Repeated genetic adaptation to high altitude in two tropical butterflies

Gabriela Montejó-Kovacevich¹, Joana I. Meier^{1,2}, Caroline N. Bacquet³, Ian A. Warren¹, Yingguang Frank Chan⁴, Marek Kucka⁴, Camilo Salazar⁵, Nicol Rueda⁵, Stephen H. Montgomery⁶, W. Owen McMillan⁷, Krzysztof M. Kozak⁷, Nicola J. Nadeau⁸, Simon Martin^{9*}, Chris D. Jiggins^{1*}.

¹*Department of Zoology, University of Cambridge, Cambridge, CB2 3EJ, UK*

²*St John's College, University of Cambridge, Cambridge, CB2 3EJ, UK*

³*Universidad Regional Amazónica de Ikiám, Tena, Ecuador*

⁴*Friedrich Miescher Laboratory of the Max Planck Society, Max Planck Ring 9, 72076 Tübingen, Germany*

⁵*Department of Biology, Faculty of Natural Sciences. Universidad del Rosario. Carrera 24 # 63C-69, Bogotá, Colombia, 111221*

⁶*School of Biological Sciences, University of Bristol, Bristol, UK*

⁷*Smithsonian Tropical Research Institute, Panamá*

⁸*Ecology and Evolutionary Biology, School of Biosciences, University of Sheffield, Sheffield, S10 2TN, UK*

⁹*Institute of Evolutionary Biology, University of Edinburgh, UK*

**S. H. Martin and Chris D. Jiggins contributed equally to this work*

Abstract

Repeated evolution can provide insight into the mechanisms that facilitate adaptation to novel or changing environments. Here we study adaptation to high altitude in two divergent tropical butterflies, *H. erato* and *H. melpomene*, which have repeatedly and independently adapted to high elevations on either side of the Andean mountains. We sequenced 518 whole genomes from elevational transects and found many regions under selection at high altitude, with repeated genetic differentiation across multiple replicates, including allopatric comparisons. In contrast, there is little 'molecular parallelism' between *H. erato* and *H. melpomene*. With a further 85 whole genomes of five close relatives, we find that a large proportion divergent regions have arisen from standing variation and putative adaptive introgression from high-altitude specialist species. Taken together our study supports a key role of standing genetic variation and gene flow from pre-adapted species in promoting parallel genetic local adaptation to the environment.

Introduction

Understanding how organisms adapt to the environment is a central goal of evolutionary biology and highly relevant given the pace of global change. One approach is to explore the repeatability of local adaptation in the wild in order to understand whether phenotypic and genetic changes are predictable. On the one hand, repeated adaptation to similar environments can act as a ‘natural experiment’ and provide the means to identify the targets of selection, by distinguishing locally adaptive from neutral or globally beneficial changes¹. On the other hand, these scenarios can allow us to test whether the same loci are repeatedly targeted across populations and species². Despite many studies reporting repeated adaptation involving the same genes or alleles across lineages^{3–5}, which we here term ‘molecular parallelism’, we know relatively little about the evolutionary mechanisms that facilitate it.

Three main mechanisms can give rise to molecular parallelism in repeated adaptation (Fig.1). Genetic variation upon which selection repeatedly acts may arise via independent mutations at the same gene or locus⁶. Beneficial variants may be recruited from ancestral standing variation⁷ or shared across populations of the same species via migration and gene flow⁸. Lastly, gene flow between species can facilitate the introgression of adaptive alleles^{9–11}. A combination of these mechanisms may also be at play, for instance the high altitude adaptation Tibetan-EPAS1 haplotype was introgressed from Denisovan hominins but remained as neutral standing variation before positive selection occurred¹².

The likelihood of molecular parallelism and the relative importance of each mechanism may be largely dependent on the divergence between the lineages under study (Fig. 1), but this has seldom been empirically tested³. For instance, populations that diverged recently and retain a large pool of standing genetic variation tend to reuse preexisting alleles during repeated adaptation, as seen in freshwater adaptation in sticklebacks¹³, crypsis in beach mice¹⁴, or coastal ecotypes of bottlenose dolphins⁷. Similarly, organisms that readily hybridise in the wild are more likely to share beneficial alleles via adaptive introgression. This allows populations to rapidly adapt to, for instance, novel anthropogenic stressors such as pollutants¹⁵ or insecticides¹⁶.

Anthropogenic change is forcing organisms to move, adapt, or die, with many predicted to expand their ranges towards the highlands to escape warming and degrading lowland habitats¹⁷. Despite insects making up about half of all species that have ever been described¹⁸, we know very little about the genomics or predictability of adaptation to altitude

(but see^{19,20}), especially in the tropics. The type and genomic architecture of the trait under study may also determine its predictability²¹. Phenotypes controlled by few, large-effect loci typically show predictable genetic paths of evolution, such as melanic colouration in mammals, fish, and birds^{22–25}, perhaps due to selective constraints on genetic pathways^{21,26,27}. Organisms adapting to complex environmental challenges that face multifarious selective regimes may show less predictable patterns, with functional redundancy between genes allowing for different combinations of alleles to achieve similar phenotypes^{28,29}. Thus, understanding the relative importance of these mechanisms in determining the predictability of adaptation to the environment could inform future research into conservation strategies to protect biodiversity³⁰.

Here, we study the genetic basis of repeated adaptation to high altitude in two divergent tropical butterflies, *H. erato* and *H. melpomene*. These aposematic, toxic species have very wide ranges and co-mimic each other across South America, commonly found from sea level to around 1600 m³¹. In contrast, most other species in this genus have specialised to highland Andean or lowland Amazonian habitats, with topography and climate shown to correlate with speciation rates across the clade³². Phenotypic differences between highland and lowland populations of *H. erato* and *H. melpomene* have recently been identified, such as in wig shape^{33,34} or heat tolerance³⁵. We search for signatures of local adaptation to high altitude with extensive sampling that harnesses the power of natural spatial replication within and across sides of the Andes, in order to assess the extent of molecular parallelism in adaptation to high altitude. We quantify parallelism at multiple levels of divergence: (i) replicate transects within sides of the Andes connected via gene flow, (ii) trans-Andean allopatric transects with no gene flow at these latitudes, and (iii) two species that diverged 12 million years ago (Fig. 1A). Furthermore, we test whether the same haplotypes are under selection across transects and search for the mutational origin of candidate adaptive alleles. Overall, this large empirical study deepens our understanding of how organisms adapt to new environments and identifies both standing genetic variation and adaptive introgression from pre-adapted species as important mechanisms facilitating local adaptation.

Genetic basis of repeated local adaptation

A. Molecular parallelism across transects and species

B. Potential mechanisms of molecular parallelism

Figure 1. The study of repeated adaptation to the environment and the mechanisms potentially facilitating it. A) We hypothesise that increasing divergence between the lineages under study reduces the likelihood of molecular parallelism (same genes or alleles) underlying repeated adaptation to the environment. In this study, we test this hypothesis by sampling replicate (within sides of the Andes) or allopatric (across sides of the Andes) altitudinal transects of the same species, i.e. connected via gene flow or not (divergence times indicated in Million years ago, Mya), and all replicated in two divergent *Heliconius* species. B) Three main mechanisms can give rise to the genetic variation upon which selection acts repeatedly, giving rise to molecular parallelism: (i) adaptive *de-novo* mutations independently arise in two or more lineages, (ii) existing shared standing variation is repeatedly selected across lineages or shared via gene flow within species, and (iii) adaptive alleles are shared via gene flow across species (adaptive introgression). We tested for the relative importance of these mechanisms with a range of analyses on the relevant transect comparisons. Illustrative tree including four highland populations from four transects (with the Andes preventing gene flow between replicate transects across sides) of our focal species, either *H. erato* or *H. melpomene*, and a lineage of a related high-altitude specialist species from which adaptive introgression is plausible. Sample sizes for the full datasets (including lowland populations when present) are shown in brackets.

Results and Discussion

Divergence and diversity across elevations and transects

To study adaptation to altitude in *H. erato* and *H. melpomene*, we used whole-genome data from 518 re-sequenced individuals, 444 of which were newly sequenced for this study. Samples were collected from 111 different locations, which we grouped into 30 populations, corresponding to four transects: Colombia West/East, and Ecuador West/East (Fig. 2, Table S1). In each transect, populations were either in the highlands (~1200 m) or lowlands (~200m), itself divided into nearby or distant lowland sites (Fig. 2 A). The Andes act as a barrier to gene flow at these latitudes, with populations on opposite sides of the Andes thought to have split ~0.9 and ~2 million years ago for *H. erato* and *H. melpomene*, respectively^{36–38}. Individuals of each species clustered strongly into Western and Eastern groups in genome wide PCAs (S.I., Fig. S1). In PCAs that only included populations from replicate transects of the same side of the Andes (two per species), structuring by altitude was absent in all but one comparison, *H. erato* East, where the *H. erato* highland population in Colombia corresponds to a different colour pattern subspecies and diverged moderately from other populations (S.I., Fig. S1). Intraspecific pairwise differentiation between populations on the same side of the Andes increased with geographic distance but was generally low ($F_{st} < 0.1$; S.I., Fig. S2). The effective replication over space and extensive gene flow within transects provide a powerful setting to study the genomics of parallel adaptation to altitude in the wild.

Genetic distance between populations can be increased by environmental differences that affect dispersal or survival of locally adapted migrants^{39,40}. We tested for such isolation by environment, in our case altitude, with pairwise F_{st} across all populations of the same side of the Andes and species. At similar geographic distances, genetic differentiation was higher when comparing highland vs. lowland populations than when comparing lowland vs lowland populations. Isolation by altitude could be driven by local adaptation reducing gene flow between elevations or due to many adaptive sites diverging across the genome. This difference was stronger in *H. erato* than in *H. melpomene* (S.I., Fig. S2). F_{st} was generally highest when comparing two highland populations at short distances, despite using a topographically informed 'least-cost path' as our measure of geographic distance (S.I. Fig. S3). This could indicate topographical barriers decreasing gene flow, beyond what was captured by the least-cost path, or local adaptation in the highlands leading to increased selection against migrants. A pattern of isolation by environment could also arise due to, for instance, different demographic histories and higher levels of inbreeding in range-edge populations^{41–43}. We found no consistent differences in nucleotide diversity (π) between

elevations (S.I., Fig. S4). Tajima's D was negative across populations, characteristic of population expansion, but generally less negative in highland populations, suggesting less pronounced expansion or more recent/ongoing contractions in the highlands (S.I., Fig. S4). Thus, both heterogeneous demographic histories and selection against locally adapted migrants across elevations may lead to genome-wide isolation by environment.

