Supplementary Information for

Disentangling population history and character evolution among hybridizing lineages.

Sean P. Mullen[1]*, Nicholas W. VanKuren[2], Wei Zhang[3], Sumitha Nallu[2], Evan B. Kristiansen[1], Qiqige Wuyun[4], Kevin Liu[4], Ryan I Hill[5], Adriana D. Briscoe[6], and Marcus R. Kronforst[2]

*Sean P. Mullen
Email: smullen@bu.edu

This PDF file includes:

Supplementary Methods

De novo reference genome assembly

Paired-end and mate-pair libraries (Table S1) were constructed from genomic DNA isolated from wild-caught *L. a. astyanax* (n=5; collected along a 2km stretch of unimproved road in Pennsylvania State Game Lands #57 in Tunkhannock, PA), and sequenced (~75x) at the Beijing Genome Institute on Illumina's HiSeq platform. Raw reads from three PE libraries, 250 bp, 500 bp and 800 bp and four mate pair libraries, 2kb, 5kb, 10kb and 20kb libraries were filtered for base quality (Qual >20, low quality rate 0.2), read length (<30bp discarded), and trimmed of adapters. All trimmed reads were assembled into scaffolds using the *Platanus* assembler (version: 1.2.4; Kajitani et al. 2014) with default parameter settings. We then obtained 8.5 million long reads from PacBio sequencing from a single lab-reared *L. a. astyanax* pupae, inbred for ~6 generations, derived from the same wild population as the indivduals used for short-read sequencing.The raw reads were corrected using the *Canu* (v1.5; Koren et al. 2017) software with the -correct option resulting in 3,631,966 corrected reads with a N50 of 3,378 bp. In addition to the PacBio reads, we also generated 19,355 scaffolds with an N50 of 80kb from 2,124 BAC's (see Gallant et al. 2014 for specimen info). Finally, the scaffolds generated from *Platanus*, corrected PacBio reads and the assembled scaffolds from the BAC libraries were passed through the *Redundans* (v0.13a; Pryszcz and Gabaldón 2016) pipeline with --longreads option to generate a scaffolded homozygous genome assembly. Genome FASTA was linked to NCBI Bioproject #: PRJNA556447. Raw reads used for the assembly were uploaded to NCBI's short read archive (SRA): SUB6048049

Identifying autosomal scaffolds

We extracted 261 scaffolds above 50 kb and blasted them against 577 genes on the Z chromosome in *Heliconius melpomene* according to *H. melpomene* v2 (Davey et al. 2016). We collected scaffolds with more than one reciprocal best hit. Then we selected

five male and five female samples from *Limenitis lorquini* and *L. weidemeyerii* and estimated genome-wide read depth and read depth for each scaffold using *VCFtools*. An additional filter step was applied by adapting the method mentioned by (Vicoso et al. 2013). For each sample, the read depth of each scaffold was calibrated dividing by the genome-wide read depth, and then the mean calibrated depths of each scaffold were calculated for five female samples and five male samples, separately. We divided the male mean depth by the female mean depth and plotted their log2 values. Autosomal scaffolds should have values close to zero, whereas Z-linked scaffolds should have values close to one. We used log2 = 0.3 as a cutoff and identified ten candidate Z-linked scaffolds, eight of which passed the previous blast test. We considered the rest scaffolds as autosomal scaffolds for downstream analyses, and assigned them to *Melitaea cinxia* chromosomes using a custom BLAT pipeline (Kent 2002; Ahola et al. 2014). Ordering information was used to produce genome-wide plots.

Genome annotation

We annotated the final assembly using *MAKER* v3.01.02 (Campbell et al. 2014). We used RNA-seq data originally generated by Gallant et al (2014), which was derived from 5th instar larval and pupal wing discs (n=12 individuals) that we assembled using *Trinity* (Grabherr et al. 2011; Haas et al. 2013), as evidence for transcribed regions. In addition, we used protein sequences from the UniProt/SwissProt protein database, and GenBank or RefSeq protein models for *Danaus plexippus* (Zhan et al. 2011; GCA_000235995.2), *Papilio xuthus* (Nishikawa et al. 2015; GCF_000836235.1), *Bombyx mori* (Consortium and others 2008; GCF_000151625.1), *Vanessa tameamea* (GCF_002938995.1), *Pieris rapae* (Shen et al. 2016; GCF_001856805.1), and *Drosophila melanogaster* (Adams et al. 2000; GCF_000001215.4) as evidence for protein-coding regions. We trained *SNAP* (Korf 2004) over three rounds using this evidence, then used *SNAP*, *Augustus* v3.2 (Stanke et al. 2008) with *Heliconius melpomene* parameters, and *GeneMark-ES* 4 (Ter-Hovhannisyan et al. 2008) with MAKER to generate the final gene models. Finally, we functionally annotated predicted proteins using BLASTp of all predicted proteins against the SwissProt database and combined that information using scripts included in the *MAKER* package. We performed whole genome BUSCO analysis using *BUSCO* v3

(Waterhouse et al. 2017) using default settings and the Endopterygota database (2,440 SCOs) from OrthoDBv10 (Kriventseva et al. 2018).

Annotation of pigmentation genes

Translated nucleotide sequences of *Vanessa cardui* melanin and ommochrome-pathway genes identified in Zhang et al. (Zhang et al. 2016) were used as query sequences for *tblastn* searches of a *Limenitis arthemis astanax* wing RNA-seq transcriptome. Individual transcripts were aligned to the *V. cardui* sequence and trimmed, then the translated *L. arthemis astyanax* sequences were used as query sequences in blastn searches against the *L. arthemis astyanax* reference genome. Completeness of individual genes in the genome was verified by confirming the presence of start and stop codons, all exons, and a lack of scaffold miss-assembly. These sequences were deposited in Genbank under the accession numbers: MN842725-MN842774.

Whole genome resequencing

We generated genome re-sequencing data for 65 butterflies (Table S4) and processed the raw reads with the *Trimmomatic* Version 0.36 . The reads with high quality were aligned to the reference genome using *Bowtie2* v 2.3.0 with the option –very-sensitive-local. PCR duplicates were removed by Picard v2.8.1 (Grabherr et al. 2011; Haas et al. 2013). Indels were realigned by RealignerTargetCreator and IndelRealigner, and genotypes were called by UnifiedGenotyper in *GATK* v3.7. The population genomic data for *L. a. arthemis* and *L. a. astyanax* were previously generated by Gallant et al. 2014 (2014) (NCBI short read archive: PRJNA252628).These data are archived under SRA accession number: SUB6066536

Phylogenetic Analysis

We extracted genotype calls (44.40 Mb) with good quality (Qual > 30) from 12 individuals and constructed a genome-wide maximum-likelihood phylogeny using *RAxML* (Stamatakis 2014) with a GTRGAMMA model and 100 bootstrap replicates. We used *iTOL* to output tree images. To evaluate genome-wide patterns of genealogical

discordance, we employed Martin and Van Belleghem's (2017) topology weighting analysis using iterative sampling of subtrees. Two individuals of arthemis, astyanax, arizonensis, and archippus with the best read depth were chosen for this analysis. Maximum-likelihood trees were constructed with 100 bootstrap replicates using a general time-reversible (GTR) model for each 50kb window containing at least 200 SNPs. Trees support three different topologies were counted. Topology 1 (((*arthemis*, *astyanax*),*arizonensis*), *archippus*) corresponds to the species tree and was supported by 60,490 subtrees. Topology 2 (((*astyanax*, *arizonensis*), *arthemis*), archippus) corresponds to monophyly of mimetic individuals and was supported by 18,741 subtrees. Topology 3 (((*arthemis*, *arizionensis*), *astyanax*), *archippus*) represents a sister relationship between allopatric, mimetic *arizonensis* and non-mimetic *arthemis*, and was supported by 17,889 subtrees.

