, in contrast to an absence of evidence for nuclear gene flow (*32*). With few genomic comparisons available to date, it is perhaps too early to say whether introgression is a major feature of adaptive radiations in general, but evidence thus far suggests this to be the case.

Our results raise the question of why some genomic regions cross species boundaries and others do not. In the *erato* clade, we found a strong correlation between recombination rate and introgression probability. Similar associations with topology also exist between sister species in the *melpomene* clade (*8*). Associations between recombination and introgression in hybridizing populations of fishes and monkey flowers (*Mimulus* spp.) support the role of linked selection on a highly polygenic landscape of interspecific incompatibilities (*18, 33, 34*). Our results establish that this relationship persists and may indeed be strengthened with time since introgression. While hybridization is ongoing, many introgressed blocks are constantly reintroduced into the population. If linked to weakly deleterious alleles, introgressed loci will finally be purged by linked selection only long after introgression ceases.

Recombination rate alone cannot account for differential introgression, so we must delve into specific regions to elucidate their function and relevance to speciation. It is critical, therefore, to have tools that can confidently identify introgressed loci, and much effort has gone into developing such methods (*11, 35*). Our test using internal branch lengths in triplet gene trees is based on coalescent theory and takes advantage of the discriminatory power of a property of gene trees not explicitly accounted for by other methods. QuIBL allows us to assess probability of introgression for each locus in each species triplet (*9*). Here, we used this method to identify the evolutionary origin of a convergent inversion that has undergone multiple independent introgression events and to show that genomic regions with discordant topologies arose mostly through hybridization. Just as sex aids adaptation within species, occasional introgression and recombination among species can have major long-term effects on the genome, contributing variation that could fuel rapid adaptive divergence and radiation.

REFERENCES AND NOTES

1. D. Schluter, *The Ecology of Adaptive Radiation* (Oxford Univ. Press, 2000).
2. S. Lamichhaney *et al.*, *Nature* **518**, 371–375 (2015).
3. J. B. Pease, D. C. Haak, M. W. Hahn, L. C. Moyle, *PLOS Biol.* **14**, e1002379 (2016).
4. K. M. Kozak *et al.*, *Syst. Biol.* **64**, 505–524 (2015).
5. K. M. Kozak, O. McMillan, M. Joron, C. D. Jiggins, Genome-wide admixture is common across the Heliconius radiation. bioRxiv 414201 [Preprint]. 11 September 2018.
6. A. V. Z. Brower, I. J. G. Orduña, *Cladistics* **34**, 151–166 (2018).
7. S. H. Martin *et al.*, *Genome Res.* **23**, 1817–1828 (2013).
8. S. H. Martin, J. W. Davey, C. Salazar, C. D. Jiggins, *PLOS Biol.* **17**, e2006288 (2019).
9. Materials and methods are available as supplementary materials.
10. A. G. Clark *et al.*, *Nature* **450**, 203–218 (2007).
11. N. Patterson *et al.*, *Genetics* **192**, 1065–1093 (2012).
12. D. Wen, L. Nakhleh, *Syst. Biol.* **61**, 439–457 (2017).
13. J. W. Davey *et al.*, *Evol. Lett.* **1**, 138–154 (2017).
14. S. M. Van Belleghem *et al.*, *Nat. Ecol. Evol.* **1**, 52 (2017).
15. C. D. Jiggins, *The Ecology and Evolution of Heliconius Butterflies* (Oxford Univ. Press, 2017).
16. J. A. Coyne, H. A. Orr, *Speciation* (Sinauer, 2004).
17. D. J. Begun, C. F. Aquadro, *Nature* **356**, 519–520 (1992).
18. M. Schumer *et al.*, *Science* **360**, 656–660 (2018).
19. J. W. Davey *et al.*, *G3: Genes Genomes Genetics* **6**, 695–708 (2016).
20. T. Dobzhansky, *Genetics and the Origin of Species* (Columbia Univ. Press, 1937).
21. H. A. Orr, M. Turelli, *Genetics* **143**, 613–616 (1996).
22. C. D. Jiggins *et al.*, *Evolution* **55**, 1631–1638 (2001).
23. R. E. Naisbit, C. D. Jiggins, M. Linares, C. Salazar, J. Mallet, *Genetics* **161**, 1517–1526 (2002).
24. S. M. Van Belleghem *et al.*, *Mol. Ecol.* **27**, 3852–3872 (2018).
25. M. Joron *et al.*, *PLOS Biol.* **4**, e303–e310 (2006).
26. N. J. Nadeau *et al.*, *Nature* **534**, 106–110 (2016).
27. P. Jay *et al.*, *Curr. Biol.* **28**, 1839–1845.e3 (2018).
28. M. Joron *et al.*, *Nature* **477**, 203–206 (2011).
29. F. Roda, F. K. Mendes, M. W. Hahn, R. Hopkins, *Mol. Ecol.* **26**, 2317–2330 (2017).
30. M. C. Fontaine *et al.*, *Science* **347**, 1258524 (2015).
31. J. I. Meier *et al.*, *Nat. Commun.* **8**, 14363 (2017).
32. J. M. Good, S. Vanderpool, S. Keeble, K. Bi, *Evolution* **69**, 1961–1972 (2015).
33. Y. Brandvain, A. M. Kenney, L. Flagel, G. Coop, A. L. Sweigart, *PLOS Genet.* **10**, e1004410 (2014).

