The evolution and genetics of sexually dimorphic "dual" mimicry in the butterfly *Elymnias hypermnestra*

Dee M. Ruttenberg[1,2], Nicholas W. VanKuren[1], Sumitha Nallu[1], Shen-Horn Yen[3], Djunijanti Peggie[4], David J. Lohman[5,*], and Marcus R. Kronforst[1,*]

Affiliations:

[1] Department of Ecology & Evolution, The University of Chicago, Chicago IL 60637, USA

[2] Current Address: Department of Quantitative and Computational Biology; Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, 08544, USA

[3] Department of Biological Sciences, National Sun Yat-Sen University, Kaohsiung, 80424, Taiwan

[4] Museum Zoologicum Bogoriense, Research Center for Biology, Indonesian Institute of Sciences (LIPI), Cibinong-Bogor 16911, Indonesia

[5] Biology Department, City College of New York, City University of New York, New York, NY, 10031, USA; PhD Program in Biology, Graduate Center, City University of New York, New York, NY, 10016, USA; and Entomology Section, National Museum of Natural History, Manila, Philippines, 1000

*Corresponding authors

DJL: dlohman@ccny.cuny.edu

MRK: mkronforst@uchicago.edu

KEYWORDS: Batesian mimicry, color pattern, evolution, gene reuse, genomics, Satyrinae

ABSTRACT

Sexual dimorphism is a major component of morphological variation across the tree of life, but the mechanisms underlying phenotypic differences between sexes of a single species are poorly understood. We examined the population genomics and biogeography of the common palmfly *Elymnias hypermnestra*, a dual mimic in which female wing color patterns are either dark brown (melanic) or bright orange, mimicking toxic *Euploea* and *Danaus* species, respectively. As males always have a melanic wing color pattern, this makes *E. hypermnestra* a fascinating model organism in which populations vary in sexual dimorphism. Population structure analysis revealed that there were three genetically distinct *E. hypermnestra* populations, which we further validated by creating a phylogenomic species tree and inferring historical barriers to gene flow. This species tree demonstrated that multiple lineages with orange females do not form a monophyletic group, and the same is true of clades with melanic females. We identified two SNPs near the color patterning gene *WntA* that were significantly associated with the female color pattern polymorphism, suggesting that this gene affects sexual dimorphism. Given *WntA*'s role in color patterning across Nymphalidae, *Elymnias hypermnestra* females demonstrate the repeatability of the evolution of sexual dimorphism.

INTRODUCTION

Understanding the relationship between genetic variability and the many levels of biological diversity is a central aim of genomics. Single genes of large effect are often found to be responsible for striking examples of adaptive variation [1, 2]. Thus, much morphological diversity is derived from genetic variation at a relatively small number of genetic loci [3-6]. Mimetic butterflies are models for studying the relationship between exceptional phenotypic diversity resulting from limited genetic diversity for a number of reasons, including the manifest adaptive value of mimetic phenotypes, the fecundity and ease of rearing butterflies, and the incredible morphological diversity of butterflies [7, 8]. Unraveling the genomic and developmental basis of butterfly phenotypes has advanced understanding of the evolution of sexual dimorphism [9], mimicry [5], and evolvability [10].

The Batesian mimetic butterfly genus *Elymnias* (Lepidoptera: Nymphalidae: Satyrinae) lends itself to the study of mimicry and sexual dimorphism because its 53 recognized species can vary dramatically in the color, pattern, and wing size to mimic a variety of different model species in the families Nymphalidae, Pieridae, Papilionidae, Erebidae (Arctiinae), and Zygaenidae throughout tropical and subtropical Asia [11, 12]. Moreover, only dorsal or both dorsal and ventral wing surfaces may be mimetic, and individual species can mimic multiple models via morphological differences that vary between sexes, locales, or syntopic forms [13, 14]. Within the genus Elymnias there are several examples of allopatrically distributed species mimicking the same widespread model, thereby resembling each other, and of different populations or forms of a single species mimicking different models [13,14]. The most widespread and locally abundant species in this genus is the common palmfly, *E. hypermnestra* [11]. This species is a “dual mimic” [15]: it is sexually dimorphic and each sex resembles a dramatically different model species. All males of this palm-feeding species resemble melanic, unpalatable models in the genus *Euploea* [13, 16] (figure 2). However, female mimicry is

68 geographically variable: some disjunct populations are sexually dimorphic with orange females
69 that mimic *Danaus*, while other populations are monomorphic, and melanic females mimic
70 *Euploea* models along with the males (figures 1 and 2). Orange and melanic females do not co-
71 occur. Naïve, captive insectivorous birds (*Pycnonotus sinensis formosae*, *Zosterops japonicus*
72 *simplex*, and *Copsychus malabaricus*) with no prior exposure to the model or mimic readily
73 consume adult males, orange females, and melanic females representing four E. hypermnestra
74 subspecies, indicating the species is a palatable Batesian mimic (S.-H. Yen, unpublished data).
75 This species provides a unique opportunity to study the genomic basis of dual mimicry to assess
76 whether the trait is controlled by loci known to control sexual dimorphism [2, 17], mimicry [6,
77 18, 19], or both. In addition, the experimental advantages of this variable and widespread species
78 might allow identification of loci that play an important role in the tremendous morphological
79 diversity of its congeners.

80 Here, we examine the evolution and biogeography of sexually dimorphic dual mimicry in
81 E. *hypermnestra.* Orange females of *E. h. tinctoria* (Thailand) and *E. h. baliensis* (Bali) produce
82 orange patterns using different combinations of ommochrome pigments, suggesting independent
83 evolution of orange morphs in these two geographically distant populations [20]. However, the
84 evolutionary history and current population structure of *E. hypermnestra* were unknown, making
85 it impossible to distinguish between single- and multi-origin scenarios. Moreover, while
86 researchers have identified many genes that control the development of mimetic color patterning
87 in butterflies [5], including *doublesex*, responsible for female-limited polymorphic mimicry in
88 *Papilio polytes* [2, 21], the genes controlling sexually dimorphic dual mimicry are not
89 understood. Since *E. hypermnestra* is dimorphic in some regions and monomorphic in others,
90 this species has the potential to elucidate how sex-specific effects emerge and contribute to
91 phenotypic variation. We assembled a high-quality reference genome and then resequenced low-
92 coverage reads from 45 individuals representing 18 subspecies across the species' range. This

allowed us to address three questions: 1) What is the population history and current population genetic structure of *E. hypermnestra*? 2) Does the orange female color pattern have a single evolutionary origin? 3) What gene(s) are responsible for whether a population is dimorphic with orange females or monomorphic with melanic females?

RESULTS

Genetic structure of E. hypermnestra *populations*

We first assembled a reference genome for *E. hypermnestra baliensis* to facilitate downstream analyses. Using k-mer analysis [22] and SCO content evaluation [23,24], we found that the *E. hypermnestra* reference genome presented here is among the best assembled, most complete, and least redundant nymphalid genomes available (electronic supplementary material, table S1). To better understand natural variation in *Elymnias hypermnestra* across its large distribution spanning *ca.* 55 longitudinal degrees from western India to eastern Indonesia, we resequenced the genomes of 45 samples with at least ~20X coverage representing 18 subspecies across Asia (figure 1*a*, electronic supplementary material, table S2) . We called SNPs in our resequenced data relative to the reference genome. This genome-wide SNP data indicated substantial genetic structure. The samples formed three distinct clusters in a principal component analysis of these data (figure 1*b*). We calculated fixation (F_{ST}) indices between each pair of subspecies and found the same three populations (electronic supplementary material, figure S1). The same three groups were also identified by ADMIXTURE [25] (figure 1*c*). Increasing the number of putative populations increased the likelihood of the admixture model, but the results assuming 2-5 populations all had comparable cross-validation errors (electronic supplementary material, figure S2).