Demographic analyses using G-PhoCS

We inferred demographic parameters such as population sizes, divergence times as well as migration rates using *G-PhoCS* (Gronau et al. 2011), which employs a Markov Chain Monte Carlo (MCMC) sampling strategy. We selected seven samples with good sequencing depth, including: AZ11, IV2, RIH2093, RIH2125, V2, GA4 and VT44. We filtered genomic scaffolds smaller than our N50 (2.16Mb), and applied additional filters (see Table S5) to exclude repetitive elements, highly conserved regions, and/or genomic regions situated closely to protein coding genes potential under selection. After applying filters, we extracted 1,732 putatively neutral loci for use in the downstream G-PhoCS analysis. It is important to also note that we excluded the scaffold containing *WntA* from this analysis entirely because of *a priori* evidence for selection in this genomic region. To infer the demographic history of this radiation, we performed 200,000 MCMC iterations using the default Gamma distribution settings. We then viewed and estimated the MCMC traces using *Tracer* v1.6 (http://beast.bio.ed.ac.uk/Tracer). We assumed an average mutation rate of 3.0×10^{-9} and an average generation time of 0.25 per year and calibrated the raw estimates accordingly (Freedman et al. 2014). We conducted analyses for one no-migration model (Table S6) and 16 models with potential migration bands according to Table S7. We covered each potential migration band twice and determined significant

migration bands with all the 95% HPD lower bounds above zero in independent tests. We performed a full model test with all the significant migration bands (Table S8).

Population genomic analysis

We applied the Patterson's *D*-statistic to characterize genome-wide patterns of introgression among the three ingroup taxa *Limenitis a. arthemis*, *L. a. astyanax*, *L. a. arizonensis* using *L. archippus* as a designated outgroup. We used all observed ABBA and BABA sites, regardless of ancestral state, and the *D*-statistic was calculated following Durand et al. (2011) as

$$D(P_1, P_2, P_3, O) = \frac{\sum_{i=1}^{n}[(1-\hat{P}_{i1})\hat{P}_{i2}\hat{P}_{i3}(1-\hat{P}_{i4}) - \hat{P}_{i1}(1-\hat{P}_{i2})\hat{P}_{i3}(1-\hat{P}_{i4})]}{\sum_{i=1}^{n}[(1-\hat{P}_{i1})\hat{P}_{i2}\hat{P}_{i3}(1-\hat{P}_{i4}) + \hat{P}_{i1}(1-\hat{P}_{i2})\hat{P}_{i3}(1-\hat{P}_{i4})]} \quad (1)$$

where P_1, P_2, P_3 and P_4 refer to four taxa and $\hat{P}_{ij}$ refers to the SNP frequency in the corresponding population. We chose a window block size of 50 kb to explore patterns of allele-sharing across the genome, and employed the jackknife approach to calculate the standard error using an R package, bootstrap ver. 2012.04. We used a smaller fixed window size of 5 kb for the *D*-statistic across the *WntA* scaffold to obtain a more fine-grained portrait of allele-sharing in the region of the genome known to be responsible for mimetic color pattern variation. We also calculated mean pairwise sequence divergence (d_{XY}) across the whole genome (50kb window size), and within the *WntA* region (5kb window size), among *Limenitis a. arthemis*, *L. a. astyanax*, *L. a. arizonensis*, using the following equation:

$$d_{xy} = \frac{1}{n}\sum_{i=1}^{n}\hat{p}_{ix}(1-\hat{p}_{iy}) + \hat{p}_{iy}(1-\hat{p}_{ix}) \quad (2)$$

where $\hat{P}$ refers to the reference allele frequency in the corresponding population.

Hemiplasy Risk Factor

We calculated the hemiplasy risk factor following Guerrero and Hahn (2018). We generated Bayesian phylogenies for 700 randomly-chosen 100 kb windows using MrBayes3.2 (Ronquist et al., 2012) under a GTRGAMMA model of nucleotide substitution for 1,000,000 MCMC generations. We used the results from these 700

windows to calculate a Bayesian concordance tree with BUCKY v1.4.4, discarding the first 25% of MrBayes trees and otherwise default parameters (Ané et al., 2006; Larget et al., 2010) and input the resulting tree with branch lengths in coalescent units into the *PePo* R package provided by Guerrero and Hahn (2018).

Tests for Introgression

PhyloNet-HMM (Liu et al. 2014), a statistical introgression mapping model, was used to distinguish between heterogeneous genomic signatures left by point mutations, genetic drift and lineage sorting, recombination, and gene flow. Phylonet-HMM utilizes a combined statistical model that integrates the multi-species network coalescent model (Yu et al. 2012), a finite sites substitution model such as the General Time Reversible model of nucleotide substitutions (Rodriguez et al. 1990), and a hidden Markov model (HMM). The species phylogeny and aligned genome resequencing data (for the seven individuals with the best sequencing depth; see G-PhoCS methods) were used as input, and we tested three different species network hypotheses (Fig. 4) that each had a single reticulation. PhyloNet-HMM outputs an annotation of each site along the aligned genomes with an introgression probability, which in turn is used to assess confidence of detected introgression region. Introgression probabilities were determined based on a modified posterior decoding probability, which was calculated by averaging the posterior decoding probabilities of Hidden Markov Model (HMM) states corresponding to local coalescent histories. Here, we used the Phylonet-HMM implementation that is provided in the recently released PhyloNet version 3.6 (Than et al. 2008; Wen et al. 2018). This implementation utilizes a slightly different transition parameterization compared to the earlier model proposed by Liu et al. (2014), which is implemented in an earlier version 0.1 of the PhyloNet-HMM software. For running PhyloNet-HMM, we used the default settings as specified in the following PhyloNet configuration file.

```
NEXUS
BEGIN NETWORKS;
Network net = <network>;
END;
Begin DATA;
dimensions ntax=<number of taxa> nchar=<length of
```

```
sequence>; format datatype=dna symbols="ACTG" missing=?
gap=-;
matrix
1 <sequence 1>
2 <sequence 2>
.....
;
END;
BEGIN PHYLONET;
HmmCommand net -allelemap <allele mapping> -outputdirectory
"<output dir>" -threads 1 -numberofruns 10 -iterations 300
-noplots;
END;
```

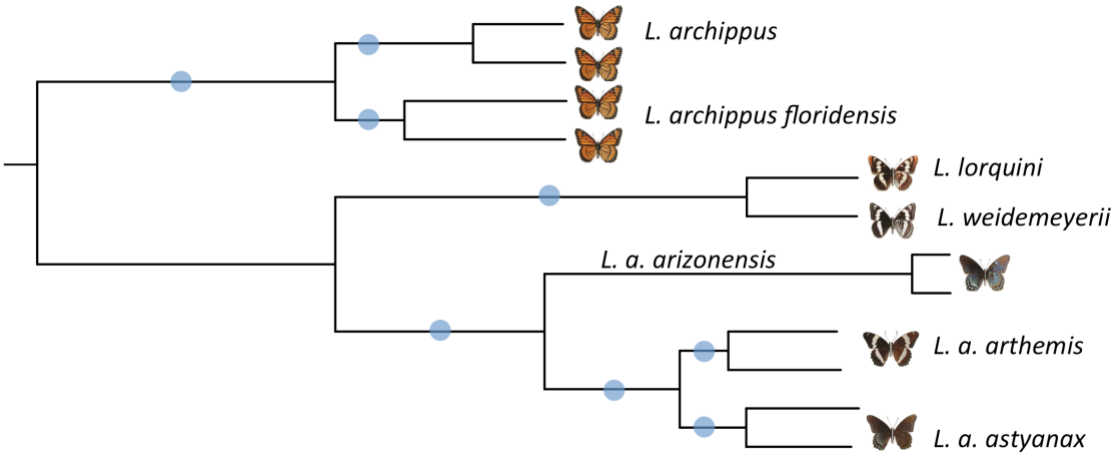

Fig. S1. Maximum-likelihood bootstrap (n=100) tree, for 12 individuals with the best sequence depth, generated using genome-wide concatenated SNPs (Q>30) and GTRGAMMA model implemented in RAxML(Stamatakis 2014) ; tree was rooted using genome sequence data from *Heliconius sara* (not shown). Blue circles indicate branches with >95% bootstap support.