Figure 2. Sampling design. A) Elevation map of the 30 populations sampled for this study in four geographical transects (Colombia West/East, Ecuador West/East). More details of each population can be found in Table S1, number of whole-genome sequences included per elevation and transect is indicated above each map (*H. erato*, *H. melpomene*). The black scale bar represents 25km. *Heliconius* subspecies present are depicted per transect (credit: McGuire Center for Lepidoptera and Biodiversity, Florida Museum of Natural History), note that only Colombia East has different subspecies in the highlands compared to the lowlands. B) Plot depicts trio sampling scheme, with mean altitudinal and geographical distance of the three population types (high, low, low distant) for both species. C) Mean genome-wide Population Branch Statistic (PBS) trees averaged across the four transects per species and respective total sample size in brackets.

Parallel high-altitude differentiation detected with Population Branch Statistics

To identify genomic regions with high-altitude specific differentiation we calculated Population Branch Statistics (PBS) for three transects in each species, and F_{st} for the remaining two transects with more limited sampling (*H. erato* Colombia West, *H. melpomene* Ecuador West; Fig. 2A). PBS was originally developed to study high-altitude adaptation in humans⁴⁴ and can distinguish between global and lineage specific differentiation by constructing a trifurcating population tree based on F_{st} that includes a geographically distant population^{15,42,45–49}. By attributing fractions of differentiation to each branch, PBS identifies genomic regions disproportionately diverged in the focal population, consistent with positive selection or relaxed constraint^{44,50}.

When assessing genome-wide average PBS trees the longest PBS branches corresponded to the low distant populations of both species, as expected under a model of neutral isolation by distance (Fig. 2B). *H. erato* had a consistently longer high-altitude branch compared to the lowland, but geographically nearby population, which could be indicative of increased drift or extensive local adaptation to altitude throughout the genome. Indeed, we detected many regions across the genome strongly differentiated in high altitude populations across transects and species (Fig. 3). We defined High Differentiation Regions (“HDRs” hereafter) by adding a ± 50 kb buffer around outlier windows, i.e. those with $zPBS_{high}$ (high-altitude branch) or zF_{st} values above 4 (standardised Z-transformed, equivalent to > 4 SD), and merging overlapping intervals into discrete regions. The transects for which only two populations were sampled (F_{st}), had a higher number of HDRs: 400 and 405 HDRs, covering 11.4% and 17.1% of the genome for *H. erato* and *H. melpomene*, respectively (compared to, on average, 229 PBS-based HDRs covering 8%; details on S.I., Note S1). This likely reflects the property of PBS to discern between population-specific and globally differentiated alleles.

To test for molecular parallelism in local adaptation, we assessed whether the same individual HDRs were repeatedly found at high altitude across replicate (same side of the Andes) or allopatric (opposite side of the Andes) transects within each species (Fig. 1A). In *H. erato*, 45% (± 3.8 SD) and 38% (± 4.4 SD) of HDRs overlapped between replicate transects within the Western and the Eastern Andes, respectively (shared HDRs, SHDRs hereafter; Fig. 3D blue, green). Of those SHDRs, more than a third were also shared across allopatric transects not connected via gene flow, 15% of the total HDRs (allopatric SHDRs hereafter; Fig. 3D red). *H. melpomene* had a slightly lower percentage of HDRs shared within sides of the Andes (West 33% ± 11 SD; East 27% ± 11 SD), but very few shared

across sides of the Andes (allopatric SHDRs: 4% of the total). We then tested if the observed level of sharing was higher than expected under a null distribution of genomic regions, obtained by assessing sharing in randomly placed blocks across the genome of the same size and number as observed HDRs per transect. *H. erato* HDR sharing was higher than predicted by our simulations in all three comparisons (replicate Eastern/Western and all transects, Fig. 3B), whereas the low levels of allopatric parallelism in *H. melpomene* did not differ from the null distribution (Fig. 3B, red). The difference between the observed and simulated extent of molecular parallelism of HDRs was robust to the buffer size around outlier windows used for assessing overlap (S.I. Note S1). Our selected buffer of $\pm 50\text{kb}$ is conservative given that well-studied regulatory regions of colour pattern loci in these species can be up to 100kb from the coding sequence itself⁵¹. On average, 15.6% of SHDRs detected overlapped between species, but this fell within the simulated overlap levels given the number and size of SHDRs in each species (mean overlap = $14.06\% \pm 1.19$).

Together, we show that there are high levels of parallelism in high-altitude differentiation within species, but little between the two species. On the one hand, molecular parallelism within species could be facilitated by the high levels of nucleotide diversity observed, indicating a large pool of shared variants upon which selection can repeatedly act. Furthermore, gene flow between populations across replicate transects (within sides of the Andes) could facilitate the recruitment of new or standing adaptive alleles, as expected from theory⁵² and seen in other systems such as maize, *Arabidopsis*, or sticklebacks^{3,8,53}. On the other hand, the lack of significant molecular parallelism in altitude candidate loci between *H. erato* and *H. melpomene*, whose clades diverged 12 million years ago³⁸, is in stark contrast with colour patterns⁵⁴, where near-perfect local Müllerian mimics have arisen repeatedly in both species through independent mutations at a handful of conserved loci^{51,55}. This difference in the extent of molecular parallelism might be explained by the nature of the trait under study: adaptation to altitude is multifarious and, as shown here, involves many genes. Genetic redundancy in polygenic adaptation may lead to evolution taking different paths to reach similar solutions, as shown for quantitative iridescence colouration in these two *Heliconius* species⁵⁶ or in recent experimental evolution of thermal adaptation in *Drosophila*^{28,57}. In contrast, the major effect loci that underlie switches in discrete colour patterning may favour genetic predictability even across divergent taxa, especially when selection is strong and reaching adaptive peaks requires large phenotypic shifts.

Figure 3. Molecular parallelism in PBS/ F_{st} regions of differentiation across eight altitude transects of *H. erato* and *H. melpomene*. A) Number of high-differentiation regions per species (HDRs, including ± 50 kb buffers), in blue/green if shared across replicate transects within sides of the Andes (SHDR: blue=within West, green=within East) and in red those additionally shared across allopatric transects, i.e. shared across all four transects (also SHDR). B) Vertical lines represent percentage of outlier windows shared across transects (jackknife resampling confidence intervals as dashed lines), compared to 10,000 simulations (grey distributions). C) Comparison of SHDR (coloured) and genome-wide (grey) recombination rates (cM/Mb). D-E Patterns of highland-specific differentiation (zPBS_{high}) across the genome in four transects of *H. erato* (D) and *H. melpomene* (E). In the two transects where only two populations were sampled zF_{st} is presented. Horizontal dashed line indicates threshold of 4 standard deviations from the mean. HDRs private to one transect are highlighted in dark grey.

Highly differentiated regions show additional signatures of selection and, generally, are not associated with low recombination rates

We tested whether highly differentiated genomic regions shared across transects (SHDRs) showed additional evidence of positive selection by computing difference in nucleotide diversity ($\Delta\pi$) across elevations ($\Delta\pi = \pi_{\text{high}} - \pi_{\text{low}}$), deviation from neutrality in site frequency distributions (Tajima's D), and absolute divergence (D_{xy}) for the same 5 kb windows. Processes other than positive selection, such as background selection, can decrease within-population diversity and thus lead to increased relative differentiation (F_{st}), especially in the absence of gene flow between populations^{58,59}. Thus, it is important to test for enrichment of different selection statistics to strengthen our inference of locally adaptive loci. A reduced $\Delta\pi$ compared to the background would indicate that a selective sweep in the highlands reduced nucleotide diversity compared to the lowlands. In selective sweep regions, Tajima's D is expected to be low, as regions with selected haplotypes that rapidly increased in frequency would have an excess of rare alleles. Finally, absolute sequence divergence (D_{xy}), is expected to be high in old selective sweeps or variants, and less affected by genetic variation within populations than relative measures of differentiation such as F_{st} ⁶⁰.

SHDRs were considered outliers for $\Delta\pi$, Tajima's D, or D_{xy} , if the observed maximum or minimum values within SHDRs <10th percentile (or >90th in Tajima's D) of the simulated values, obtained from 10000 permutations that randomly placed blocks of equal number and size to observed HDRs across the genome. Of the SHDRs differentiated on both Ecuadorian and Colombian transects but in one side of the Andes only, i.e. across replicate transects, on average 74% and 48% of *H. erato* and *H. melpomene* SHDRs, respectively, were at least outliers for one other statistic, in addition to zPBS/zF_{st} (Fig. 2 A grey). Of the *H. erato* and *H. melpomene* SHDRs shared across all transects (allopatric SHDRs), 94% and 86% had at least one additional outlier statistic, respectively (Fig. 4A, C). In *H. erato*, SHDRs were often outliers for both, high D_{xy} and reduced Tajima's D (36% of SHDRs with additional outlier statistics, on average). In contrast, *H. melpomene* SHDRs were rarely outliers for Tajima's D, whereas 22-33% of them were outliers for both $\Delta\pi$ and D_{xy} . This could point towards different selection histories in each species, with *H. erato* showing signatures of recent or ongoing selective sweeps leading to an excess of rare alleles (negative Tajima's D). The Andean split is dated ~0.5-1 million years older in *H. melpomene* populations^{36,37} and their altitudinal range is wider than that of *H. erato*, which is rarely found >1500 m at these latitudes. Thus, it is possible that *H. melpomene* SHDRs represent more ancient sweeps, reflected in the high prevalence of SHDRs outliers for D_{xy} . Alternatively, high absolute

sequence divergence (D_{xy}) between elevations could be indicative of selected haplotypes arising through adaptive introgression from other species into the highland populations.