Repeated evolution of the Danaus *mimetic color patterns in* Elymnias hypermnestra

A 6-locus intraspecific phylogeny of *E. hypermnestra* suggested that neither orange nor melanic females were monophyletic, but support values on this tree were low (electronic supplementary material, appendix S1). We therefore inferred a species tree with ASTRAL using gene trees from 3,000 unlinked, autosomal 10 kb windows. This tree was also inconsistent with either orange or melanic female morphs forming a monophyletic group (figure 2), as there were 5 melanic and 4 orange lineages. Trees inferred from Z-linked windows or complete mtDNA genomes (electronic supplementary material, figure S3) were topologically similar to the species tree inferred from autosomal loci (figure 2). The *E. h. hainana* subspecies/genetic population was distinctive in the PCA (figure 1*b*) and in the species tree, where all five samples form a strongly supported branch (figure 2). However, samples of this subspecies were not monophyletic in the 6-locus tree (electronic supplementary material, appendix S1), underscoring the potential bias of inferring intraspecific phylogeny using only few protein-coding markers [26].

We developed a coalescent model using the phylogeny of *E. hypermnestra* to examine when a given number of gene flow events are likely to have occurred during the evolutionary history of *E. hypermnestra*. Our species tree suggested that most subspecies are monophyletic (figure 2), so, while we recognize that subspecies are not necessarily monophyletic groups [27], we treated each subspecies as a group for computational ease. We found residual covariance between taxa in our model that was best explained by gene flow (electronic supplementary material, figure S4*a*). Some gene flow events were between melanic clades in different regions, but others were between melanic and orange clades, suggesting that gene flow was not only between subspecies with the same female morphs, consistent with results from *Papilio polytes* [28].

Finally, we used EEMS [29] to visualize estimated relative migration rates across geographic space (electronic supplementary material, figure S4*b*). Our results were consistent with biogeographic patterns evident in the species tree (figure 2). EEMS predicted several strong barriers to gene flow. The strongest barrier coincides with Wallace's Line, a well-known

biogeographic demarcation that separates Bali (orange females) from Lombok (melanic females) and extends northward between Borneo and Sulawesi [30]. A second barrier separates Sumatra (melanic females) from Java (orange females), and the third barrier separates melanic *E. h. hainana* from all other populations. Intraspecific genetic diversity was highest on the Asian mainland and decreased from west to east along the Indo-Australian Archipelago (electronic supplementary material, figure S5).

A genome-wide association study of the orange Danaus*-like color pattern suggests reuse of WntA*

To identify the genetic locus or loci associated with orange and melanic female color patterns in *E. hypermnestra*, we performed genome wide association mapping of female color patterns using the full SNP call set from the 45 re-sequenced samples. If male butterflies were sequenced, their collection locality was used to infer the female color pattern from that area. We performed the GWAS using GEMMA [31] because it incorporates population structure and relatedness among samples. We saw no peaks in the unaligned genome without an equivalent in the aligned genome (electronic supplementary material, figure S5).

While many sites fell above the 1% false discovery rate (FDR), correcting for multiple testing, these sites had relatively little linkage disequilibrium. Importantly, the most strongly associated sites had no other neighbors (figure 3). This was consistent with our gene tree of the 200 bp region surrounding these SNPs in which neither orange nor melanic color patterns were monophyletic (electronic supplementary material, figure S3*c*).

The two most strongly associated sites were 3 base pairs apart; both exceed 1% FDR (figure 3*a*). Adding the population structure (as measured by the first principal component from the PCA) as a covariate removed neither of these sites (electronic supplementary material, figure S7). Looking at the genotype of the two sites on the scaffold (figure 3*b*, electronic supplementary material, figure S7), we observed that they predicted wing pattern almost perfectly (electronic

supplementary material, table S6). These sites were 150 kb away from *WntA*, a patterning gene that has repeatedly been shown to be involved in melanization across the family Nymphalidae [32].

DISCUSSION

Repeated evolution of a mimetic color pattern

Females of the dual mimic *E. hypermnestra* either resemble *Euploea* with a melanic color pattern similar to males, or have an orange color pattern mimicking *Danaus*. Our analyses shed light on the evolutionary and genetic mechanisms responsible for the geographic mosaic of female color pattern in this facultatively sexually dimorphic species.

Our analysis of population structure suggested the presence of three genetic populations in *E. hypermnestra*. The first group represented the described subspecies *Elymnias hypermnestra hainana* found in Taiwan, southern China including Hainan, northern Vietnam, and central Laos (figure 1). The second genetic population comprised *E. hypermnestra* found on Java, the Lesser Sunda Islands, and Seram. The third included individuals from South Asia including Sri Lanka, Indochina south of *hainana*, and Sumatra (figure 1). The geographic border between *E. h. hainana* and the rest of *E. hypermnestra*'s range coincides with the Hoang Lien Son Range and surrounding high elevation areas in the "Tail of the Himalayas." The other border between genetic populations lies between Java and Sumatra. While these are currently separate land masses, the two islands were conjoined during Pleistocene low sea stands together with Borneo and the Thai-Malay peninsula to form a single land mass along the edge of the continental shelf, Sundaland [30]. Thus, the border of these populations lacks an obvious barrier to dispersal, though this area is frequently associated with genetic discontinuities within and between other butterfly species [Lohman, unpublished data]. While all *E. h. hainana* females are melanic, the other two populations include areas with orange females and areas with melanic females, which could be explained by the convergent evolution of color patterns in disjunct locales.

We were able to trace the evolutionary history of the orange/melanic transition using phylogenetic analysis. As suggested by our species tree, the orange and melanic morphs of *Elymnias hypermnestra* did not form monophyletic groups. This is not uncommon in butterflies – for instance, a single morph of *Heliconius* may prevail in a given region, but actually comprise distinct *Heliconius* species that are only monophyletic at color pattern loci [33]. It is still unclear why variability between melanic and orange morphs of *E. hypermnestra* evolved and how it is maintained. The lack of monophyletic female color patterns in *E. hypermnestra* may result from a geographic mosaic of selection to mimic the most common unpalatable model in a region. While differences in *Danaus* and *Euploea* local abundance have not been demonstrated [Yen, unpublished data], they tend to live in different habitats [11]. Characterization of the host plants, predators, and butterfly communities where different female forms live may shed light on this issue, including assessment of model species abundance. Moreover, studying geographic variability in the chemical ecology of the mimicry ring may provide insight on the relationship between mimetic morphs and their models [34].