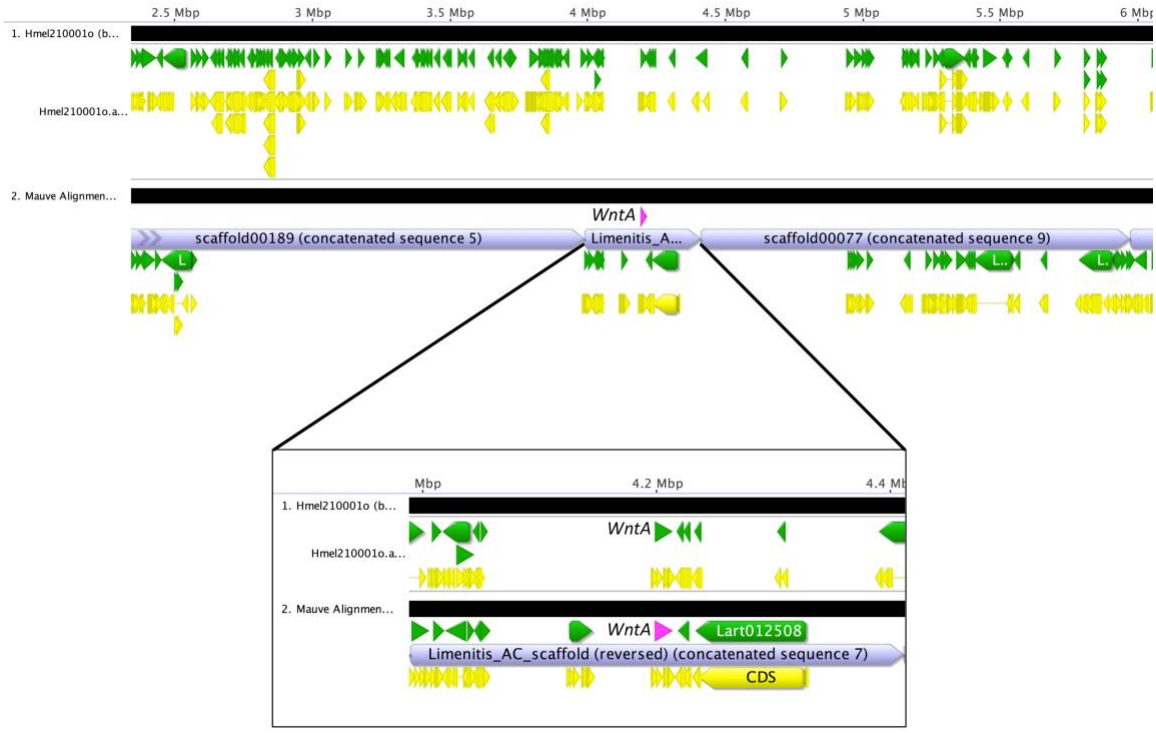

Fig. S2. Mauve alignment of the *Heliconius melpomene* scaffold (*Hmel*210001o) housing the *WntA* locus, the *Limenitis WntA* scaffold, and other *Limenitis* scaffolds previously mapped and reordered to this region of the *Hmel* genome using a custom BLAT pipeline. Images show the *H. melpomene* chromosome on top and the *Limenitis* scaffold below. Gene models (green) and coding sequences (CDS, yellow) are show for each genome. Inset panel shows alignment around the *WntA* coding sequence demonstrating high concordance between the two assemblies in this region.

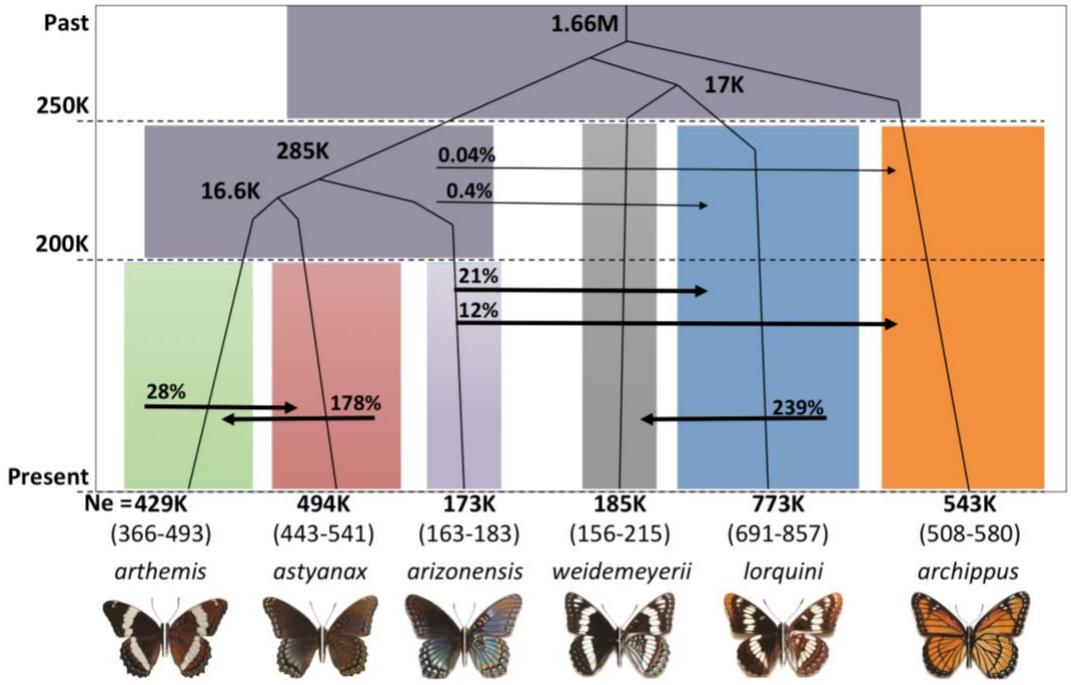

Fig. S3. Graphical summary of the full G-PhoCS demographic model (see also Table S7). Numbers at internal nodes represent estimated ancestral effective population sizes (*Ne*). Mean and 95% CIs for current effective population sizes shown above each *Limenitis* taxon. Divergence time estimates in years (K=thousand, M=million) are noted on the Y-axis. Arrows between lineages reflect probability of migration between bands at each time point. Note that migration estimates (M) outputted by G-PhoCS are scaled by tau (age*mu/generation time) to obtain total migration rates. Values over 100% are possible under scenarios where rates of gene flow are very high relative to the duration of time corresponding to a particular migration band.