As an additional independent line of evidence that SHDRs are under selection, we checked for overlaps with selection statistics and altitude-associated regions obtained from an altitudinal transect in southern Ecuador on the East of the Andes, sequenced with a new linked-read technology called ‘haplotagging’ (Meier *et al.*, 2021). We found that, on average, 59% (± 10 SD) of Eastern SHDRs of each species overlapped with at least one additional outlier selection statistic estimated with the haplotagging dataset, whereas, as expected, fewer Western SHDRs did (32% ± 8 SD, on average; S.I., Note S2, Fig. S5). In contrast, SHDRs shared in all transects showed high levels of overlap with haplotagging-derived selection statistics in all transects (52% ± 15 SD, on average).

We tested if SHDRs were associated with low-recombining regions. Regions of high differentiation and low recombination could be indicative of purifying selection against deleterious mutations (background selection) or maladaptive introgression^{59,61,62}. Background selection has been shown to be a major driver of differentiation landscapes between populations with little gene flow⁶⁰ but is less plausible when populations readily exchange genetic material, as in this study^{63,64}. However, in low recombining regions, selection may be more efficient due to reduced effective gene flow and segregation of co-adapted alleles, and thus facilitate the maintenance of locally adaptive loci^{59,65,66}. Several strongly selected *Heliconius* colour pattern loci have been previously associated with regions of low recombination (Fig. 3 chromosomes 15 and 18)^{55,67}. Nevertheless, here we found that recombination rate at SHDRs did not differ from background levels, except in comparisons that included strongly selected colour pattern loci (Fig. 3C, S.I. Fig. S6). Overall, these additional signatures of selection strongly support the action of repeated divergent selection at high altitude rather than background selection driving the differentiation detected at SHDRs.

Figure 4. Signatures of positive selection across shared high differentiation regions (SHDRs). Number of SHDR with additional outlier selection statistics in *H. erato* (A) and *H. melpomene* (C), statistics included were nucleotide diversity difference between highlands and lowlands ($\Delta\pi$), Tajima's D, and absolute genetic differentiation (Dxy). Shading indicates number of statistics that were above 90th percentile of simulations, white=1 (only zPBS), light grey=2, dark grey=3, and black=4 statistics. Example close-ups of regional zPBS highland values and statistic patterns in Eastern SHDR (#005, B). Each line represents the values for one of the two Eastern transects, solid lines are transects in Colombia and dashed in Ecuador. In this example, all three additional selection statistics ranked as outliers among simulations. Green shading highlights the region of the eastern SHDR with $zPBS_{high} > 4$.

Known genes of interest overlap with SHDRs

We retrieved 908 and 747 genes overlapping with SHDRs in *H. erato* and *H. melpomene*, respectively. With so many potential targets of selection within SHDRs, we do not attempt to infer biological function or adaptive significance from the whole gene set. Instead, we checked for overlaps with regions recently associated with wing shape variation across an altitudinal cline of Southern Ecuador in *H. erato* and *H. melpomene*³⁴. Rounder wings are generally associated with high altitude across 13 species of *Heliconius*³³ and subtle altitude-associated wing shape variation is highly heritable in *H. erato* and *H. melpomene*³⁴. We found that five out of 12 previously identified candidate wing shape loci³⁴, overlapped with SHDRs in *H. erato*, three of which corresponded to SHDRs detected in all transects. In contrast, only two wing shape loci (out of 16) overlapped with *H. melpomene* SHDRs, one of which was an allopatric SHDR. The number of overlaps between candidate wing shape loci

and SHDRs in *H. erato* (n=5) was higher than the 90th quantile of 10,000 permutations, but not in *H. melpomene* (n=2; S.I. Fig S7).

One gene on chromosome 13 stood out, *rugose*, as it was associated with wing shape in both *H. erato* and *H. melpomene*³⁴ and overlapped with SHDRs shared in all transects in both species. In *Drosophila* mutants, *rugose* has been shown to affect social interactions, locomotion, and hyperactivity⁶⁸. The highland incipient species of the *H. erato* clade, *H. himera*, has been shown to fly for more hours per day than lowland *H. erato*, suggesting a potentially important role of locomotion to adapt to highland habitats⁶⁹. Additionally, we found that an *H. erato* Eastern SHDR (Fig. 4B) overlapped with a locus recently identified to be differentiated across many pairs of subspecies in several *Heliconius* species and shown to affect wing beat frequency in *Drosophila*⁷⁰. Thus, future studies could focus on functionally testing some of these candidates and ascertain the potentially adaptive functions of candidate regions.

Same haplotypes underlie parallel adaptation to altitude

High altitude differentiation at the same locus could be driven by the same or different haplotypes under selection. For instance, different *de-novo* mutations at one locus were recently found to confer parallel adaptation to toxic soils in *Arabidopsis*, although most parallel regions were sourced from a common pool of standing alleles⁷¹. To test whether our candidate regions shared the same haplotypes, we performed local Principal Component Analyses (PCA) with outlier windows of each SHDR (Fig. 5A). While ‘global’ PCAs tend to show relatedness between individuals due to geographic structure or partial reproductive isolation, local PCAs of smaller genomic regions can highlight divergent haplotypes due to, for instance, structural variation or positive selection leading to similar haplotypes in adapted individuals^{72,73}. Here we assessed whether genetic variation across individuals at SHDRs (local PCA PC1) could be significantly explained by altitude while accounting for genome-wide (‘global’) structuring (Fig. 5 A) to test for evidence for shared allelic basis for altitude adaptation.

PCAs in each SHDR were performed with individuals from all elevations in transects connected via gene flow (replicate Colombia/Ecuador transects). Local PCAs at SHDRs often showed individual clustering that differed from the neutral geographic expectations (whole-genome ‘global’ PCAs that included Western or Eastern transects, S.I. Fig. S1), and the first axes of variation tended to explain a much larger proportion of the variation observed (PC1 explained 55% ±20 SD compared to 19% in global PCAs, on average; S.I.

Table S3). Out of the four genome-wide PCAs including individuals from replicate transects within sides of the Andes, altitude only explained clustering along PC1 in the Eastern *H. erato* transects (S.I. Table S3, Figure S1). This can be explained by a different highland colour-pattern subspecies in Colombia, *H. e. dignus*, reducing gene flow across the genome (Fig. 2). In contrast, we found that a large proportion of local SHDR PCAs had PC1s that correlated significantly with individual altitude ($P < 0.05$, *H. erato*: East=48%, *H. melpomene*: West=74%, East=66%, S.I. Table S3), except in *H. erato* West where only 25% did (Fig. 5B, S.I. Fig. S8-11). Altitude explained, on average, 12% of the variation in local PC1 while controlling for the global PC1 (altitude partial R^2 , *H. erato*: West= 0.05, East=0.16, *H. melpomene*: West=0.10, East=0.15, S.I. Table S3).

Local PCAs can also highlight divergent haplotypes, putatively associated with inversions, by clustering individuals that possess homozygous or heterozygous haplotypes in those regions^{74,75}. We found several *H. erato* SHDRs within a high differentiation block in chromosome 2 (6.5 Mbp, Fig. 5 B), ~0.75 Mbp downstream from a recently identified inversion exclusively present in lowland individuals of southern Ecuador⁷⁶. We performed an additional local PCA across the large putative inversion and found a three-cluster pattern, consistent with the presence of the three inversion genotypes (Fig. 5 B), and a neighbour-joining tree with outgroups supported its appearance in the Eastern lowlands (S.I. Fig. S12). Local PCAs of SHDRs within the inversion region correlated more strongly with altitude than putatively inverted-only haplotypes, indicating that adaptive loci remain differentiated in the highlands and may pre-date the inversion event in the lowlands (S.I. Fig. S8). This is consistent with a model in which the inversion could enhance local adaptation by reducing gene flow between elevations at pre-existing locally-adapted alleles⁷⁷.

Overall, the majority of SHDRs involve the same alleles across transects connected via gene flow. Those SHDRs that do not correlate with altitude could represent cases where different de-novo mutations arose at the same or nearby loci or facilitated by independent introgression events, or indeed may represent regions of differentiation not under positive selection (i.e. false positives). Furthermore, the large putative inversion found exclusively in the lowlands may represent a case of structural variation facilitating adaptation in the highlands⁷⁷. Recent studies on environmental adaptation in seaweed flies and sunflowers, among others, have demonstrated a key role for inversions in maintaining adaptive alleles together and facilitating the evolution of locally adapted ecotypes^{72,73}. By studying individual clustering across differentiated loci we have shown that the same alleles often drive parallelisms involved in local adaptation. We thus next turned to identifying the source of the genetic variation causing molecular parallelism across populations.

A. Same alleles at SHDR repeatedly selected within sides of Andes

B.

C. *H. erato* (East) chromosome 2 putative inversion

Figure 5. Allele sharing SHDRs and large putative inversion in chromosome 2 of *H. erato* Eastern transects. A) Example analysis to test whether same alleles underlie SHDRs (here depicted *H. erato* Eastern SHDR #77). First, outlier windows (zPBS>4) in either the Eastern Colombia (solid black line) or the Eastern Ecuadorian (dotted grey line) transect are selected (grey panel), and a local PCA with those sites is performed. Then we test whether PC1, the axis explaining most of the variation, is significantly explained by the altitude at which individuals were collected, while controlling for the global PC1 (i.e. neutral population structure). Each point represents an individual, their shape represents transect of origin (Colombia filled, Ecuador empty symbols) and their colour the altitude at which individuals were collected. B) Number of SHDRs where altitude is a significant predictor of local PCA axis 1. C) zPBS highland differentiation in Eastern Colombia (solid black line) and Ecuadorian (dotted grey line) transects across a 6.5 Mbp region of *H. erato* chromosome 2. zPBS lines are regionally smoothed with rolling means of 200 windows, thus some individual outlier windows are higher in value (Fig. 3). SHDRs are shown as vertical segments, colored by whether they represent SHDRs within Eastern transects (green) or allopatric shared across all transects (red, SHDRs shown in Fig. 3). The Pearson correlation coefficient between the putative inversion local PCA PC1 and altitude is shown (**** $p < 0.0001$). Homokaryotes for the wildtype arrangement are named wt/wt, heterokaryotes are inv/wt and inversion homokaryotes are labelled inv/inv. The most common arrangement clustered with the Western transects of *H. erato* and outgroups of other species in a neighbour-joining tree, and thus was considered the most likely, non-inverted haplotype (wt/wt) (S.I. Fig. S12).