WntA *and the orange/melanic shift*

To identify genetic factors underlying the shift between orange and melanic color patterns in *E. hypermnestra,* we performed a genome-wide association study (GWAS) of female color pattern. In GWAS analyses of similar systems, there are usually large peaks of many linked sites [2]. This raises the question of why there is apparently little linkage disequilibrium (LD) in this system. One possibility is that LD is lost because of filtering. On average, we identified one polymorphic site every 100 bp. Another possible explanation is that, unlike most previous functional genomics studies on Lepidoptera, this study sampled butterflies across a wide geographic range with strong population structure. Most other work was done within a narrower geographic range. For instance, all butterflies sampled in Kunte et. al. [2] were from a single F3 generation. When we compared our GWAS (figure 4) to results of other studies with

222 geographically extensive sampling (such as those on *Arabidopsis*), we found similarly rapid
223 linkage decay resulting in narrow peaks [35, 36].

224 Many previous studies demonstrate that *WntA* is associated with color patterning in other
225 nymphalid butterflies. In *Heliconius*, *WntA* is related to a color pattern transitions among
226 different species, and is typically expressed in regions of the butterfly wing that are melanic in
227 mature adults [37]. Moreover, linkage mapping has shown that *WntA* is associated with a similar
228 transition in *Limenitis arthemis*; in this case, an ancient *cis*-regulatory element mediates a
229 transition from a mimetic white banded to a non-mimetic, unbanded form [38, 39]. These data on
230 these two SNPs in *E. hypermnestra* were consistent with them being *cis*-regulatory elements
231 regulating *WntA* 150 kb downstream. While this is an unusually long distance between a
232 regulatory element and its target, it is not unprecedented. Regulatory elements have even been
233 found megabases away from the genes they regulate [40], and *optix* enhancers have been shown
234 to be up to 220 kb away in *Heliconius* [41, 42]. We found pronounced similarities between
235 *WntA*'s known effects on wing patterning in butterflies and the phenotype observed in *Elymnias*
236 *hypermnestra*. For instance, Mazo-Vargas *et al.* [32] created CRISPR *WntA* knockouts for a
237 variety of nymphalids and found two conserved characteristics of *WntA*. First, *WntA* typically
238 acts on the Basalis (B), the Central Symmetry System (CSS), and the Marginal Band System
239 (MBS), three regions of butterfly wings which are conserved across nymphalids. Moreover,
240 *WntA* is typically expressed in melanic regions, likely because it is associated with upregulation
241 of melanin. Both traits were found in the orange/melanic switch in *Elymnias hypermnestra*
242 (figure 2), further suggesting that *WntA* is involved in this transition of female color pattern.

243 The potential involvement of *WntA* in *E. hypermnestra* mimicry polymorphism suggests
244 that the gene functions somewhat differently than in *Heliconius* or *Limenitis*. Mimicry in *E.*
245 *hypermnestra* is sexually dimorphic: while females may be orange, males are always melanic
246 [13]. This implies that polymorphism affects females differently than males. Several mechanisms
247 are plausible: by upregulating *WntA* in melanic females; downregulating *WntA* in orange

females; or changing the spatial pattern of *WntA* expression. This is an unusual example of a single gene involved in both sexually dimorphic and non-sexually dimorphic mimicry. This suggests a slightly different role for *WntA* in this system than in others, where *WntA* affects both sexes. Future functional genomics work can elucidate the specific nature of *WntA* on this variation. Two other peaks in our GWAS stood out, one on chromosome 20 and one on chromosome 6. Many of the genes have unknown functions, suggesting an angle for further research (electronic supplementary material, table S6).

Predictability of evolution

Studies on wing patterns in Nymphalidae have revealed that a common toolkit of genes, including *optix*, *cortex*, and *WntA*, underlie wing patterning and support the hypothesis that evolutionary outcomes can be predictable [2, 10, 37, 38]. This study complements work on the predictability of evolution in two critical ways. For one, *Elymnias* diverged from the clade with *Limenitis* and *Heliconius* over 80 million years ago [43], making this one of the oldest cases of gene re-use in Nymphalidae that has been studied. Moreover, this demonstrates how sexual dimorphism can create variation with a single component of the toolkit: the same gene, *WntA*, seems to underlie sexually monomorphic variation and sexually dimorphic variation. This variation, in turn, allows for a greater phenotypic diversity than single genes of large effect would establish alone. The seemingly adaptive variability between sexes and among populations of *Elymnias hypermnestra* has provided a fascinating natural experiment to study the genomic basis and evolution of a novel sexually dimorphic trait.

METHODS

Reference genome assembly and quality

The *E. hypermnestra* reference genome was generated from two *E. hypermnestra baliensis* females from Bali. We isolated DNA from thorax tissue using a phenol–chloroform extraction method and constructed Illumina paired-end (PE) libraries with insert sizes 250 and

500 bp using the KAPA Hyper Prep Kit (KR0961 – v1.14) from 2 μg genomic DNA [44]. We constructed mate pair (MP) libraries with insert sizes of 2 kb, 6 kb, and 15 kb using the Nextera Mate Pair Library Prep kit (FC-132-1001) and 4 μg genomic DNA (electronic supplementary material, table S3). The five, unique barcoded libraries were pooled in a ratio of 59:30:6:3:2 and sequenced 2x100 bp on a single lane of Illumina HiSeq 4000 (electronic supplementary material, table S2). We trimmed low-quality regions and adapters from raw PE reads using Trimmomatic v0.36 [45] where bases in the reads that were below a quality score of 15 were trimmed using a sliding window of 4 bp and all reads less than 36 bp in length were discarded. We used Platanus v1.2.4 [44] to trim adapter sequences and low quality regions from mate pair reads. Trimmed libraries were assembled using the default settings of Platanus v1.2.4 and the assembly was polished using Redundans v0.13a (default settings; 46). We removed scaffolds <5 kb from this assembly, generated a species-specific repeat library, and masked repeats using RepeatScout 1.0.5 and RepeatMasker 4.0.8 [47, 48], respectively, to produce the final assembly. We estimated genome size and heterozygosity using 21-mer frequencies in the raw 250 bp PE library using GenomeScope [22].

We assessed the quality of our assemblies and other well-assembled nymphalid genomes using BUSCO v3 and the endopterygota gene set (2,440 single-copy orthologs) from OrthoDB v9 [23, 24]. The accessions of the assemblies tested are in the supplementary table. We assigned *E. hypermnestra* scaffolds to *Melitaea cinxia* chromosomes using RaGOO [49, 50]. This pipeline assigned 206/947 scaffolds (542 Mb/566 Mb) to chromosomes.

Finally, we generated a preliminary gene annotation set for the *E. hypermnestra* genome using MAKER v3.01.02 [51, 52]. We used *de novo* transcripts from *Bicyclus anynana* (NCBI BioProject) as evidence for transcription, as no transcriptome data exist for *Elymnias*. We downloaded raw reads from BioProject PRJEB10924 using the SRA toolkit, trimmed remaining adapters using Trimmomatic, and assembled transcripts using Trinity v2.8.0 [53] with default settings. Furthermore, we used protein sequences from the UniProt/SwissProt protein database

[54], and RefSeq protein models for *Danaus plexippus, Papilio xuthus, Bombyx mori, Vanessa tameamea, Pieris rapae,* and *Drosophila melanogaster* as evidence for protein-coding regions. We trained SNAP using this evidence, then used SNAP, Augustus v3.2 with *Heliconius melpomene* parameters, and GeneMark-ES 4 with MAKER to generate the final gene models (55). We functionally annotated predicted proteins using BLASTp against the Uniprot/SwissProt database and combined that information using scripts included in MAKER.