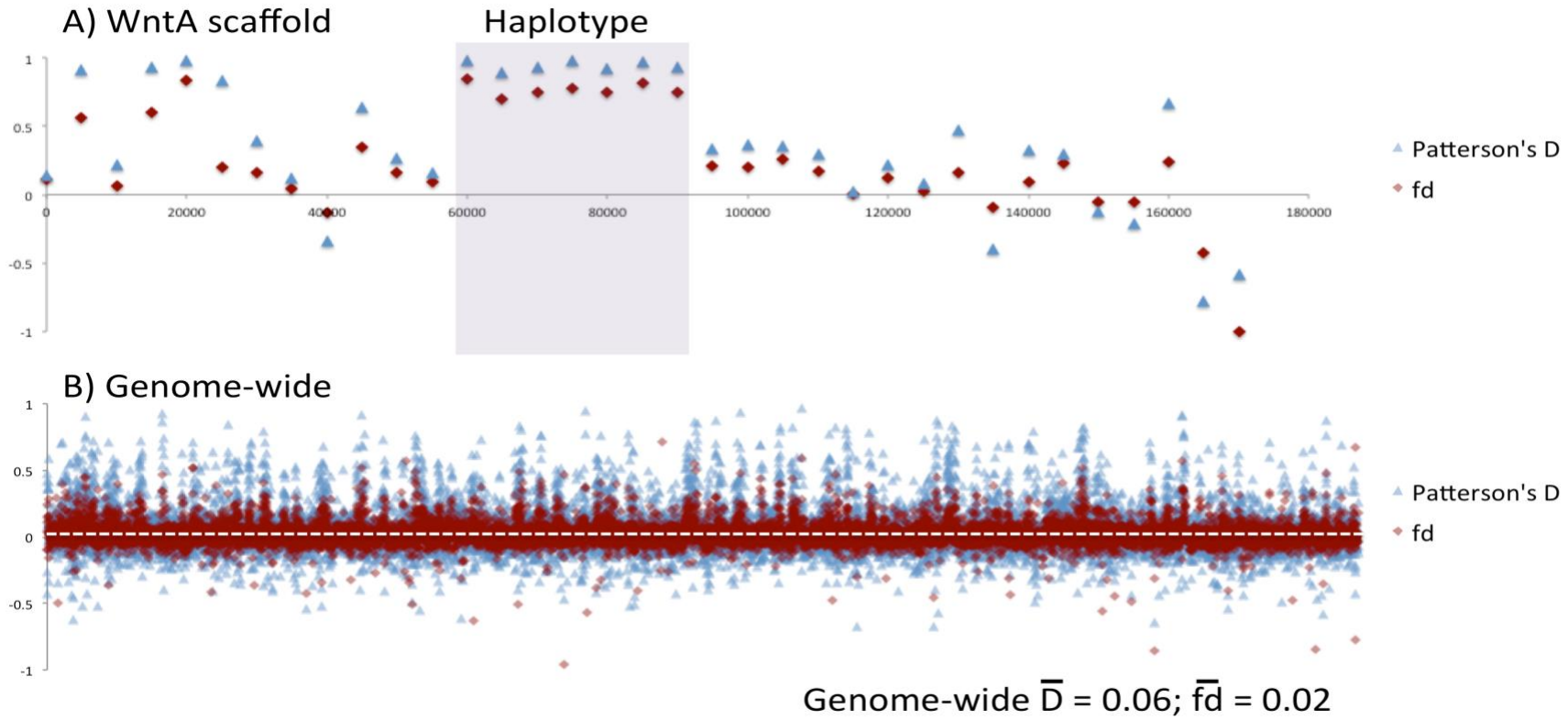

Fig. S4. A) Plot of Patterson's D (Durand et al. 2011) and Martin et al.'s (2014) fd calculated in 5kb windows across the WntA scaffold. The haplotype associated with mimetic variation is highlighted in light grey. B) Genome-wide estimates of D and *fd* calculated in 50Kb windows, excluding the WntA focal scaffold. Values for D and fd are on the y-axis and each windowed point (n = 6000), determined by sliding across individual scaffolds ordered largest to smallest, is plotted on the X-axis. The genome-wide average for D (0.06 +/- 0.0022) is shown as a dashed white line.

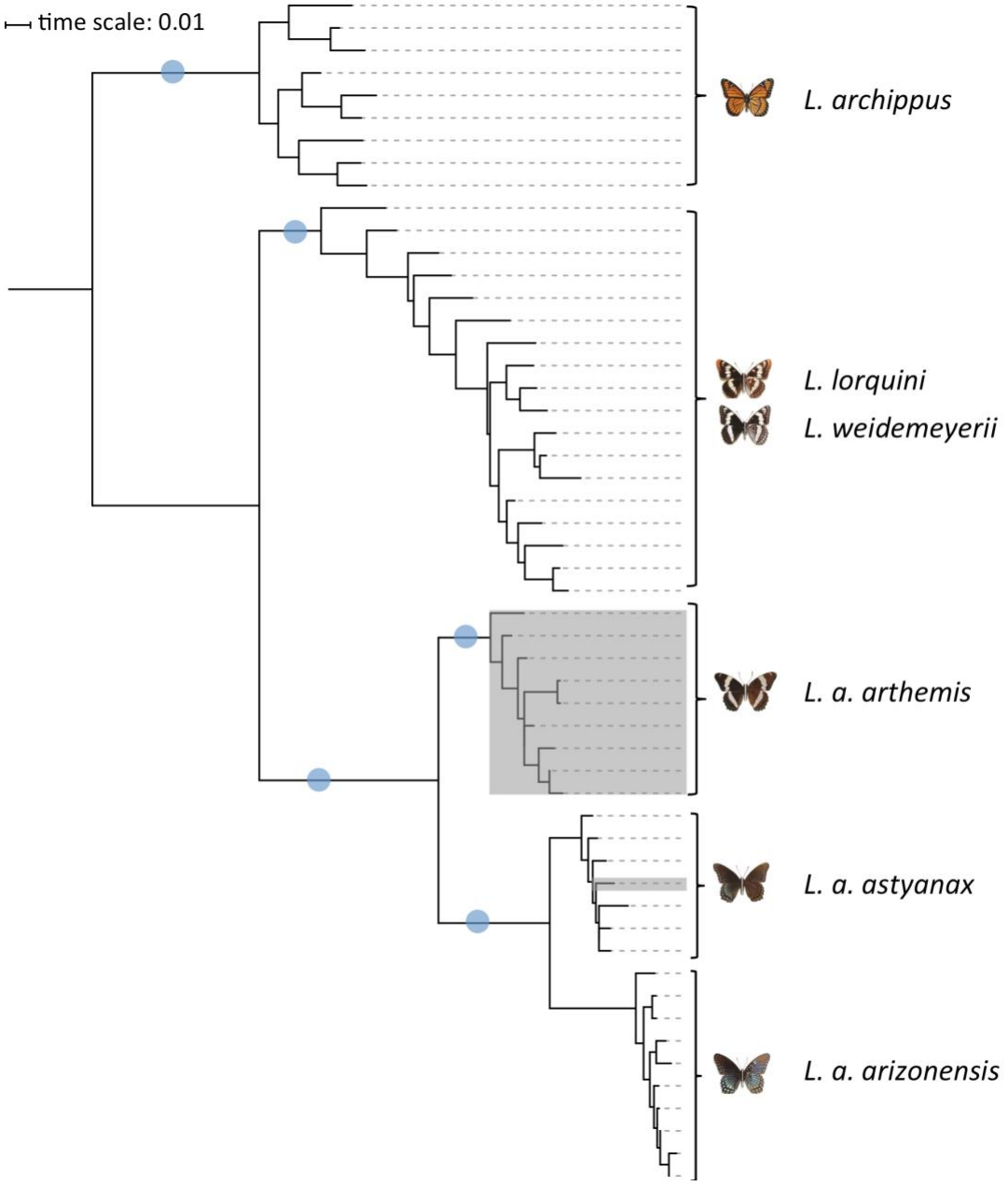

Fig. S5. Maximum-likelihood bootstrap (N=100) tree for the *WntA* protein-coding gene. Gray shading denotes all white-banded *L. a. arthemis* individuals; note that one individual sample of *L. a. arthemis* groups with the clade containing mimetic *L. a. astyanax*. Blue circles indicate branches with >95% support.