The source of parallelism: standing variation and adaptive introgression with high-altitude relatives

The presence of the same putatively adaptive haplotypes on several transects could either reflect: (i) standing variation being repeatedly selected at high altitudes or shared via intraspecific gene flow, or (ii) recruitment of adaptations from other high-altitude adapted species through introgression. With five high-depth individuals per population of *H. erato* and *H. melpomene*, and 116 additional whole-genomes of high-altitude specialist species and outgroups, we tested for signatures of shared standing variation within species and of adaptive introgression in shared high differentiation regions (SHDRs).

To test for excess allele sharing at SHDR, we calculated the F_{DM} statistic in 50 kb windows across the genome^{78,79}. For each test, we used a tree with four populations (((P1, P2), P3), O), where P1/P2 reflect the lowland and highland populations, respectively, and P3 is an allopatric high-altitude population or a sympatric high-altitude specialist species (Fig. 6A). Positive F_{DM} values indicate excess allele sharing between P3 and P2 (i.e. between non-sister high-altitude lineages), and negative values indicate excess allele sharing between P3 and P1 (i.e. between non-sister high- and low-altitude lineages, Fig. 6A). We then tested if SHDRs are enriched for outlier positive F_{DM} (i.e. excess allele sharing with the highlands), using the distribution of absolute negative F_{DM} across SHDRs as a null (see Methods for details). This specifically tests whether genomic regions that are differentiated in high-altitude populations (SHDRs) are systematically enriched for alleles shared with allopatric high-altitude populations of the same species or with sympatric specialist high-altitude species.

We first assessed intraspecific allele sharing between allopatric highland populations in transects on opposite sides of the Andes, which split ~0.9 Mya and ~2 Mya in *H. erato* and *H. melpomene*, respectively^{36,37}. Signatures of allele sharing likely represent shared ancestral standing variation that pre-dates the Andean split (Fig. 1B), but we cannot rule out gene flow via distant contact zones in the north and south edges of the Andes or periods of secondary contact in the past³⁷. Nearly half of *H. erato* Eastern SHDRs had outlier excess allele sharing with the Western highlands in Colombia, whereas only 9% of Western SHDRs did (Fig. 6B, S.I. Fig. S15). Both comparisons resulted in a significant enrichment of excess allele sharing between allopatric highland populations compared to sharing with the lowlands across all SHDRs (Kolmogorov-Smirnov tests $P < 0.05$ as stars in Fig. 6B, S.I. Fig. S15). *H. melpomene* only showed significant enrichment of excess allele sharing between allopatric highland populations of Ecuador in Eastern SHDRs, and the percentage of SHDRs with

excess allele sharing was generally lower than in *H. erato* (Fig. 6B). Overall, shared standing variation has been an important mechanism facilitating molecular parallelism in *H. erato*, in which trans-Andean populations share a more recent common ancestor.

We then explored allele sharing between highland populations of *H. erato* and *H. melpomene* and five sympatric high-altitude specialist relatives or those (Fig. 6C). These high altitude species are known to readily or occasionally hybridise with *H. erato* or *H. melpomene*^{37,80–82}. To quantify genome-wide evidence of allele sharing we first computed f_{branch} statistics, which test for gene flow between branches of a phylogeny⁸³. As expected, we found evidence of excess allele sharing between all relevant pairs (details in S.I. Note S3, Fig. S13-14)⁸⁴. Excess allele sharing at SHDR between highland populations and sympatric high-altitude specialist species likely represent cases of adaptive introgression. Generally, the more closely related putative donor and recipient species were, the higher proportion of SHDRs that showed excess allele sharing (Fig. 6B, S.I. Fig. S16). For instance, levels of allele sharing in SHDRs were much higher between highland *H. erato* and the closely related *H. himera* than with the distantly related *H. telesiphe* (Fig. 6B). Interestingly, in Eastern *H. melpomene*, SHDR sharing with a closely related sympatric species, *H. timareta*, was more prevalent in SHDR than shared variation with allopatric highland populations of its own species (S.I. Fig. S16D). Admixture between adjacent *H. melpomene* and *H. timareta* populations is well-documented, with strongly selected colour pattern loci having been shared across the species barrier^{81,82,85}

Levels of putative adaptive introgression at SHDRs were also high between *H. himera* and highland *H. erato*, and significantly enriched across SHDRs in Colombia (Fig. 6 B). *H. himera* is a closely related species that split from within the Eastern *H. erato* clade 215,000 - 527,000 years ago³⁷, with pre-mating isolation and a range of divergent life-history phenotypes adapted to the highland dry forests it inhabits^{86–88}. Admixture is predominantly from *H. himera* into *H. erato*³⁷, supporting our hypothesis that the high levels of excess allele sharing at SHDRs between the two may represent cases of adaptive introgression into *H. erato*. Our study is the first to show that potentially adaptive alleles other than colour pattern loci have been shared between high-altitude specialist species and *H. erato* and *H. melpomene*, likely facilitating their expansion into the highlands.

The 6.5Mbp putative inversion detected in chromosome 2 of Eastern *H. erato* individuals showed high levels of allele sharing between highland *H. erato* populations and *H. himera*, whereas there was no excess allele sharing with either the highlands or the lowlands when the putative donor species was a distantly related species (S.I. Fig. S17). A neighbour-

joining tree of this region revealed that lowland distant individuals that clustered in the local PCA formed a monophyletic group divergent from all other *H. erato* populations, including allopatric populations in the west of the Andes (S.I. Fig. S12). This suggests that the inversion may have arisen anciently, prior to the western and eastern Andean split of *H. erato*. Its maintenance in the lowland populations may protect locally adaptive alleles from maladaptive migration load and/or promote the accumulation of novel, locally adaptive mutations⁷³. Furthermore, its absence in the highlands allows for ongoing gene flow between highland *H. erato* populations and the closely related highland specialist *H. himera*. Supergene evolution in another species of this genus, *H. numata*, has been linked to the introgression of a chromosomal inversion⁸⁹, highlighting the role of structural variation and hybridisation in providing novel genetic architectures that can promote adaptation. Future work could investigate the potential role of this inversion in maintaining locally beneficial allele clusters and their associated adaptive phenotypes.

Figure 6. Many SHDRs were sourced from standing variation and putative adaptive introgression from highland-specialist species. A) Tree used to estimate F_{DM} values per 50kb window across the genome in each comparison, which when positive represents excess allele sharing between P2, a highland *H. erato* or *H. melpomene* population, and P3, an allopatric highland population of the same species or a sympatric high-altitude specialist species, compared to a lowland population (P2). Colours of P3 populations or species indicate the potential mechanism driving the excess allele sharing, either intraspecific shared ancestral standing variation (blue) or adaptive introgression from closely (orange) or distantly related (yellow) high-altitude specialist species. B) Excess allele sharing at SHDRs between P3 (putative highland donors, left y axes) and P2 (putative highland recipients, right y axes) across the *H. erato* (top) or *H. melpomene* (bottom) comparisons (phylogeny from Kozak *et al.*, 2015)¹. Left panel shows mean maximum F_{DM} ($\pm$ S.E.) across SHDRs (western SHDRs if the putative recipient was on the West of the Andes, and vice versa) of the Colombian (solid triangles) and Ecuadorian (unfilled triangles) transects. Background mean maximum F_{DM} values were obtained from 1000 block permutations across the genome and shown in grey. Stars represent comparisons where distribution of maximum F_{DM} (excess allele sharing with the highlands) was significantly higher than absolute minimum F_{DM} (excess allele sharing with the lowlands) distribution across all SHDRs (Kolmogorov-Smirnov tests $P < 0.05$, S.I. Fig. S15-15). Right panels show percentage of SHDRs with evidence of excess allele sharing between P2 and P3, considered significant if they had outlier maximum F_{DM} ($>90^{th}$ percentile of absolute minimum F_{DM} across all SHDRs). Abbreviations not depicted: "era" *H. erato*, "mel" *H. melpomene*. C) Putative donor (P3) high-altitude specialist species.

Conclusions

By studying recently and anciently diverged populations at different altitudes within and across sides of the Andes of two species we have uncovered (i) strong signatures of high-altitude differentiation in narrow regions across the genome, consistent with positive selection, (ii) high levels of molecular parallelism between transects of the same species but no sharing across species, and (iii) an important role of standing variation and adaptive introgression from high-altitude specialist species in adaptation to these environments. The overall lack of molecular parallelism across species points towards genetic redundancy of polygenic evolution that allows different combinations of alleles to confer adaptation to the same environments²⁸. The evolutionary success of *H. erato* and *H. melpomene* in colonising a wide range of elevations has likely been facilitated by abundant genetic diversity, as well as by intra- and interspecific gene flow allowing for the sharing of pre-existing adaptive alleles. Together, our study highlights the value of extensive replication across space and large whole-genome datasets for understanding the molecular underpinnings of local adaptation in the wild. Both standing genetic variation and recent hybridization can supply the selection targets required for adaptation to a new environment, which emphasizes the importance of preserving gene flow and connectivity between populations if organisms are to adapt to everchanging environmental pressures.

Methods

STUDY SYSTEM AND WILD BUTTERFLY COLLECTION

H. erato and *H. melpomene* can be found across most of the Neotropics and have Müllerian aposematic mimicry to advertise their toxicity to predators, thus share colour pattern when inhabiting the same areas⁹⁰. They can be found continuously coexisting across elevational transects ranging from sea level up to 1600 m along the Andean mountains, and *H. melpomene* can be found across elevations up to 1800 m. Butterflies were collected from 111 different locations, which we grouped into 30 populations, corresponding to four transects Colombia West/East, and Ecuador West/East (Table S1). In each transect populations were either in the highlands (altitude mean=1235 m), lowlands (altitude mean=364 m), or distant lowlands (altitude mean=95 m) to control for genetic drift due to isolation by distance (Fig. 2A). The Andes acts as a barrier to gene flow, as elevations in these latitudes are too high for butterflies to fly across and have been for at least 8 million years⁹¹, which pre-dates the expansion of both species across these latitudes^{36,37}. All but one of these transects (Colombia East) had the same subspecies, i.e. geographic colour morph, in the three elevations, to avoid differentiation due to highly divergent colour pattern loci (Fig. 2A). Detached wings were stored in glassine envelopes and bodies in EtOH (96%) vials. We additionally collected high-altitude specialist relatives of *H. erato* and *H. melpomene* that have potential for admixture between them. The *H. erato* relatives were *H. himera* and *H. telesiphe* from the Eastern Andes, and *H. clysonymus* which is found on both sides of the Andes. The *H. melpomene* relatives were *H. timareta* and *H. cydno*, from the Eastern and Western Andes, respectively. More distantly related outgroups were also sampled, *H. eleuchia* and *H. hecale* for *H. erato* and *H. melpomene*, respectively.