Whole genome resequencing and quality control

Adult *E. hypermnestra* were collected in the wild and preserved in ethanol and/or by freezing at -80˚ C (Table S1) before genomic DNA was extracted from thorax tissue using a phenol–chloroform DNA extraction protocol. We constructed ~250 bp paired-end libraries using the KAPA Hyper Prep Kit (KAPA Biosystems) and sequenced them to ~20X coverage using 2 x 80 bp Illumina NextSeq 500 (Table S1). We trimmed adapters and low-quality regions from raw resequencing reads using TrimGalore 0.6.1 and cutadapt v1.18 [56], then removed reads containing overrepresented sequences (identified using FastQC). We mapped reads to the *E. hypermnestra* reference genome using Bowtie2 v2.3.0-beta7 with parameter "--very-sensitive-local" [57]. We marked duplicate reads using PicardTools v2.8.1 and realigned around indels using the Genome Analysis ToolKit's (GATK, v3.8) RealignerTargetCreator and IndelRealigner. Finally, we called SNPs using the GATK UnifiedGenotyper with default settings except for the following values: heterozygosity prior = 0.02; minimum allowable base quality score = 30; and minimum mapping quality = 20 [58]. We removed genotypes with phred-scaled quality < 10. We then produced FASTA formatted genome sequences for each individual using the GATK FastaAlternateReferenceMaker [59]. Our data reached ~20X coverage on average and had average mapping rates of 94.76% (electronic supplementary material, table S3).

Population structure analyses

We inferred *E. hypermnestra* population structure with ADMIXTURE 1.3.0 [25]. We first performed linkage-disequilibrium-based pruning on our SNP dataset using plink v1.90, including only SNPs with $r^2 < 0.10$ in 50-bp sliding windows with 10-bp steps according to plink's --indep-pairwise utility. This yielded 108,189 SNPs. We ran ADMIXTURE with 10-fold cross-validation for parameters $k = 2$ through 10. We looked at the cross-validation error and the value of k that minimized the residuals (electronic supplementary material, figure S2, 60). We performed principal component analysis on the same filtered data set using plink [61].

Phylogenetic analyses

Since linkage disequilibrium returns to background levels over ~50 kb in *Heliconius* [62], we split the *E. hypermnestra* genome into non-overlapping 10 kb windows, kept every fifth window, then extracted alignments of sequences for each window from individual fastas with GATK. We tested for recombination within each alignment using PhiPack [63], then filtered out windows with recombination p values > 1e-10 and at least 100 informative sites. PhiPack uses patterns of polymorphism to infer the probability of past recombination events; as p-values decrease, the probability of recombination in the tested window increases. We randomly selected 3,000 autosomal alignments that passed these filters and inferred an unpartitioned gene tree from each using IQ-TREE, which selected the best model with ModelFinder and estimated branch support using 1000 ultrafast bootstraps [64-66]. Finally, we inferred a species tree using the default settings of ASTRAL-III [67], which computeed a consensus topology with support values derived from the fraction of gene trees that support a particular four-taxon topology (quartet scores).

Genome-wide association for color

We filtered out SNPs with 10 or more missing alleles or minor allele frequency < 0.10 from the unpruned data set for a total of 5.4 million SNPs. We assigned phenotypes to each sample based on that population's female wing color pattern (electronic supplementary material, table S2), then computed site-wise Wald $\chi 2$ test *p*-values using GEMMA v0.98, including GEMMA's centered kinship matrix as a covariate [31]. We calculated genome-wide cutoff scores using the false discovery rate method [68]. GWA results were plotted by ordering *E. hypermnestra* scaffolds to the *Melitaea cinxia* chromosome-level assembly [49].

Acknowledgements

Specimen collection in Thailand was authorized by permits from the National Research Council of Thailand and the Department of National Parks, Wildlife and Plant Conservation; fieldwork in Indonesia was conducted under an MoU between CCNY and RCB—LIPI with permits from RISTEK and other pertinent authorities; specimen collection in Vietnam was conducted under an MoU between CCNY and Cat Tien National Park. Additional specimens from the Museum of Comparative Zoology were sequenced for this study. Fieldwork was funded by grants 9285-13 and WW-227R-17 from the Committee for Exploration and Research of the National Geographic Society to DJL. This work was funded by NSF grants DEB-1120380 and DEB-1541557 to DJL, MOST grant 108-2621-B-110-004-MY3 to SHY, and NIH grant GM131828 to MRK.

Author's Contributions

D.J.L. and M.R.K. conceived and designed the study; D.M.R., N.W.V., and S.N. performed analyses and collected data; S.H.Y., D.P., D.J.L. collected specimens; N.W.V., D.J.L. and M.R.K. directed the project; D.M.R., N.W.V., and D.J.L. wrote the manuscript with input from all co-authors.

Data Accessibility Statement

The reference genome and sequence data generated for this study are publicly available at NCBI under BioProject accessions PRJNA660054 and PRJNA660057.

Competing Interests

The authors declare they have no competing interests.

REFERENCES

1. Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP. 2006 A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* **313**, 101-104. (doi:10.1126/science.1126121)
2. Kunte K, Zhang W, Tenger-Trolander A, Palmer DH, Martin A, Reed RD, Mullen SP, Kronforst MR. 2014 *doublesex* is a mimicry supergene. *Nature* **507**, 229-232. (doi:10.1038/nature13112)
3. Gilbert LE. 2004 Adaptive novelty through introgression in *Heliconius* wing patterns: evidence for a shared genetic "tool box" from synthetic hybrid zones and a theory of diversification. *Ecology and Evolution Taking Flight: Butterflies as Model Systems,* 281–318. University of Chicago Press, Chicago
4. Carroll SB. 2008 Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25-36. (doi:10.1016/j.cell.2008.06.030)
5. Kronforst MR, Papa R. 2015 The functional basis of wing patterning in *Heliconius* butterflies: the molecules behind mimicry. *Genetics* **200**, 1-19. (doi:10.1534/genetics.114.172387)
6. Deshmukh R, Baral S, Gandhimathi A, Kuwalekar M, Kunte K. 2018 Mimicry in butterflies: co-option and a bag of magnificent developmental genetic tricks. *Wires Dev. Biol.* **7**, 1-21. (doi:UNSP e29110.1002/wdev.291)