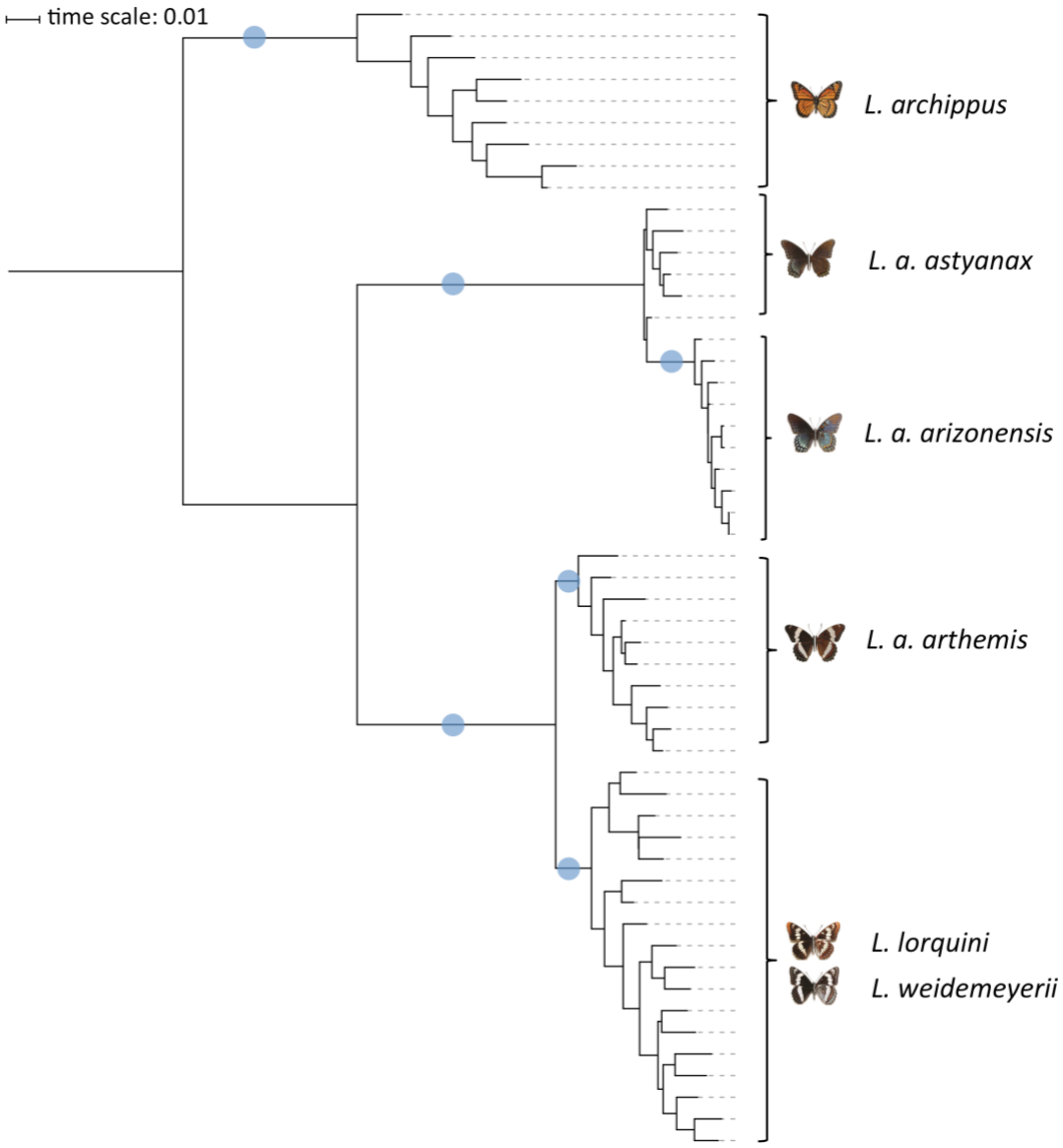

Fig. S6. Maximum-likelihood bootstrap (N=100) tree for the associated haplotype upstream of *WntA*. Blue circles indicate branches with >95% support. Monophyly of the two mimetic taxa, *L. a. astyanax* and *L. a. arizonensis* is strongly support.

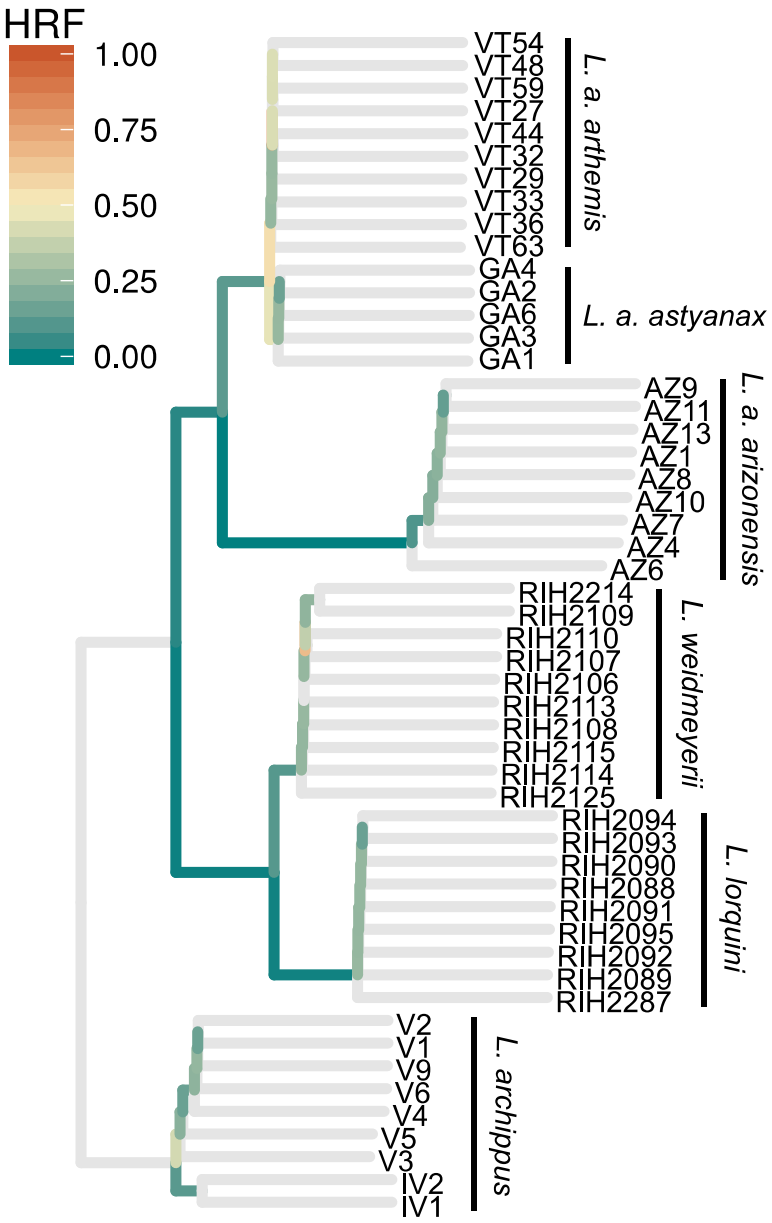

Fig. S7. Bayesian consensus phylogeny based on 700 randomly-chosen 100Kb genomic windows showing the calculated hemiplasy risk (following Guerrero and Hahn 2018) across branches of the Limenitis phylogeny.

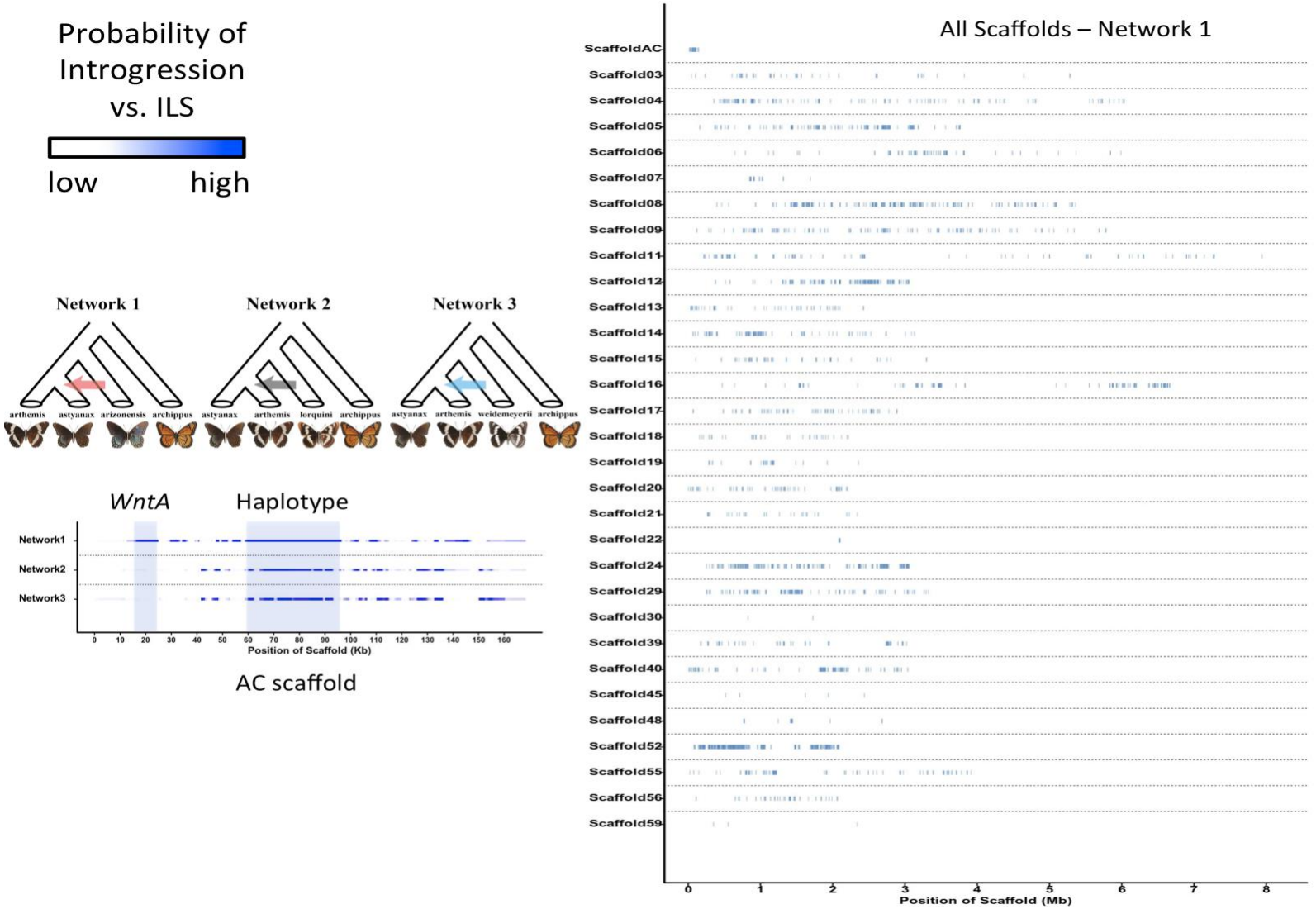

Fig. S8 PhyloNet-HMM analysis of 30 largest *Limenitis* scaffolds. Per-site introgression probabilities inferred using PhyloNet-HMM for network 1 are shown, where probabilities between 0 and 1 are colored using a continuous gradient from white to blue, respectively.