WHOLE-GENOME SEQUENCING

Whole genome sequence data from 518 individuals was analysed in this study, 444 were newly sequenced here, while the rest were obtained from published studies (n=74). Of the newly sequenced individuals 365 were sequenced at low-medium depth with BGI (~6X), and 79 were sequenced at high depth with Novogene (~18X-30X), at least 5 per population. For the high-altitude specialist species dataset and outgroup species, we obtained high-depth whole genome sequencing data for 116 individuals, 63 of which were newly sequenced for this study at ~20X depth with BGI. A full list of individuals, localities, and accession numbers can be found in Supplementary Table 1. Individuals with *H. melpomene malleti* phenotypes (Fig. 1) were genotyped with a restriction digest, following Nadeau *et al.* 2014⁵⁴, to identify cryptic individuals of the species *H. timareta ssp. nov.*, which are indistinguishable phenotypically from *H. m. malleti*. We extracted DNA with QIAGEN DNeasy Blood and

Tissue extraction kits, including RNA removal, and confirmed DNA integrity and concentration (minimum of 10 ng/μL) using Qubit. DNA samples were stored at -20°C until library preparation. For the individuals that were sequenced with low-medium depth, a secondary purification was performed with magnetic SpeedBeads™ (Sigma) and we prepared Nextera DNA libraries (Illumina, Inc.) with purified Tn5 transposase⁹². PCR extension with an i7-index primer (N701–N783) and the N501 i5-index primer was performed to barcode the samples. Library purification and size selection was done using magnetic SpeedBeads™ (Sigma). We confirmed adaptor lengths through TapeStation High sensitivity T1000 (Agilent Technologies, CA, USA) and gel electrophoresis. Pooled libraries were sequenced by The Beijing Genomics Institute (China) using HiSeq X Ten (Illumina). Library preparation and sequencing of the high-depth *H. erato* and *H. melpomene* individuals was carried with HiSeq X platform (150bp paired-end) by Novogene.

STATISTICAL ANALYSES

All non-genomic analyses were run in R V2.13 (R Development Core Team 2011) and graphics were generated with the package *ggplot2*⁹³.

Read mapping and genotype calling

We aligned the sequence data of all individuals of the two focal species and their relatives to their corresponding reference genomes, either *H. melpomene* version 2.5^{94,95} or *H. erato demophoon*⁵⁵, obtained from Lepbase⁹⁴, using bwa mem (v 0.7.15 Li, 2013). We used samtools (v 1.9 Li et al., 2009) to sort and index the alignment files. Duplicates were removed using the MarkDuplicates program in Picard tools (v 1.92 Broad Institute, 2018). Genome-wide mean sequencing depth was calculated with samtools (v 1.9 Li et al., 2009). Mean sequencing depth was very similar across areas of *H. erato* (mean=8.93, Fig. S18 and was generally higher for *H. melpomene* (mean= 12.3), but more variable, especially in Colombia where many sequences were obtained from published studies (Supplementary Table 1). Most of the analyses described below for *H. erato* and *H. melpomene* were performed with genotype likelihoods in ANGSD and low or variable sequencing depths are thus accounted for.

However, for our phylogenetic datasets combining our samples with other species for phylogenetic tree reconstruction and tests of admixture, we restricted the *H. erato* and *H. melpomene* samples to the five individuals per population with high sequencing depth. We mapped the high-altitude specialist relatives and outgroups of *H. erato* and *H. melpomene* to the respective reference genomes as explained above (Supplementary Table 1). We used a

genotype calling approach with GATK v. 3.7⁹⁹ to obtain a vcf file each for the *H. erato* and *H. melpomene* clade. Genotypes were called with HaplotypeCaller for each individual and variants were then called with GenotypeGVCFs across all individuals combined. The vcf files were filtered with vcftools v. 0.1.15¹⁰⁰ to remove genotypes with less than 3 reads, monomorphic sites, multi-allelic sites, insertions and deletions (indels), and sites with more than 50% missing data.

Isolation by distance and Isolation by environment

To study Isolation by Distance (IBD) and Isolation by Environment (IBE), we first calculated all pairwise genetic differentiation (F_{st}) between all populations on each side of the Andes that had at least 5 individuals each, i.e. *H. erato* west ($n_{\text{population}}=7$), *H. erato* east ($n_{\text{population}}=11$), *H. melpomene* west ($n_{\text{population}}=7$), *H. melpomene* east ($n_{\text{population}}=9$), hereafter “side-species replicate” (Table S1). We calculated pairwise population genetic distance with the function `calculate.all.pairwise.Fst()` from the R package BEDASSLE³⁹. This requires a matrix of allele count data, with populations as rows and number of biallelic unlinked loci sampled as columns, which we obtained with ANGSD and custom scripts. First, we obtained a list of highly polymorphic SNPs per side-species replicate, by (i) heavily filtering 10 random individuals and obtaining minor allele frequencies (`-doMaf 1`) by forcing the major allele to match the reference state, so that it is the same across all populations (`-doMajorMinor 4`) (ii) extracting the sites and the major/minor allele frequencies, (iii) creating an indexed sites file (`angsd sites index`), (iv) subsetting so that sites are at least 2kb apart, to prune for linkage disequilibrium. This list of sites and regions was then used to obtain minor/major allele frequency counts with all individuals per population and forcing the major/minor allele to match the ones given by the sites file (`-doMajorMinor 3`). From the resulting allele frequencies per population, we calculated allele counts, by multiplying by the number of individuals per population and the number of chromosomes samples (2, diploid). We obtained the required allele count matrix by concatenating all populations per side-species replicate, and only keeping loci with allele counts for all populations.

Geographical distance between populations measured as a straight line through the landscape is not biologically representative of organisms moving through space. To account for topographic complexity, we obtained topographic least cost paths with the R package *topodistance*¹⁰¹. With historical records from the Earthcape database¹⁰², we created a binary habitat suitability raster based on the elevational range of each species, so that least cost paths between populations never included elevations that these butterflies do not inhabit. Then we used the function `topoLCP()` to get the least cost path distance between

populations (Fig. S3, Wang, 2020), and use this distance as a proxy of isolation by distance between populations.

Differentiation and Selection statistics

To search for signatures of local adaptation to high altitude we use a measure of lineage-specific differentiation, population branch statistics (PBS)⁴⁴. PBS is a summary statistic based on pairwise genetic differentiation (*Fst*) among three populations, two of which are located closely geographically (high, low) and one distant outgroup (low distant). For each population trio (high, low, low distant Fig. 1 A), we computed PBS with ANGSD¹⁰³. We first obtained genotype likelihoods to calculate the site-frequency spectrum (SFS) per population. Then, we computed 2D-SFS for each population pair (high-low, high-low.distant, low-low.distant) with the function *realSFS*. We then used the 2D-SFS as a prior for the joint allele frequency probabilities at each site, which are used to compute per-site pairwise Hudson's *Fst*¹⁰⁴ as interpreted by Bhatia (Bhatia et al., 2013, *realSFS* *fst* index -whichFst 1) among the three populations (*Fst*_{high-low}, *Fst*_{high-low.distant}, *Fst*_{low-low.distant}) and PBS per population (PBS_{high}, PBS_{low}, PBS_{low.distant}). This is achieved by first transforming pairwise *Fst* values into relative divergence times:

$$T = -\log(1 - Fst)$$

To then obtain PBS for a given population (here the highlands):

$$PBS_{high} = \frac{T_{high-low} + T_{high-low.distant} - T_{low-low.distant}}{2}$$

This quantifies the magnitude of allele frequency change in the highland lineage since its divergence from the lowland and lowland distant populations. Regions of the genome with large PBS values represent loci that have undergone population-specific sequence differentiation consistent with positive selection. We computed weighted *Fst* and PBS averages per 5kb window size and 1kb steps (*realSFS* *fst* stats2), with the same window centres across datasets (-type 0). For transects where only two out of three populations had any individuals (Colombia west for *H. erato* and Ecuador west for *H. melpomene* Fig. 1 A), we computed pairwise *Fst* only. Finally, we normalized PBS and *Fst* window values with z-scores, i.e. number of standard deviations from the mean, so that divergence is comparable across transects.

HDR parallelism between replicates

To measure genetic parallelisms in adaptation to high altitude, we quantified overlap of outlier windows and adjacent regions, i.e. high differentiation regions (HDR), across replicate and allopatric transects and species. We considered outlier windows to be those with values above 4 standard deviations from their mean (zPBS_{high}>4, zFst>4). We then expanded

outlier window positions 50kb upstream and 50kb downstream from the window center (HDR) and checked for overlaps with other transects, either within or across sides of the Andes for each species, to highlight parallelism when visualizing patterns of genome-wide divergence. High-differentiation regions with any overlap with other transects are termed shared HDRs (SHDRs). To check for HDR overlaps between species, we mapped the *H. melpomene* windows (starts and end positions) to the *H. erato* reference genome using a chainfile from Meier et al. (2021)⁷⁶ and the liftover utility (Hinrichs, 2006).

To test whether the level of parallelism observed between transects within and across sides of the Andes was higher than expected by chance, we used the R package *intervals*¹⁰⁶. We first created outlier window intervals (HDRs), by obtaining the start and end positions of continuous blocks of outlier windows (± 50 kb buffers) with the function `Intervals()` (options `type="Z"`, closed). We obtained the observed proportion of total intervals that overlapped, at any of their positions, with outlier intervals in the other transect within sides of the Andes, or with outlier intervals in both, the other transect within the same side of the Andes and the two transects on the other side of the Andes (allopatric sharing). We calculated confidence intervals around the proportion of overlapping intervals by performing random jackknife block resampling across the genome. We then simulated 10,000 randomized distributions of outlier-window intervals across the genome per transect, per species ($n=8$). In each simulation set, we randomly placed the same number of HDR intervals and of the same size as the observed outlier-window intervals for those transects. With these, we estimated the proportion of simulated intervals that overlapped with observed HDR outlier intervals within and across sides of the Andes, obtaining as a result a null distribution of interval overlap proportions.