34. H. F. Gante *et al.*, . *Mol. Ecol.* **25**, 6143–6161 (2016).
35. S. H. Martin, J. W. Davey, C. D. Jiggins, *Mol. Biol. Evol.* **32**, 244–257 (2015).
36. N. B. Edelman, P. B. Frandsen, M. Miyagi, B. Clavijo, J. Davey, R. B. Dikow, G. García-Accinelli, S. M. Van Belleghem, N. Patterson, D. E. Neafsey, R. Challis, S. Kumar, G. R. P. Moreira, C. Salazar, M. Chouteau, B. A. Counterman, R. Papa, M. Blaxter, R. D. Reed, K. K. Dasmahapatra, M. Kronforst, M. Joron, C. D. Jiggins, W. O. McMillan, F. Di Palma, A. J. Blumberg, J. Wakeley, D. Jaffe, J. Mallet, Multispecies, whole-genome alignment and subalignment data for: Genomic architecture and introgression shape a butterfly radiation, Dryad (2019); https://doi.org/10.5061/dryad.b7bj832.
37. N. B. Edelman, P. B. Frandsen, M. Miyagi, B. Clavijo, J. Davey, R. B. Dikow, G. García-Accinelli, S. M. Van Belleghem, N. Patterson, D. E. Neafsey, R. Challis, S. Kumar, G. R. P. Moreira, C. Salazar, M. Chouteau, B. A. Counterman, R. Papa, M. Blaxter, R. D. Reed, K. K. Dasmahapatra, M. Kronforst, M. Joron, C. D. Jiggins, W. O. McMillan, F. Di Palma, A. J. Blumberg, J. Wakeley, D. Jaffe, J. Mallet, Code used in data analysis for: Genomic architecture and introgression shape a butterfly radiation, Zenodo (2019); https://doi.org/10.5281/zenodo.3401692.
38. N. B. Edelman, P. B. Frandsen, M. Miyagi, B. Clavijo, J. Davey, R. B. Dikow, G. García-Accinelli, S. M. Van Belleghem, N. Patterson, D. E. Neafsey, R. Challis, S. Kumar, G. R. P. Moreira, C. Salazar, M. Chouteau, B. A. Counterman, R. Papa, M. Blaxter, R. D. Reed, K. K. Dasmahapatra, M. Kronforst, M. Joron, C. D. Jiggins, W. O. McMillan, F. Di Palma, A. J. Blumberg, J. Wakeley, D. Jaffe, J. Mallet, Code used in data analysis for: Genomic architecture and introgression shape a butterfly radiation, Zenodo (2019); https://doi.org/10.5281/zenodo.3401698.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/366/6465/594/suppl/DC1
Materials and Methods
Figs. S1 to S95
Tables S1 to S14
References (*39*–*91*)

View/request a protocol for this paper from *Bio-protocol*.

Science

Genomic architecture and introgression shape a butterfly radiation

Nathaniel B. Edelman, Paul B. Frandsen, Michael Miyagi, Bernardo Clavijo, John Davey, Rebecca B. Dikow, Gonzalo García-Accinelli, Steven M. Van Belleghem, Nick Patterson, Daniel E. Neafsey, Richard Challis, Sujai Kumar, Gilson R. P. Moreira, Camilo Salazar, Mathieu Chouteau, Brian A. Counterman, Riccardo Papa, Mark Blaxter, Robert D. Reed, Kanchon K. Dasmahapatra, Marcus Kronforst, Mathieu Joron, Chris D. Jiggins, W. Owen McMillan, Federica Di Palma, Andrew J. Blumberg, John Wakeley, David Jaffe and James Mallet

Following gene flow in butterfly genomes

The role of hybridization in evolution and species radiations has long been debated. In *Heliconius* butterflies, introgression was a major factor in their radiation, and the genetic variation it imparted into species is variable across the genome. Edelman *et al.* developed a new sequencing strategy and produced 20 *Heliconius* genomes (see the Perspective by Rieseberg). They also developed a means by which to identify genetic variation that originates from incomplete lineage sorting versus hybridization. Applying this model to their newly developed genomes, they investigated the evolutionary history of the genus and, in particular, the impact of introgression.

Science, this issue p. 594; see also p. 570

ARTICLE TOOLS	http://science.sciencemag.org/content/366/6465/594
SUPPLEMENTARY MATERIALS	http://science.sciencemag.org/content/suppl/2019/10/30/366.6465.594.DC1
RELATED CONTENT	http://science.sciencemag.org/content/sci/366/6465/570.full
REFERENCES	This article cites 86 articles, 17 of which you can access for free http://science.sciencemag.org/content/366/6465/594#BIBL
PERMISSIONS	http://www.sciencemag.org/help/reprints-and-permissions

Use of this article is subject to the Terms of Service