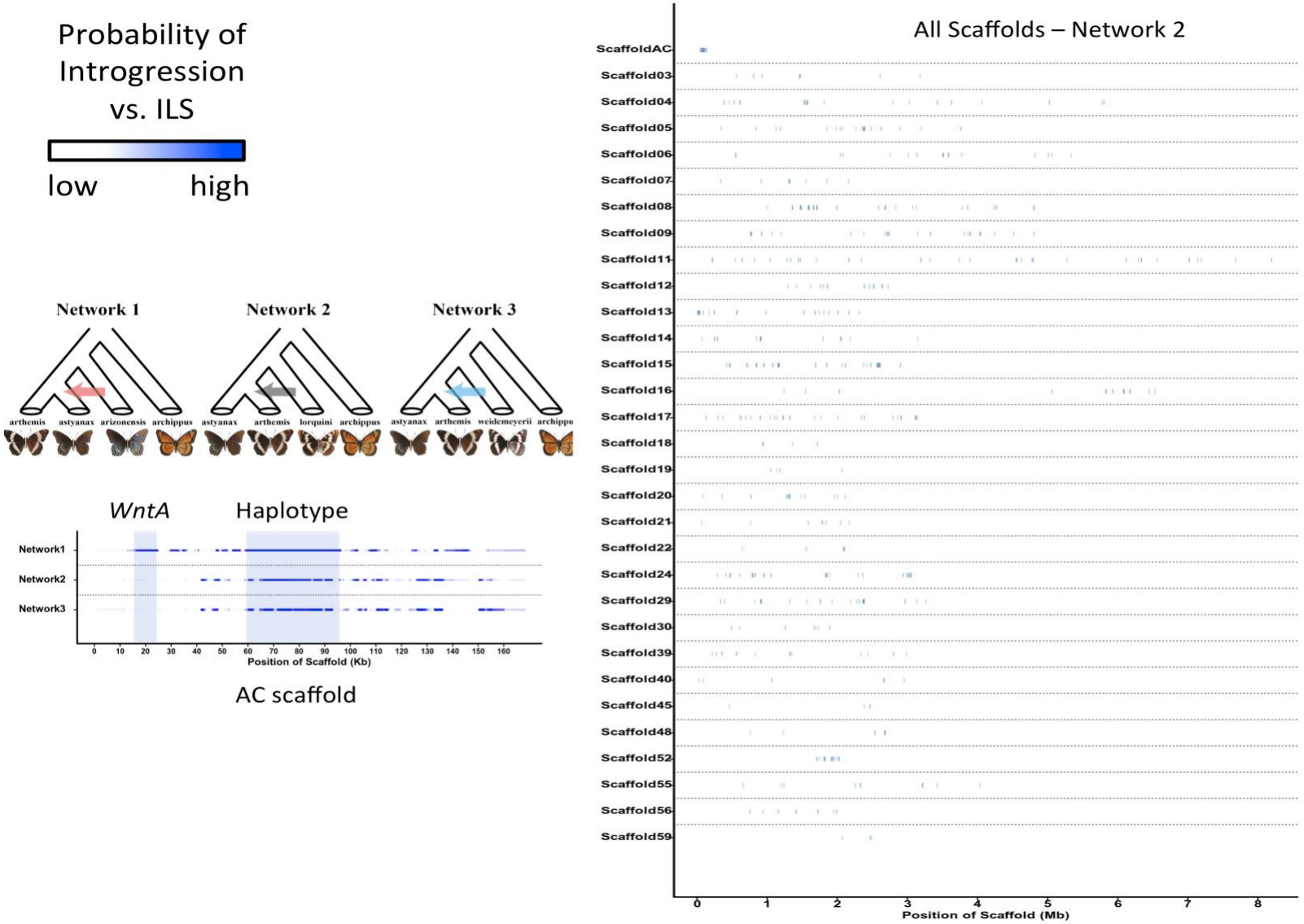

Fig. S9 PhyloNet-HMM analysis of 30 largest *Limenitis* scaffolds. Per-site introgression probabilities inferred using PhyloNet-HMM for network 2 are shown, where probabilities between 0 and 1 are colored using a continuous gradient from white to blue, respectively.

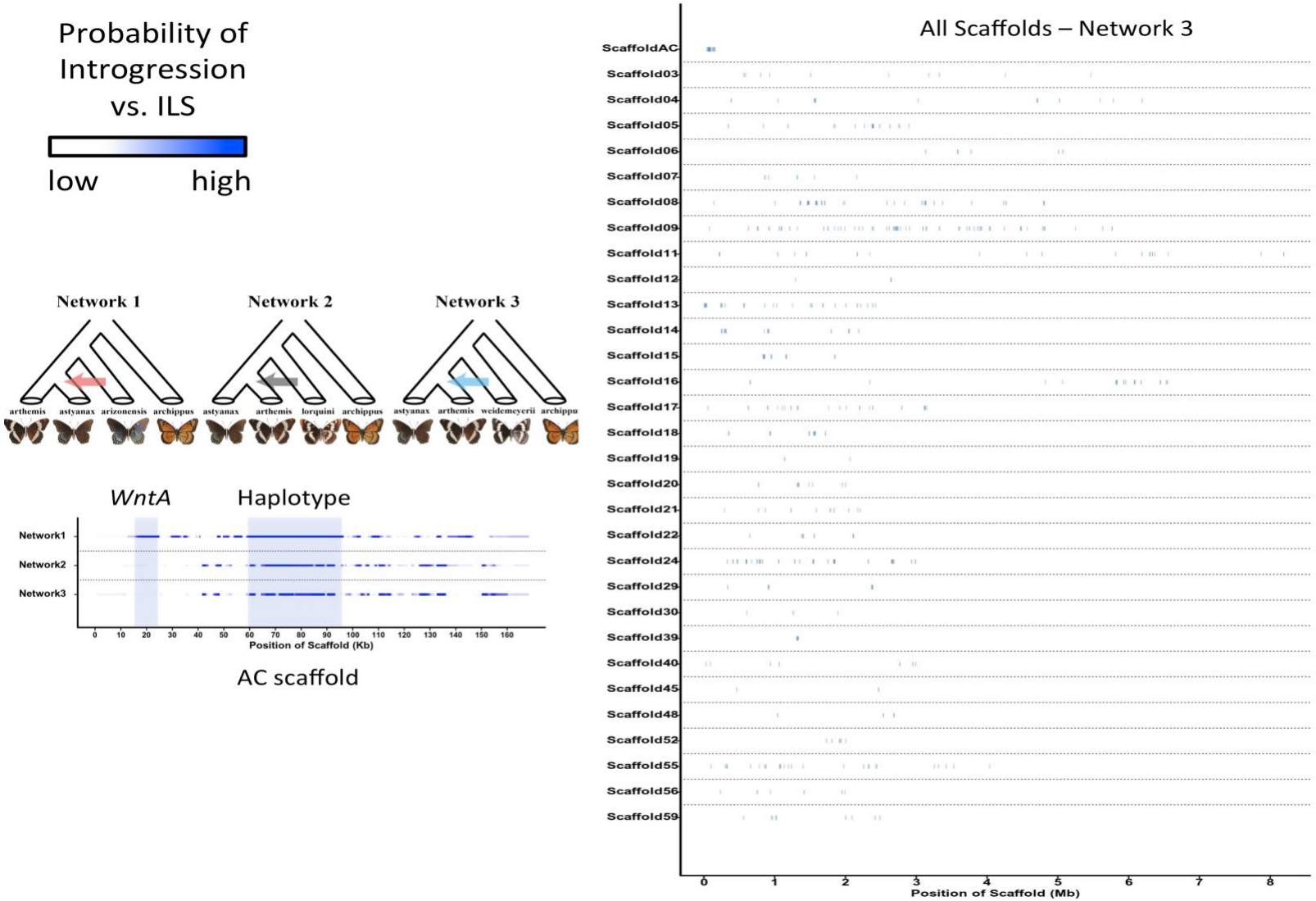

Fig. S10 PhyloNet-HMM analysis of 30 largest *Limenitis* scaffolds. Per-site introgression probabilities inferred using PhyloNet-HMM for network 3 are shown, where probabilities between 0 and 1 are colored using a continuous gradient from white to blue, respectively.