7. Timmermans MJ, Baxter SW, Clark R, Heckel DG, Vogel H, Collins S, Papanicolaou A, Fukova I, Joron M, Thompson MJ, *et al*. 2014 Comparative genomics of the mimicry switch in *Papilio dardanus*. *Proc. Biol. Sci.* **281**, 20140465. (doi:10.1098/rspb.2014.0465)
8. Jiggins CD, Wallbank RW, Hanly JJ. 2017 Waiting in the wings: what can we learn about gene co-option from the diversification of butterfly wing patterns? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **372**, 20150485. (doi:10.1098/rstb.2015.0485)
9. Kunte K. 2009 The diversity and evolution of Batesian mimicry in *Papilio* swallowtail butterflies. *Evolution* **63**, 2707-2716. (doi:10.1111/j.1558-5646.2009.00752.x)
10. VanKuren NW, Massardo D, Nallu S, Kronforst MR. 2019 Butterfly mimicry polymorphisms highlight phylogenetic limits of gene reuse in the evolution of diverse adaptations. *Mol. Biol. Evol.* **36**, 2842-2853. (doi:10.1093/molbev/msz194)
11. Wallace AR. 1869 XXI. Notes on eastern butterflies; (continued). *Transactions of the Royal Entomological Society of London* **17**, 321-349.
12. Punnett RC. 1911 "Mimicry" in Ceylon butterflies, with a suggestion as to the nature of polymorphism. *Spoila Zeylanica* **7**, 1-24 + 22 pl.
13. Wei CH, Lohman DJ, Peggie D, Yen SH. 2017 An illustrated checklist of the genus *Elymnias* Hübner, 1818 (Nymphalidae, Satyrinae). *Zookeys* **676**, 47-152. (doi:10.3897/zookeys.676.12579)
14. Lohman DJ, Sarino, Peggie D. 2020 Syntopic *Elymnias agondas aruana* female forms mimic different *Taenaris* model species (Papilionoidea: Nymphalidae: Satyrinae) on Aru, Indonesia. *Treubia* **47**, 1-12. (doi: 10.14203/treubia.v47i1.3821)
15. Vane-Wright RI. 1976 A unified classification of mimetic resemblances, *Biol. J. Linn. Soc.* **8**, 25-56. (https://doi.org/10.1111/j.1095-8312.1976.tb00240.x)
16. Butler, AG. 1871 A monograph of the Lepidoptera hitherto included in the genus *Elymnias*. *Proc. Zool. Soc. Lond.* **1871**, 518-525.

17. Allen CE, Zwaan BJ, Brakefield PM. 2011 Evolution of sexual dimorphism in the Lepidoptera. *Annu. Rev. Entomol.* **56**, 445-464. (doi:10.1146/annurev-ento-120709-144828)
18. Morris J, Navarro N, Rastas P, Rawlins LD, Sammy J, Mallet J, Dasmahapatra KK. 2019 The genetic architecture of adaptation: convergence and pleiotropy in *Heliconius* wing pattern evolution. *Heredity (Edinb)* **123**, 138-152. (doi:10.1038/s41437-018-0180-0)
19. Timmermans M, Srivathsan A, Collins S, Meier R, Vogler AP. 2020 Mimicry diversification in *Papilio dardanus* via a genomic inversion in the regulatory region of engrailed-invected. *Proc. Biol. Sci.* **287**, 20200443. (doi:10.1098/rspb.2020.0443).
20. Panettieri S, Gjinaj E, John G, Lohman DJ. 2018 Different ommochrome pigment mixtures enable sexually dimorphic Batesian mimicry in disjunct populations of the common palmfly butterfly, Elymnias hypermnestra. PLoS One 13, e0202465. (doi:10.1371/journal.pone.0202465)
21. Nishikawa H, Iijima T, Kajitani R, Yamaguchi J, Ando T, Suzuki Y, Sugano S, Fujiyama A, Kosugi S, Hirakawa H, *et al.* 2015 A genetic mechanism for female-limited Batesian mimicry in *Papilio* butterfly. *Nat. Genet.* **47**, 405-409. (doi:10.1038/ng.3241)

22. Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. 2017 GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202-2204. (doi:10.1093/bioinformatics/btx153)
23. Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simao FA, Ioannidis P, Seppey M, Loetscher A, Kriventseva EV. 2017 OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res.* **45**, D744-D749. (doi:10.1093/nar/gkw1119)
24. Waterhouse RM, Seppey M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2018 BUSCO applications from quality assessments to

454 gene prediction and phylogenomics. *Mol. Biol. Evol.* **35**, 543-548.
455 (doi:10.1093/molbev/msx319)
456 25. Alexander DH, Novembre J, Lange K. 2009 Fast model-based estimation of ancestry in
457 unrelated individuals. *Genome Res.* **19**, 1655-1664. (doi:10.1101/gr.094052.109)
458 26. Brito PH, Edwards SV. 2009 Multilocus phylogeography and phylogenetics using
459 sequence-based markers. *Genetica* **135**, 439-455. (doi:10.1007/s10709-008-9293-3)
460 27. Braby MF, Eastwood R, Murray N. 2012 The subspecies concept in butterflies: has its
461 application in taxonomy and conservation biology outlived its usefulness? *Biol. J. Linn.*
462 *Soc.* **106**, 699-716. (doi:10.1111/j.1095-8312.2012.01909.x)
463 28. Zhang W, Westerman E, Nitzany E, Palmer S, Kronforst MR. 2017 Tracing the origin and
464 evolution of supergene mimicry in butterflies. *Nat. Commun.* **8**, 1269.
465 (doi:10.1038/s41467-017-01370-1)
466 29. Petkova D, Novembre J, Stephens M. 2016 Visualizing spatial population structure with
467 estimated effective migration surfaces. *Nat. Genet.* **48**, 94-100. (doi:10.1038/ng.3464)
468 30. Lohman DJ, de Bruyn M, Page T, von Rintelen K, Hall R, Ng PKL, Shih H-T, Carvalho
469 GR, von Rintelen T. 2011 Biogeography of the Indo-Australian Archipelago. *Annu. Rev.*
470 *Ecol. Evol. Syst.* **42**, 205-226. (10.1146/annurev-ecolsys-102710-145001)
471 31. Zhou X, Stephens M. 2012 Genome-wide efficient mixed-model analysis for association
472 studies. *Nat. Genet.* **44**, 821-824. (doi:10.1038/ng.2310)
473 32. Mazo-Vargas A, Concha C, Livraghi L, Massardo D, Wallbank RWR, Zhang L, Papador
474 JD, Martinez-Najera D, Jiggins CD, Kronforst MR, *et al.* 2017 Macroevolutionary shifts
475 of *WntA* function potentiate butterfly wing-pattern diversity. *Proc. Natl. Acad. Sci. U S A*
476 **114**, 10701-10706. (doi:10.1073/pnas.1708149114)
477 33. Hines HM, Counterman BA, Papa R, Albuquerque de Moura P, Cardoso MZ, Linares M,
478 Mallet J, Reed RD, Jiggins CD, Kronforst MR, *et al*. 2011 Wing patterning gene

redefines the mimetic history of *Heliconius* butterflies. *Proc. Natl. Acad. Sci. U S A* **108**, 19666-19671. (doi:10.1073/pnas.1110096108)

34. Nishida R. 2017 Chemical ecology of poisonous butterflies: model or mimic? A paradox of sexual dimorphisms in Müllerian mimicry. *Diversity and Evolution of Butterfly Wing Patterns.* Springer, Singapore. (doi:10.1007/978-981-10-4956-9_11)

35. Nallu S, Hill JA, Don K, Sahagun C, Zhang W, Meslin C, Snell-Rood E, Clark NL, Morehouse NI, Bergelson J, *et al*. 2018 The molecular genetic basis of herbivory between butterflies and their host plants. *Nat Ecol Evol* **2**, 1418-1427. (doi:10.1038/s41559-018-0629-9)