Measures of nucleotide diversity, selective sweeps, and recombination rate

We studied genetic variation within and across populations by deriving three summary statistics from ANGSD thetas estimations¹⁰³, Tajima's D, which estimates the deviation of a sequence from neutrality, nucleotide diversity (π , or population mutation rate), and Dxy or absolute divergence, which calculates pairwise differences between sequences of two populations excluding differences between sequences within populations. ANGSD has been found to be an accurate estimator of nucleotide diversity because it includes invariant sites¹⁰⁷. Firstly, we obtained folded global site-frequency spectra for each population. Then we calculated pairwise nucleotide diversity per site (`thetaD`, `realSFS saf2theta`). Finally, we performed sliding window analysis of 5kb window size and 1kb steps (`thetaStat do_stat`) to obtain sum of pairwise differences, Tajima's D, and total effective number of sites per window. Nucleotide diversity (π) was obtained by dividing the sum of pairwise differences by

the total number of sites per window. For all transects, we calculated the difference in nucleotide diversity per window between highland and lowland distant populations (or low populations if low distant individuals unavailable), which is expected to be negative if a selective sweep led to locally reduced diversity in highland populations. Absolute divergence (D_{xy} , Nei, 1978) between high (population A) and low/low distant (population B) populations was estimated by additionally obtaining pairwise nucleotide diversity per site (θ_{AB}) for all individuals pooled from populations A and B (population AB), and then per-site D_{xy} obtained:

$$D_{xy_{site}} = \frac{\theta_{AB} \times \binom{n_{AB}}{2} - \theta_A \times \binom{n_A}{2} - \theta_B \times \binom{n_B}{2}}{n_A \times n_B}$$

With n being the number of individuals per population (A, B, or pooled AB) and θ_{AB} obtained from realSFS saf2theta. Mean D_{xy} was estimated for the same 5kb windows with 1kb steps. Finally, we obtained recombination rate for 50kb windows along the genome from the mean population recombination rate ($\rho = 4N_e r$, r = probability of recombination per generation per bp) estimated for 13 *H. erato* populations from across the range in a recent study¹⁰⁹ and for 100kb windows of four *H. melpomene* populations⁶⁷. Note that the *H. erato* genome is larger than *H. melpomene* (383 Mb and 275 Mb, respectively), hence the difference in window sizes.

Testing for significance of selection statistics in SHDRs

To test for positive selection we assessed Tajima's D, difference in nucleotide diversity across elevations $\Delta\pi$ ($\pi_{high} - \pi_{low}$), and absolute divergence (D_{xy}) within SHDRs and compared values to simulated distributions. We used the same permutation approach described for assessing HDR parallelism, by randomly placing the same number of intervals and of the same size as the observed HDRs for each transect 10000 times. We then obtained minimum Tajima's D and $\Delta\pi$ and maximum D_{xy} within each simulated SHDR and permutation, and only considered a SHDR to be an outlier for a given selection statistic if the observed maximum (Tajima's D, $\Delta\pi$) or minimum (D_{xy}) value was above the 90th or below the 10th percentile of simulated values. Number of outlier selection statistics per SHDR was tallied and compared across replicate or allopatric sharing in each species.

Global and local PCAs

To assess neutral genetic variation between individuals and populations, we performed principal component analysis (PCA) in the eastern and western transects separately for each species, i.e. two transects per PCA (Fig. 1 A). We first obtained a random subsample of 10% of windows that did not have high differentiation across populations, i.e. with

zPBS/zFst < 4, and then pruned for linked sites by only retaining 1 site for every 10kb, yielding 14995 and 8293 sites for *H. erato* and *H. melpomene*, respectively. We used the program ANGSD (v 0.933, Korneliussen et al., 2014) to obtain genotype likelihoods in beagle format (-doGlf 2) for all individuals. In *H. erato*, we excluded chromosome 2 as it contains a large inversion which could distort the neutral differentiation landscape. Genotype likelihoods were used as input for PCAngsd, which incorporates genotype uncertainty from genotype likelihoods to obtain a covariance matrix across all individuals¹¹⁰.

To assess whether the same haplotypes were involved in adaptation to altitude across replicate transects (within sides of the Andes), we performed local PCAs with outlier windows (zPBS/zFst > 4, i.e. >4 standard deviations from the mean) of each SHDR (total = 370 local PCAs). All individuals from replicate transects were included, leading to local PCAs for all Western and Eastern SHDRs of each species that included Colombia and Ecuador samples. We obtained genotype likelihoods in beagle format (-doGlf 2) as input for PCAngsd, similarly to the population structure analysis. We then assessed whether altitude was a significant predictor of individual clustering in each SHDR by building a linear model where local PCA PC1 was the response variable, and altitude and global (genome wide) PCA PC1 the predictors, to account for population structure. We considered that individuals in each replicate transect had the same haplotypes in SHDRs if altitude was a significant predictor of local PCA PC1, while controlling for global PC1 (Fig. 5). We additionally obtained the overall variation explained by the fitted linear models (R^2) for each SHDR local PCA and the relative contributions of each explanatory variable (altitude and global PCA PC1, partial R^2), estimated with the package relaimpo¹¹¹.

Measures of excess allele sharing

We used ABBA-BABA-related statistics to examine patterns of allele sharing between closely or distantly related high-altitude species and our study *H. erato* and *H. melpomene* populations. These statistics test for an excess of shared derived variation between lineages to distinguish gene flow or ancestral population structure from the incomplete lineage sorting (ILS) that can occur during a simple tree-like branching process. To examine genome-wide patterns of excess allele sharing between populations and species, we obtained F branch statistics implemented with the package *Dsuite*⁸³. Fbranch summarises and visualises patterns of excess allele sharing across phylogenetic datasets. We performed linkage-pruning to obtain a genome-wide average of excess allele sharing. Using a custom script (<https://github.com/joanam/scripts/blob/master/ldPruning.sh>) we removed sites above an LD-threshold of $R^2 > 0.1$ with plink v. 1.07¹¹². To reconstruct the backbone phylogeny for excess allele sharing tests, we extracted for each population or species the individual with highest

sequencing depth from the vcf file using vcftools v. 0.1.15¹⁰⁰. The vcf file was then converted to phylip with a custom script (<https://github.com/joanam/scripts/blob/master/vcf2phylip.py>). We reconstructed the phylogeny of the melpomene and erato clade separately with RAxML v. 8.2.9 using the GTRGAMMA model. Using this backbone tree, we used the LD-pruned vcf files of all melpomene/erato clade individuals to compute f statistics (tests of excess allele sharing) across all possible sets of three populations or species with Dsuite Dtrios. Next, we summarized these statistics with Dsuite Fbranch. The extent of gene flow in the eastern Andes between *H. melpomene* and *H. timareta* can skew genome-wide trees. Thus, to assess the levels of gene flow, we constrained the *H. melpomene* clade to be monophyletic following the species tree. In order to remove spurious signatures of excess allele sharing that are not significant, we set Fbranch values to 0 if the z-score was greater than 3 with a custom script (<https://github.com/joanam/scripts/blob/master/removeNonsignDsuite.r>). Lastly, we plotted the Fbranch statistics along the phylogeny with dttools.py of the Dsuite package.

To test for adaptive introgression from high-altitude specialist species into *H. erato* and *H. melpomene* highland populations, we computed f_{DM} , a statistic of excess allele sharing suitable for small genomic regions. This test is based on a set of four populations, including two sister taxa (P1 & P2), a close relative (P3) that may have admixed with one of these sister taxa and an outgroup (O). Here, the sister taxa represent the *H. erato* or *H. melpomene* lowland (P1) and highland populations (P2), whereas P3 represents a highland specialist species that may have contributed beneficial gene variants to the *H. erato* and *H. melpomene* highland populations. f_{DM} quantifies gene flow between P3 and P2 or between P3 and P1. In addition, we ran these tests with the allopatric erato/melpomene populations as P3, to test if at SHDRs, the same haplotypes are found in high-altitude populations on both sides of the Andes, potentially due to parallel selection on the same haplotypes. We estimated f_{DM} for non-overlapping 50 kb windows across the genome with the ABBABABAwindows.py script by (Martin et al., 2014) from https://github.com/simonhmartin/genomics_general. We considered individual SHDRs as f_{DM} outliers if their observed maximum f_{DM} value was >90th percentile of the absolute minimum f_{DM} values across all SHDRs. Additionally, we tested for overall enrichment of excess allele sharing between P3 and P2 (i.e. with the highlands) across all SHDRs, by testing with a Kolmogorov-Smirnov tests if the distribution of maximum f_{DM} values across all SHDRs (i.e. allele sharing with the highlands, P2) was significantly higher than the absolute minimum f_{DM} values across all SHDRs (i.e. allele sharing with the lowlands, P1). This was repeated for each Colombia/Ecuadorian clines with their respective potential donors (P3).

Acknowledgements

We are grateful to all the field assistants who have collected samples for this study, Narupa Reserve (Jocotoco Foundation, Ecuador), Jatun Satcha Reserve (Ecuador), and Universidad Regional Amazónica Ikiam for their support. We thank Steven van Belleghem for sharing recombination rates data for *H. erato*, and Emma Curran and Juan Enciso for providing some of the genomic sequences, and the Butterfly Genetics Lab (Cambridge) for helpful feedback.

G.M.-K. was supported by a Natural Environment Research Council Doctoral Training Partnership (NE/L002507/1). This work, N.J.N. and C.D.J. were supported by the Natural Environment Research Council (grant number: NE/R010331/1) and by a European Research Council Grant (339873) to C.D.J.. Some of the sequence data was generated under a NERC fellowship (NE/K008498/1) to NJN. Funding was provided to C.N.B. by the Spanish Agency for International Development Cooperation (AECID, grant number 2018SPE0000400194). C.S and NR were funded by Fondos Concursables Big - grant IV-FGD005/ IV-FGI006 Universidad del Rosario. SHM was supported by a NERC IRF (NE/N014936/1). Y.F.C. was supported by the European Research Council Starting Grant 639096 “HybridMiX” and the Max Planck Society. Open access funding provided by University of Cambridge. Deposited in PMC for immediate release.