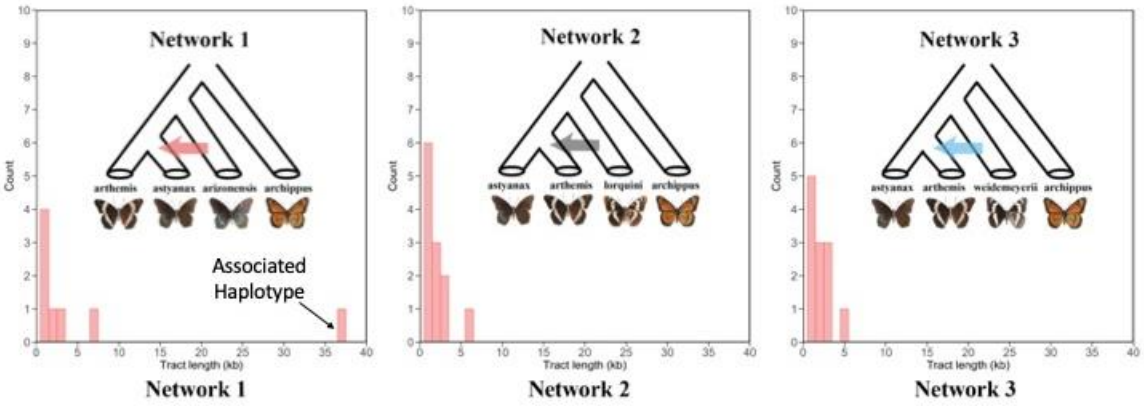

Fig. S11. Introgressed tract length histogram based on PHyloNet-HMM analysis of the *Limenitis WntA* scaffold with comparison of three species networks

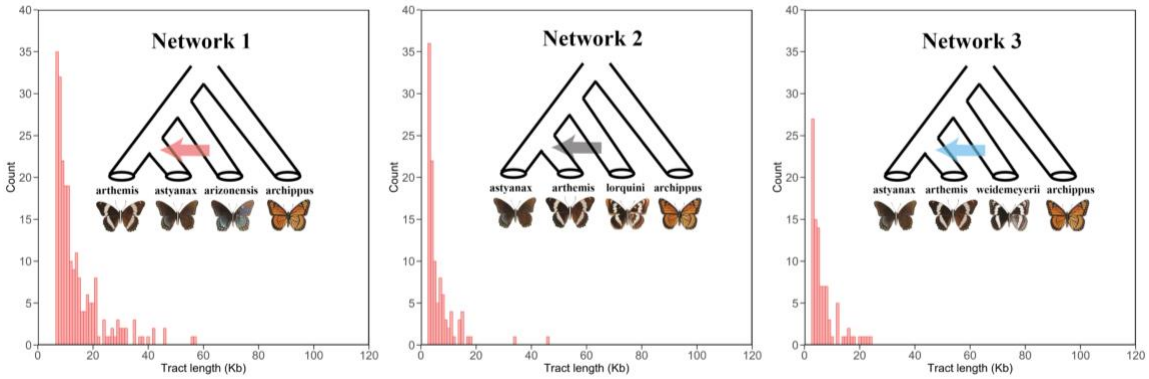

Fig. S12. Introgressed tract length histogram based on Phylo-Net HMM analyses of all non-focal genomic scaffolds for each of the three tested species networks

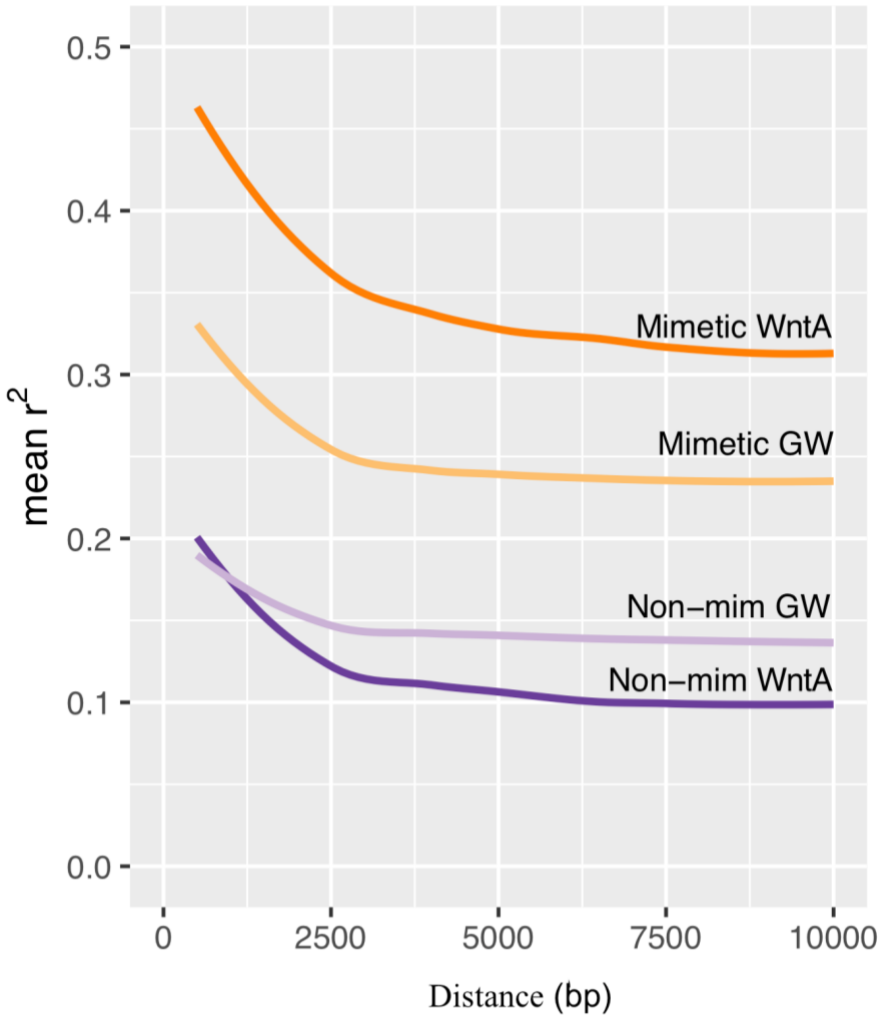

Fig. S13. Decay of linkage disequilibrium (mean r_2) as a function of physical distance across the *WntA* scaffold relative to genome-wide estimates for mimetic subspecies of *L. arthemis* (orange vs. yellow lines) and non-mimetic (dark vs. light purple) *L. a. arthemis* and *L. lorquini/L.weidemeyerii*.