36. Yuan J, Kessler SA. 2019 A genome-wide association study reveals a novel regulator of ovule number and fertility in *Arabidopsis thaliana*. *PLoS Genet.* **15**, e1007934. (doi:10.1371/journal.pgen.1007934)

37. Martin A, Papa R, Nadeau NJ, Hill RI, Counterman BA, Halder G, Jiggins CD, Kronforst MR, Long AD, McMillan WO, *et al*. 2012 Diversification of complex butterfly wing patterns by repeated regulatory evolution of a *Wnt* ligand. *Proc. Natl. Acad. Sci. U S A* **109**, 12632-12637. (doi:10.1073/pnas.1204800109)

38. Gallant JR, Imhoff VE, Martin A, Savage WK, Chamberlain NL, Pote BL, Peterson C, Smith GE, Evans B, Reed RD, *et al.* 2014 Ancient homology underlies adaptive mimetic diversity across butterflies. *Nat. Commun.* **5**, 4817. (doi:10.1038/ncomms5817)

39. Mullen SP, VanKuren NW, Zhang W, Nallu S, Kristiansen EB, Wuyun Q, Liu K, Hill RI, Briscoe AD, Kronforst MR. 2020 Disentangling population history and character evolution among hybridizing lineages. *Mol. Biol. Evol.* **37**, 1295-1305. (doi:10.1093/molbev/msaa004)

40. Chepelev I, Wei G, Wangsa D, Tang Q, Zhao K. 2012 Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. *Cell Res.* **22**, 490-503. (doi:10.1038/cr.2012.15)

505 41. Rondem KE. 2018 Characterizing the *optix* network in *Heliconius* butterfly wing color
506 patterning (Unpublished master's thesis). Cornell University, Ithaca, NY.
507 42. Lewis JJ, Geltman RC, Pollak PC, Rondem KE, Van Belleghem SM, Hubisz MJ, Munn PR,
508 Zhang L, Benson C, Mazo-Vargas A, *et al.* 2019 Parallel evolution of ancient, pleiotropic
509 enhancers underlies butterfly wing pattern mimicry. *Proc. Natl. Acad. Sci. U S A* **116**,
510 24174-24183. (doi:10.1073/pnas.1907068116)
511 43. Espeland M, Breinholt JW, Barbosa EP, Casagrande MM, Huertas B, Lamas G, Marin MA,
512 Mielke OHH, Miller JY, Nakahara S, *et al*. 2019 Four hundred shades of brown: higher
513 level phylogeny of the problematic Euptychiina (Lepidoptera, Nymphalidae, Satyrinae)
514 based on hybrid enrichment data. *Mol. Phylogenet. Evol.* **131**, 116-124.
515 (doi:10.1016/j.ympev.2018.10.039)
516 44. Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M, Harada
517 M, Nagayasu E, Maruyama H, *et al.* 2014 Efficient de novo assembly of highly
518 heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* **24**, 1384-
519 1395. (doi:10.1101/gr.170720.113)
520 45. Bolger AM, Lohse M, Usadel B. 2014 Trimmomatic: a flexible trimmer for Illumina
521 sequence data. *Bioinformatics* **30**, 2114-2120. (doi:10.1093/bioinformatics/btu170)
522 46. Pryszcz LP, Gabaldon T. 2016 Redundans: an assembly pipeline for highly heterozygous
523 genomes. *Nucleic Acids Res.* **44**, e113. (doi:10.1093/nar/gkw294)
524 47. Price AL, Jones NC, Pevzner PA. 2005 De novo identification of repeat families in large
525 genomes. *Bioinformatics* **21 Suppl 1**, i351-358. (doi:10.1093/bioinformatics/bti1018)
526 48. Smith DA, Gordon IJ, Traut W, Herren J, Collins S, Martins DJ, Saitoti K, Ireri P, Ffrench-
527 Constant R. 2016 A neo-W chromosome in a tropical butterfly links colour pattern, male-
528 killing, and speciation. *Proc. Biol. Sci.* **283**, 20160821. (doi:10.1098/rspb.2016.0821)
529 49. Ahola V, Lehtonen R, Somervuo P, Salmela L, Koskinen P, Rastas P, Valimaki N, Paulin
530 L, Kvist J, Wahlberg N, *et al*. 2014 The Glanville fritillary genome retains an ancient

karyotype and reveals selective chromosomal fusions in Lepidoptera. *Nat. Commun.* **5**, 4737. (doi:10.1038/ncomms5737)

50. Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, Lippman ZB, Schatz MC. 2019 RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biol* **20**, 224. (doi:10.1186/s13059-019-1829-6)

51. Holt C, Yandell M. 2011 MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491. (doi:10.1186/1471-2105-12-491)

52. Campbell MS, Holt C, Moore B, Yandell M. 2014 Genome annotation and curation using MAKER and MAKER-P. *Curr. Protoc. Bioinformatics* **48**, 11-39. (doi:10.1002/0471250953.bi0411s48)

53. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, *et al.* 2011 Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644-652. (doi:10.1038/nbt.1883)

54. The UniProt Consortium. 2018 UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **46**, 2699. (doi:10.1093/nar/gky092)

55. Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M. 2008 Gene prediction in novel fungal genomes using an ab initio algorithm with unsupervised training. *Genome Res.* **18**, 1979-1990. (doi:10.1101/gr.081612.108)

56. Martin, M. 2011 Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet Journal* **17**, 10-12. (https://doi.org/10.14806/ej.17.1.200)

57. Langmead B, Salzberg SL. 2012 Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357-359. (doi:10.1038/nmeth.1923)

58. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, *et al*. 2010 The Genome Analysis Toolkit: a MapReduce

framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297-1303. (doi:10.1101/gr.107524.110)

59. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, *et al.* 2011 A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet* **43**, 491-498. (doi:10.1038/ng.806)

60. Pritchard JK, Stephens M, Donnelly P. 2000 Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.

61. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, *et al*. 2007 PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559-575. (doi:10.1086/519795)

62. Baxter SW, Nadeau NJ, Maroja LS, Wilkinson P, Counterman BA, Dawson A, Beltran M, Perez-Espona S, Chamberlain N, Ferguson L, *et al*. 2010 Genomic hotspots for adaptation: the population genetics of Müllerian mimicry in the *Heliconius melpomene* clade. *PLoS Genet.* **6**, e1000794. (doi:10.1371/journal.pgen.1000794)

63. Bruen TC, Philippe H, Bryant D. 2006 A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**, 2665-2681. (doi:10.1534/genetics.105.048975)

64. Nguyen CD, Lee KJ, Carlin JB. 2015 Posterior predictive checking of multiple imputation models. *Biom. J.* **57**, 676-694. (doi:10.1002/bimj.201400034)

65. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587-589. (doi:10.1038/nmeth.4285)

66. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018 UFBoot2: Improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518-522. (doi:10.1093/molbev/msx281)

583 67. Zhang C, Rabiee M, Sayyari E, Mirarab S. 2018 ASTRAL-III: polynomial time species tree
584 reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**, 153.
585 (doi:10.1186/s12859-018-2129-y)
586 68. Benjamini Y, Hochberg Y. 1995 Controlling the false discovery rate: a practical and
587 powerful approach to multiple testing. *J. R. Statist. Soc. B* **57**, 289-300.
588
589

FIGURES

Figure 1.