In Colombia, field collections were conducted under permit no. 530 issued by the Autoridad Nacional de 539 Licencias Ambientales of Colombia (ANLA). In Ecuador, collections during November-December 2011, and September-October 2012, were done under permit 0033-FAU-MAE-DPO-PNY and exported under permits 001-FAU-MAE-DPO-PNY and 006-EXP-CIEN-FAU-DPO-PNY. Permits were obtained from Parque Nacional Yasuní, Ministerio Del Ambiente, La Dirección Provincial de Orellana. Collections in Ecuador during 2017-2019 were conducted under the permit provided by the Ministerio del Ambiente, Ecuador (MAE-DNB-CM-2017-0058).

Data availability

Sequence data is deposited at the NCBI Short Read Archive (primary accession: PRJEB36288) or elsewhere on ENA, as specified in Supplementary Table 1. Scripts have been made available in the public repository Zenodo (TBC). All records associated to the individuals used for this study are available in the *Heliconius* Earthcape database (<https://heliconius.ecdb.io/>, Jiggins et al., 2019).

References

1. Booker, T. R., Yeaman, S. & Whitlock, M. C. Global adaptation complicates the interpretation of genome scans for local adaptation. *Evol. Lett.* (2020) doi:<https://doi.org/10.1002/evl3.208>.
2. Stern, D. L. The genetic causes of convergent evolution. *Nat. Rev. Genet.* **14**, 751–764 (2013).
3. Bohutínská, M. *et al.* Genomic basis of parallel adaptation varies with divergence in *Arabidopsis* and its relatives. *Proc. Natl. Acad. Sci.* **118**, (2021).
4. Conte, G. L., Arnegard, M. E., Peichel, C. L. & Schluter, D. The probability of genetic parallelism and convergence in natural populations. *Proc. R. Soc. B Biol. Sci.* **279**, 5039–5047 (2012).
5. Martin, A. & Orgogozo, V. The Loci of Repeated Evolution: A Catalog of Genetic Hotspots of Phenotypic Variation. *Evolution* **67**, 1235–1250 (2013).
6. Besnard, G. *et al.* Phylogenomics of C4 photosynthesis in sedges (Cyperaceae): multiple appearances and genetic convergence. *Mol. Biol. Evol.* **26**, 1909–1919 (2009).
7. Louis, M. *et al.* Selection on ancestral genetic variation fuels repeated ecotype formation in bottlenose dolphins. *Sci. Adv.* **7**, eabg1245 (2021).
8. Wang, L. *et al.* Molecular parallelism underlies convergent highland adaptation of maize landraces. *Mol. Biol. Evol.* **38**, 3567–3580 (2021).
9. Calfee, E., Agra, M. N., Palacio, M. A., Ramírez, S. R. & Coop, G. Selection and hybridization shaped the rapid spread of African honey bee ancestry in the Americas. *PLOS Genet.* **16**, e1009038 (2020).
10. Jones, M. R. *et al.* Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science* **360**, 1355–1358 (2018).
11. Meier, J. I. *et al.* Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* **8**, 14363 (2017).
12. Zhang, X. *et al.* The history and evolution of the Denisovan-EPAS1 haplotype in Tibetans. *Proc. Natl. Acad. Sci.* **118**, (2021).
13. Jones, F. C. *et al.* The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61 (2012).
14. Wooldridge, T. *et al.* A novel enhancer of Agouti contributes to parallel evolution of cryptically colored beach mice. (2021).
15. Oziolor, E. M. *et al.* Adaptive introgression enables evolutionary rescue from extreme environmental pollution. *Science* **364**, 455–457 (2019).
16. Valencia-Montoya, W. A. *et al.* Adaptive introgression across semipermeable species boundaries between local *Helicoverpa zea* and invasive *Helicoverpa armigera* moths. *Mol. Biol. Evol.* **37**, 2568–2583 (2020).
17. Nogués-Bravo, D. *et al.* Cracking the code of biodiversity responses to past climate change. *Trends Ecol. Evol.* **33**, 765–776 (2018).
18. Mayhew, P. J. Why are there so many insect species? Perspectives from fossils and phylogenies. *Biol. Rev.* **82**, 425–454 (2007).
19. Ding, D. *et al.* Genetic variation in PTPN1 contributes to metabolic adaptation to high-altitude hypoxia in Tibetan migratory locusts. *Nat. Commun.* **9**, 4991 (2018).
20. Zhang, Q.-L. *et al.* Comparative transcriptomic analysis of Tibetan *Gynaephora* to explore the genetic basis of insect adaptation to divergent altitude environments. *Sci. Rep.* **7**, 16972 (2017).

21. Yeaman, S., Gerstein, A. C., Hodgins, K. A. & Whitlock, M. C. Quantifying how constraints limit the diversity of viable routes to adaptation. *PLOS Genet.* **14**, e1007717 (2018).
22. Gross, J. B., Borowsky, R. & Tabin, C. J. A Novel Role for Mc1r in the Parallel Evolution of Depigmentation in Independent Populations of the Cavefish *Astyanax mexicanus*. *PLoS Genet.* **5**, e1000326 (2009).
23. Harris, R. B. *et al.* The population genetics of crypsis in vertebrates: recent insights from mice, hares, and lizards. *Heredity* **124**, 1–14 (2020).
24. Mundy, N. I. A window on the genetics of evolution: MC1R and plumage colouration in birds. *Proc. R. Soc. Lond. B Biol. Sci.* **272**, 1633–1640 (2005).
25. Orteu, A. & Jiggins, C. D. The genomics of coloration provides insights into adaptive evolution. *Nat. Rev. Genet.* **21**, 461–475 (2020).
26. Castro, J. P. *et al.* An integrative genomic analysis of the Longshanks selection experiment for longer limbs in mice. *eLife* **8**, e42014 (2019).
27. Orr, H. A. The Probability of Parallel Evolution. *Evolution* **59**, 216–220 (2005).
28. Barghi, N., Hermisson, J. & Schlötterer, C. Polygenic adaptation: a unifying framework to understand positive selection. *Nat. Rev. Genet.* **21**, 769–781 (2020).
29. Hancock, A. M., Alkorta-Aranburu, G., Witonsky, D. B. & Di Rienzo, A. Adaptations to new environments in humans: the role of subtle allele frequency shifts. *Philos. Trans. R. Soc. B Biol. Sci.* **365**, 2459–2468 (2010).
30. Capblancq, T., Fitzpatrick, M. C., Bay, R. A., Exposito-Alonso, M. & Keller, S. R. Genomic prediction of (mal) adaptation across current and future climatic landscapes. *Annu. Rev. Ecol. Evol. Syst.* **51**, 245–269 (2020).
31. Rosser, N., Phillimore, A. B., Huertas, B., Willmott, K. R. & Mallet, J. Testing historical explanations for gradients in species richness in heliconiine butterflies of tropical America. *Biol. J. Linn. Soc.* **105**, 479–497 (2012).
32. Rueda-M, N., Salgado-Roa, F. C., Gantiva-Q, C., Pardo-Diaz, C. & Salazar, C. Environmental drivers of diversification and hybridization in Neotropical butterflies. *Front. Ecol. Evol.* 746.
33. Montejo-Kovacevich, G. *et al.* Altitude and life-history shape the evolution of *Heliconius* wings. *Evolution* **73**, 2436–2450 (2019).
34. Montejo-Kovacevich, G. *et al.* Genomics of altitude-associated wing shape in two tropical butterflies. *Mol. Ecol.* **0**, (2021).
35. Montejo-Kovacevich, G. *et al.* Microclimate buffering and thermal tolerance across elevations in a tropical butterfly. *J. Exp. Biol.* **223**, (2020).
36. Moest, M. *et al.* Selective sweeps on novel and introgressed variation shape mimicry loci in a butterfly adaptive radiation. *PLOS Biol.* **18**, e3000597 (2020).
37. Van Belleghem, S. M. *et al.* Selection and isolation define a heterogeneous divergence landscape between hybridizing *Heliconius* butterflies. *Evolution* (2021).
38. Kozak, K. M. *et al.* Multilocus species trees show the recent adaptive radiation of the mimetic *heliconius* butterflies. *Syst. Biol.* **64**, 505–524 (2015).
39. Bradburd, G. S., Ralph, P. L. & Coop, G. M. Disentangling the effects of geographic and ecological isolation on genetic differentiation. *Evolution* **67**, 3258–3273 (2013).
40. Wang, I. J. & Bradburd, G. S. Isolation by environment. *Mol. Ecol.* **23**, 5649–5662 (2014).

41. González-Martínez, S. C., Ridout, K. & Pannell, J. R. Range Expansion Compromises Adaptive Evolution in an Outcrossing Plant. *Curr. Biol.* **27**, 2544–2551.e4 (2017).
42. Härmälä, T. & Savolainen, O. Genomic Patterns of Local Adaptation under Gene Flow in *Arabidopsis lyrata*. *Mol. Biol. Evol.* **36**, 2557–2571 (2019).
43. Jasper, R. J. & Yeaman, S. *Local adaptation can cause both peaks and troughs in nucleotide diversity within populations.*
<http://biorxiv.org/lookup/doi/10.1101/2020.06.03.132662> (2020)
doi:10.1101/2020.06.03.132662.
44. Yi, X. *et al.* Sequencing of 50 Human Exomes Reveals Adaptation to High Altitude. *Science* **329**, 75–78 (2010).
45. Amorim, C. E. G. *et al.* Genetic signature of natural selection in first Americans. *Proc. Natl. Acad. Sci.* (2017) doi:10.1073/pnas.1620541114.
46. Delmore, K. *et al.* The evolutionary history and genomics of European blackcap migration. *eLife* **9**, e54462 (2020).
47. Lindo, J. *et al.* A time transect of exomes from a Native American population before and after European contact. *Nat. Commun.* **7**, 13175 (2016).
48. Vijay, N. *et al.* Genomewide patterns of variation in genetic diversity are shared among populations, species and higher-order taxa. *Mol. Ecol.* **26**, 4284–4295 (2017).
49. Wallberg, A., Pirk, C. W., Allsopp, M. H. & Webster, M. T. Identification of Multiple Loci Associated with Social Parasitism in Honeybees. *PLOS Genet.* **12**, e1006097 (2016).
50. Librado, P. & Orlando, L. Detecting Signatures of Positive Selection along Defined Branches of a Population Tree Using LSD. *Mol. Biol. Evol.* **35**, 1520–1535 (2018).
51. Nadeau, N. J. Genes controlling mimetic colour pattern variation in butterflies. *Curr. Opin. Insect Sci.* **17**, 24–31 (2016).
52. Ralph, P. L. & Coop, G. Convergent evolution during local adaptation to patchy landscapes. *PLoS Genet.* **11**, e1005630 (2015).
53. Fang, B., Kempainen, P., Momigliano, P., Feng, X. & Merilä, J. On the causes of geographically heterogeneous parallel evolution in sticklebacks. *Nat. Ecol. Evol.* **4**, 1105–1115 (2020).
54. Nadeau, N. J. *et al.* Population genomics of parallel hybrid zones in the mimetic butterflies, *H. melpomene* and *H. erato*. *Genome Res.* **24**, 1316–1333 (2014).
55. Van Belleghem, S. M. *et al.* Complex modular architecture around a simple toolkit of wing pattern genes. *Nat. Ecol. Evol.* **1**, (2017).
56. Brien, M. N. *et al.* *The genetic basis of structural colour variation in mimetic Heliconius butterflies.* 2021.04.21.440746
<https://www.biorxiv.org/content/10.1101/2021.04.21.440746v1> (2021)
doi:10.1101/2021.04.21.440746.
57. Barghi, N. *et al.* Genetic redundancy fuels polygenic adaptation in *Drosophila*. *PLOS Biol.* **17**, e3000128 (2019).
58. Charlesworth, B. Measures of divergence between populations and the effect of forces that reduce variability. *Mol. Biol. Evol.* **15**, 538–543 (1998).
59. Nachman, M. W. & Payseur, B. A. Recombination rate variation and speciation: theoretical predictions and empirical results from rabbits and mice. *Philos. Trans. R. Soc. B Biol. Sci.* **367**, 409–421 (2012).