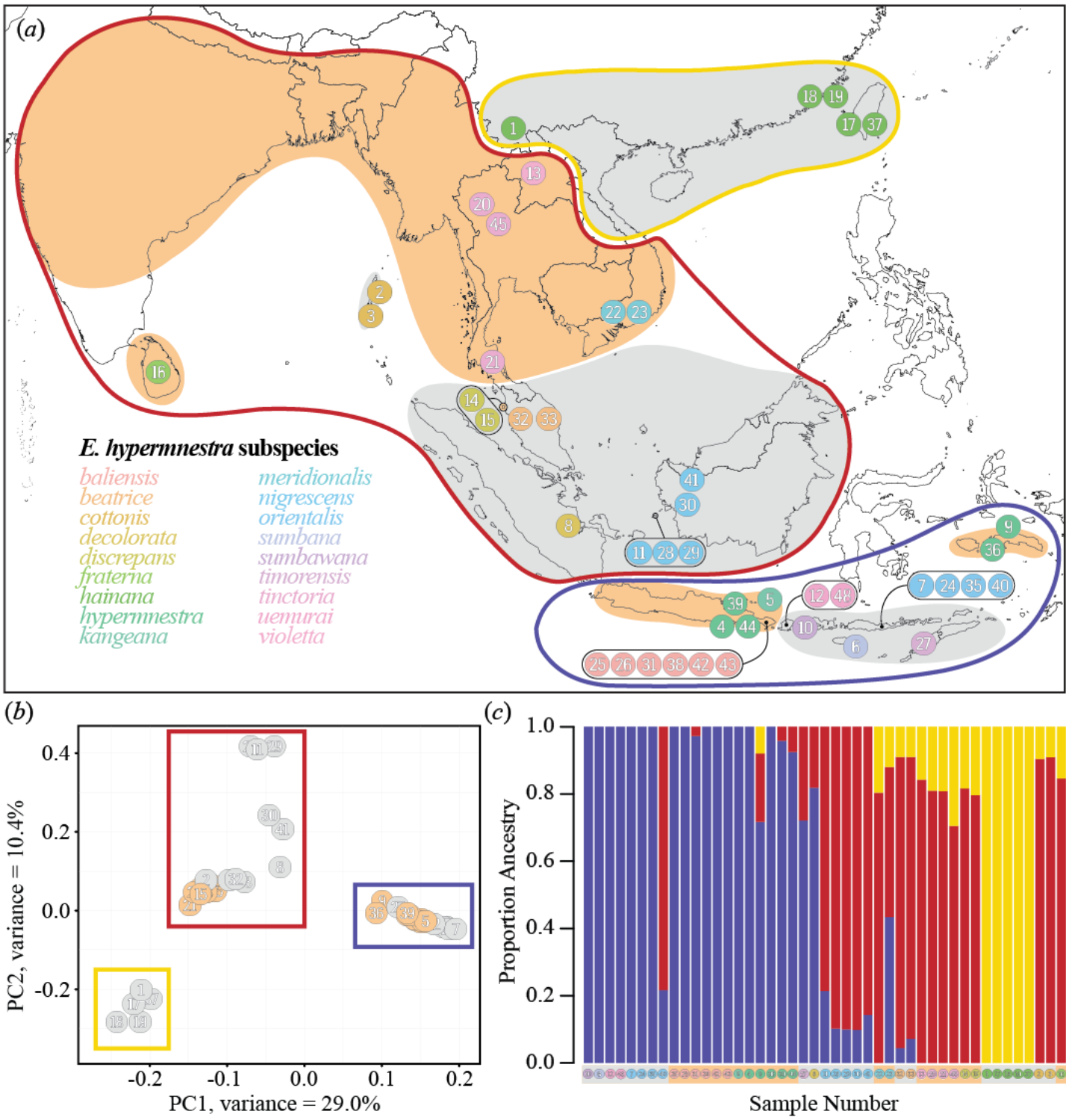

Figure 2.

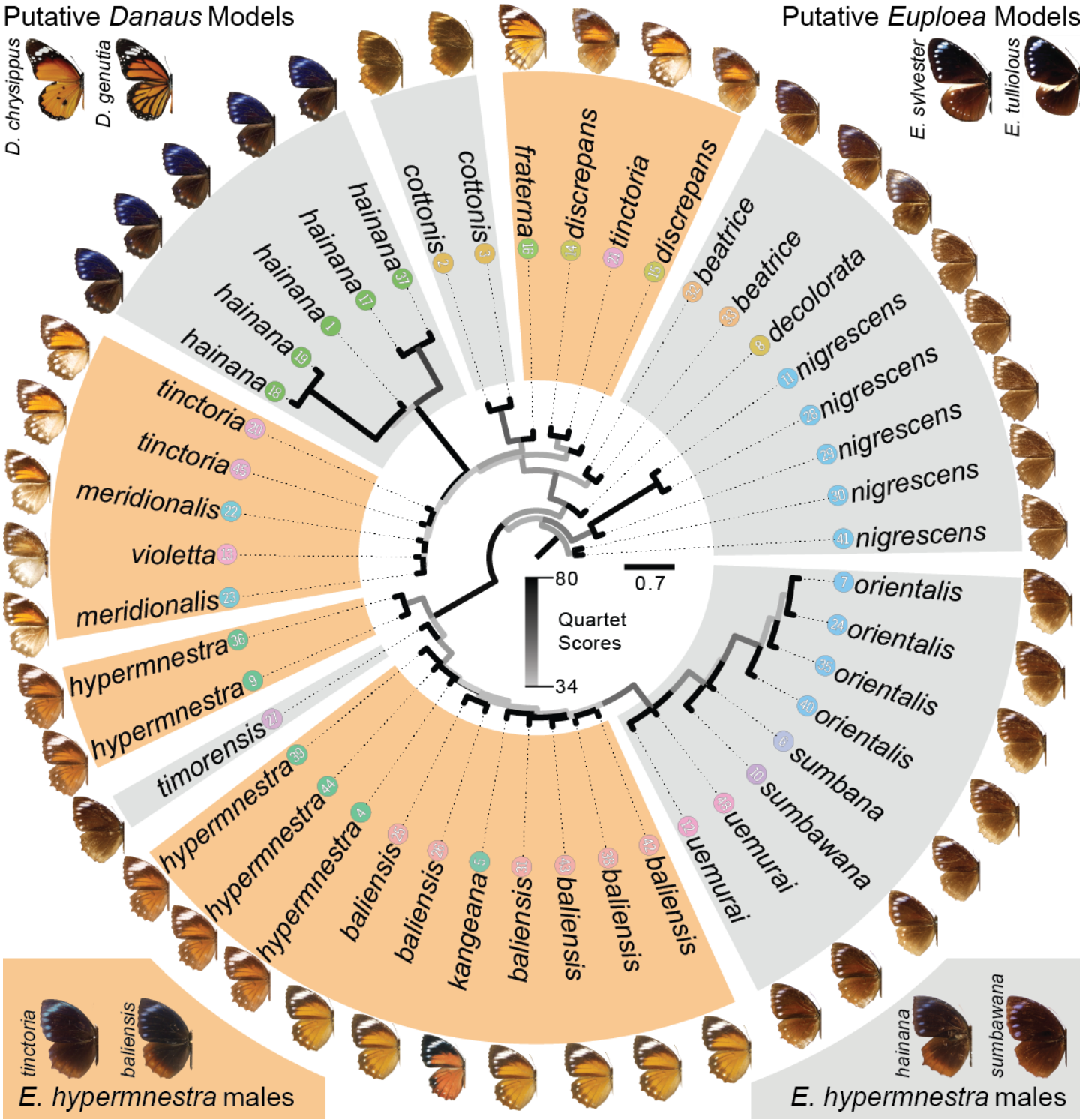
Putative *Danaus* Models
Putative *Euploea* Models
D. chrysippus
D. genutia
E. sylvester
E. tulliolous
hainana 37
hainana 17
hainana 1
hainana 19
hainana 18
cottonis 2
cottonis 3
fraterna 16
discrepans 14
tinctoria 20
discrepans 15
beatrice 32
beatrice 33
decolorata 8
nigrescens 11
nigrescens 28
nigrescens 29
nigrescens 30
nigrescens 41
tinctoria 20
tinctoria 45
meridionalis 22
violetta 13
meridionalis 23
hypermnestra 36
hypermnestra 9
timorensis 21
hypermnestra 39
hypermnestra 44
hypermnestra 4
baliensis 25
baliensis 26
kangeana 5
baliensis 31
baliensis 43
baliensis 38
baliensis 42
orientalis 7
orientalis 24
orientalis 35
orientalis 40
sumbana 6
sumbawana 10
uemurai 43
uemurai 12
80
0.7
Quartet Scores
34
tinctoria
baliensis
E. hypermnestra males
hainana
sumbawana
E. hypermnestra males

Figure 3

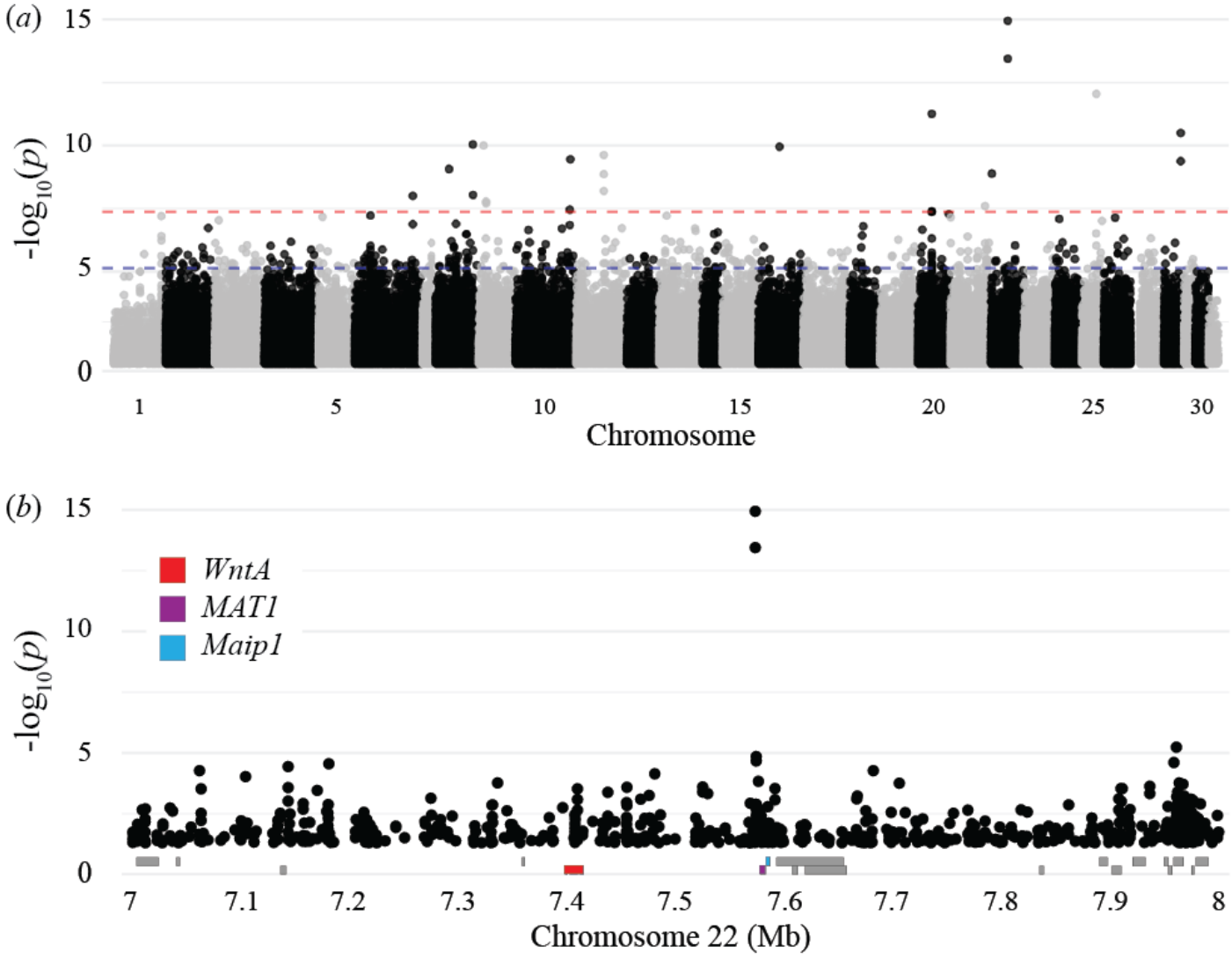

FIGURE LEGENDS

Figure 1. *Elymnias hypermnestra* comprises three genetically and geographically distinct populations. (a) The geographic distribution of 48 *E. hypermnestra* populations representing 15 subspecies. Orange females and melanic females are indicated with background colors, demonstrating disjunct distributions of each color pattern. Collection locations of each specimen used in this study are indicated with its sample ID (electronic supplementary material, table S1), which is colored to indicate its subspecies. The dark outlines on the map indicate genetically distinct populations, as inferred by (b) principal component analysis. The points in this plot indicate sample ID and color pattern. The same three populations are indicated by an (c) ADMIXTURE plot. The sample ID and color pattern are indicated below each bar.

Figure 2. An ASTRAL species tree of *Elymnias hypermnestra* based on 3000 random autosomal 10 kb windows infers multiple clades of melanic and orange female forms. Branch color indicates quartet score branch support. The sample IDs correspond to the same numbers in Fig. 1, and their color indicates subspecies affiliation. Orange or dark backgrounds indicate the female color pattern of the lineage, and representative images of females of the same subspecies as each sample are shown around the periphery. Images of the putative model species mimicked by orange and melanic females are provided at the top. Representative males of four subspecies are shown at the bottom.

Figure 3. (a) Association between *Elymnias hypermnestra* female color pattern and genetic variation. *p*-values are from SNP-wise Wald tests. Blue and red dashed lines represent the 10% and 1% false discovery rates (FDR), respectively. The full GWA results (with unplaced scaffolds) are shown in electronic supplementary material, figure S6. (b) An enlargement of chromosome 22 in Figure 4a; depicting the region of the two SNPs most significantly associated with female color pattern. The locations of 3 nearby genes in the *Melitaea cinxia* reference genome are shown below the plot.