60. Cruickshank, T. E. & Hahn, M. W. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* **23**, 3133–3157 (2014).
61. Charlesworth, B., Nordborg, M. & Charlesworth, D. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. *Genet. Res.* **70**, 155–174 (1997).
62. Martin, S. H. & Jiggins, C. D. Interpreting the genomic landscape of introgression. *Curr. Opin. Genet. Dev.* **47**, 69–74 (2017).
63. Burri, R. Interpreting differentiation landscapes in the light of long-term linked selection. *Evol. Lett.* **1**, 118–131 (2017).
64. Matthey-Doret, R. & Whitlock, M. C. Background selection and FST: consequences for detecting local adaptation. *Mol. Ecol.* **28**, 3902–3914 (2019).
65. Meier, J. I., Marques, D. A., Wagner, C. E., Excoffier, L. & Seehausen, O. Genomics of parallel ecological speciation in Lake Victoria cichlids. *Mol. Biol. Evol.* **35**, 1489–1506 (2018).
66. Yeaman, S. & Whitlock, M. C. The genetic architecture of adaptation under migration–selection balance. *Evol. Int. J. Org. Evol.* **65**, 1897–1911 (2011).
67. Martin, S. H., Davey, J. W., Salazar, C. & Jiggins, C. D. Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLoS Biol.* **17**, e2006288 (2019).
68. Wise, A. *et al.* Drosophila mutants of the autism candidate gene neurobeachin (rugose) exhibit neuro-developmental disorders, aberrant synaptic properties, altered locomotion, and impaired adult social behavior and activity patterns. *J. Neurogenet.* **29**, 135–143 (2015).
69. Davison, A., McMillan, W. O., Griffin, A. S., Jiggins, C. D. & Mallet, J. L. B. Behavioral and physiological differences between two parapatric *Heliconius* species. *Biotropica* (1999) doi:10.1111/j.1744-7429.1999.tb00415.x.
70. Zhang, Y. *et al.* A widely diverged locus involved in locomotor adaptation in *Heliconius* butterflies. *Sci. Adv.* **7**, eabh2340 (2021).
71. Konečná, V. *et al.* Parallel adaptation in autopolyploid *Arabidopsis arenosa* is dominated by repeated recruitment of shared alleles. *Nat. Commun.* **12**, 4979 (2021).
72. Mérot, C. *et al.* Locally adaptive inversions modulate genetic variation at different geographic scales in a seaweed fly. *Mol. Biol. Evol.*
73. Todesco, M. *et al.* Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature* **584**, 602–607 (2020).
74. Li, H. & Ralph, P. Local PCA Shows How the Effect of Population Structure Differs Along the Genome. *Genetics* **211**, 289–304 (2019).
75. Mérot, C., Oomen, R. A., Tigano, A. & Wellenreuther, M. A Roadmap for Understanding the Evolutionary Significance of Structural Genomic Variation. *Trends Ecol. Evol.* **35**, 561–572 (2020).
76. Meier, J. I. *et al.* Haplotype tagging reveals parallel formation of hybrid races in two butterfly species. *Proc. Natl. Acad. Sci.* **118**, (2021).
77. Kirkpatrick, M. & Barton, N. Chromosome inversions, local adaptation and speciation. *Genetics* **173**, 419–434 (2006).
78. Malinsky, M. *et al.* Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science* **350**, 1493–1498 (2015).
79. Martin, S. H., Davey, J. W. & Jiggins, C. D. Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci. *Mol. Biol. Evol.* **32**, 244–257 (2015).

80. Bull, V. *et al.* Polyphyly and gene flow between non-sibling *Heliconius* species. *BMC Biol.* **4**, 1–17 (2006).
81. Martin, S. *et al.* Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res.* (2013) doi:10.1101/gr.159426.113.
82. Pardo-Diaz, C. *et al.* Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genet.* **8**, e1002752 (2012).
83. Malinsky, M., Matschiner, M. & Svardal, H. Dsuite-Fast D-statistics and related admixture evidence from VCF files. *Mol. Ecol. Resour.* **21**, 584–595 (2021).
84. Kozak, K. M., Joron, M., McMillan, W. O. & Jiggins, C. D. Rampant genome-wide admixture across the *Heliconius* radiation. *Genome Biol. Evol.* (2021).
85. Dasmahapatra, K. K. *et al.* Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487**, 94 (2012).
86. Jiggins, C. D., Mcmillan, O., Neukirchen, W., Mallet, J. & Nw, L. What can hybrid zones tell us about speciation? The case of *Heliconius erato* and *H. himera* (Lepidoptera : Nymphalidae). *Biol. J. Linn. Soc.* 221–242 (1996) doi:10.1111/j.1095-8312.1996.tb01464.x.
87. Merrill, R. M., Chia, A. & Nadeau, N. J. Divergent warning patterns contribute to assortative mating between incipient *Heliconius* species. *Ecol. Evol.* **4**, 911–917 (2014).
88. Montgomery, S. H. & Merrill, R. M. Divergence in brain composition during the early stages of ecological specialization in *Heliconius* butterflies. *J. Evol. Biol.* **30**, 571–582 (2017).
89. Jay, P. *et al.* Supergene evolution triggered by the introgression of a chromosomal inversion. *Curr. Biol.* **28**, 1839–1845 (2018).
90. Jiggins, C. D. *The Ecology and Evolution of Heliconius Butterflies*. (Oxford University Press, 2016).
91. Hoorn, C. *et al.* Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *science* **330**, 927–931 (2010).
92. Picelli, S. *et al.* Tn5 transposase and tagmentation procedures for massively scaled sequencing projects. *Genome Res.* **24**, 2033–2040 (2014).
93. Ginestet, C. ggplot2: Elegant Graphics for Data Analysis. *J. R. Stat. Soc. Ser. A Stat. Soc.* (2011) doi:10.1111/j.1467-985X.2010.00676_9.x.
94. Challis, R. J., Kumar, S., Dasmahapatra, K. K., Jiggins, C. D. & Blaxter, M. *Lepbase: the Lepidopteran genome database*. <http://biorxiv.org/lookup/doi/10.1101/056994> (2016) doi:10.1101/056994.
95. Davey, J. W. *et al.* Major Improvements to the *Heliconius melpomene* Genome Assembly Used to Confirm 10 Chromosome Fusion Events in 6 Million Years of Butterfly Evolution. *G3 GenesGenomesGenetics* g3.115.023655 (2016) doi:10.1534/g3.115.023655.
96. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv Prepr. ArXiv13033997* (2013).
97. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinforma. Oxf. Engl.* **25**, 2078–2079 (2009).
98. Broad Institute. Picard tools. *Broad Inst. GitHub Repos.* (2018).
99. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
100. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).

101. Wang, I. J. Topographic path analysis for modelling dispersal and functional connectivity: Calculating topographic distances using the topoDistance R package. *Methods Ecol. Evol.* **11**, 265–272 (2020).
102. Jiggins, C. D., Salazar, P. A. & Montejo-Kovacevich, G. Heliconiine Butterfly Collection Records from University of Cambridge. Department of Zoology, Cambridge. (2019).
103. Korneliussen, T. S., Albrechtsen, A. & Nielsen, R. ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics* **15**, 356 (2014).
104. Hudson, R. R., Slatkin, M. & Maddison, W. P. Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**, 583–589 (1992).
105. Bhatia, G., Patterson, N., Sankararaman, S. & Price, A. L. Estimating and interpreting FST: the impact of rare variants. *Genome Res.* **23**, 1514–1521 (2013).
106. Bourgon, R. *intervals: Tools for working with points and intervals*. <https://CRAN.R-project.org/package=intervals> (2020).
107. Korunes, K. L. & Samuk, K. pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *bioRxiv* 2020.06.27.175091 (2020) doi:10.1101/2020.06.27.175091.
108. Nei, M. The theory of genetic distance and evolution of human races. *Jpn. J. Hum. Genet.* **23**, 341–369 (1978).
109. Belleghem, S. M. V. *et al.* Selection and gene flow define polygenic barriers between incipient butterfly species. *bioRxiv* 2020.04.09.034470 (2020) doi:10.1101/2020.04.09.034470.
110. Meisner, J. & Albrechtsen, A. Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data. *Genetics* **210**, 719–731 (2018).
111. Grömping, U. Relative importance for linear regression in R: The package **relaimpo**. *J. Stat. Softw.* (2006) doi:10.18637/jss.v017.i01.
112. Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, s13742-015 (2015).
113. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
114. Malinsky, M. *et al.* Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nat. Ecol. Evol.* **2**, 1940–1955 (2018).