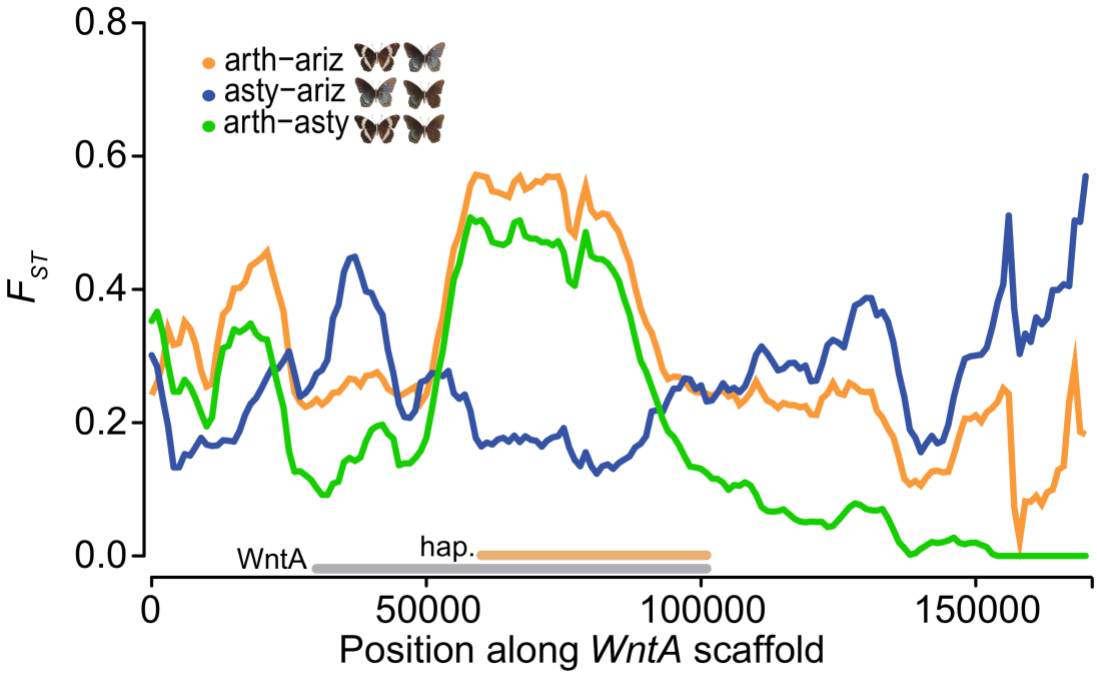

Fig. S14. Sliding window (5kb) mean FST values across the *Limenitis* genomic scaffold housing the *WntA* gene and the upstream region associated with differences in color pattern. Note the high level of divergence centered on the associated haplotype (hap) between mimetic and non-mimetic subspecies of *L. a. arthemis* vs. lower estimated divergence between the two mimetic subspecies (*L. a. astyanax* and *L. a. arizonensis*).

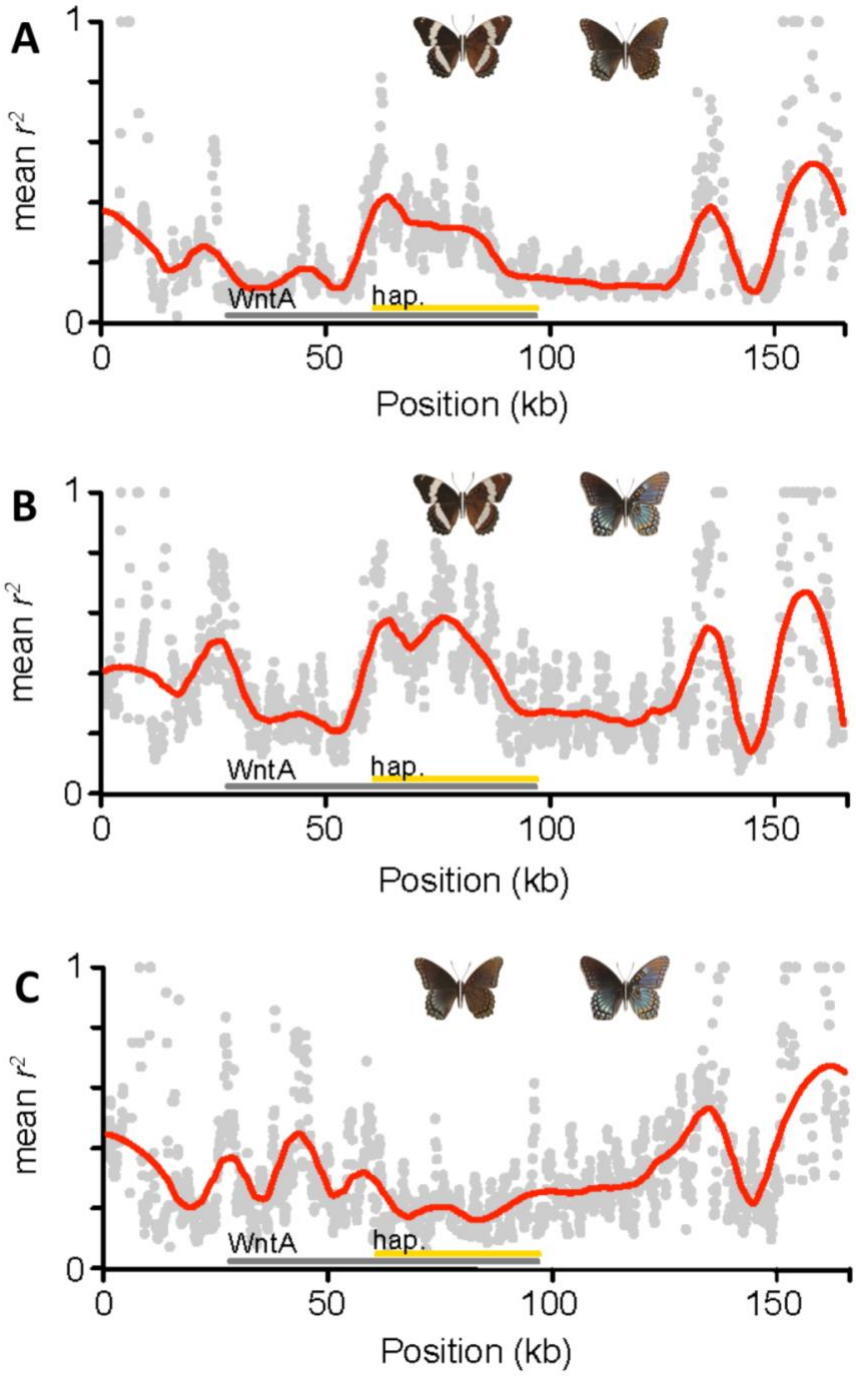

Fig. S15. Mean r2 estimates between all SNPs in 500bp sliding windows (50bp steps) across the *WntA* scaffold for comparisons between A) *L. a. arthemis* and *L. a. astyanax*, B) *L. a. arthemis* and *L. a. arizonensis*, and C) *L. a. astyanax* and *L. a. arizonensis*. Raw points overlaid with a loess best fit line.

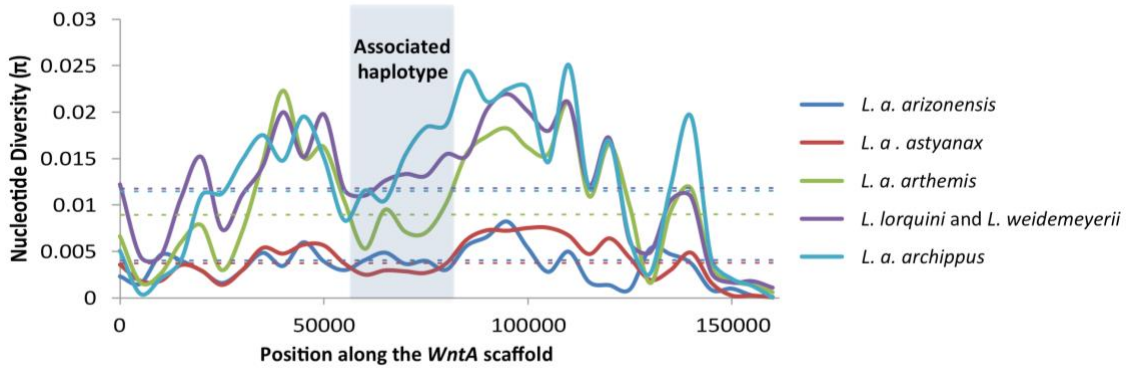

Fig. S16. Levels of nucleotide diversity (π) across the *WntA* scaffold (solid lines). Dashed lines represent genome-wide mean values of π for each *Limenitis* taxon.

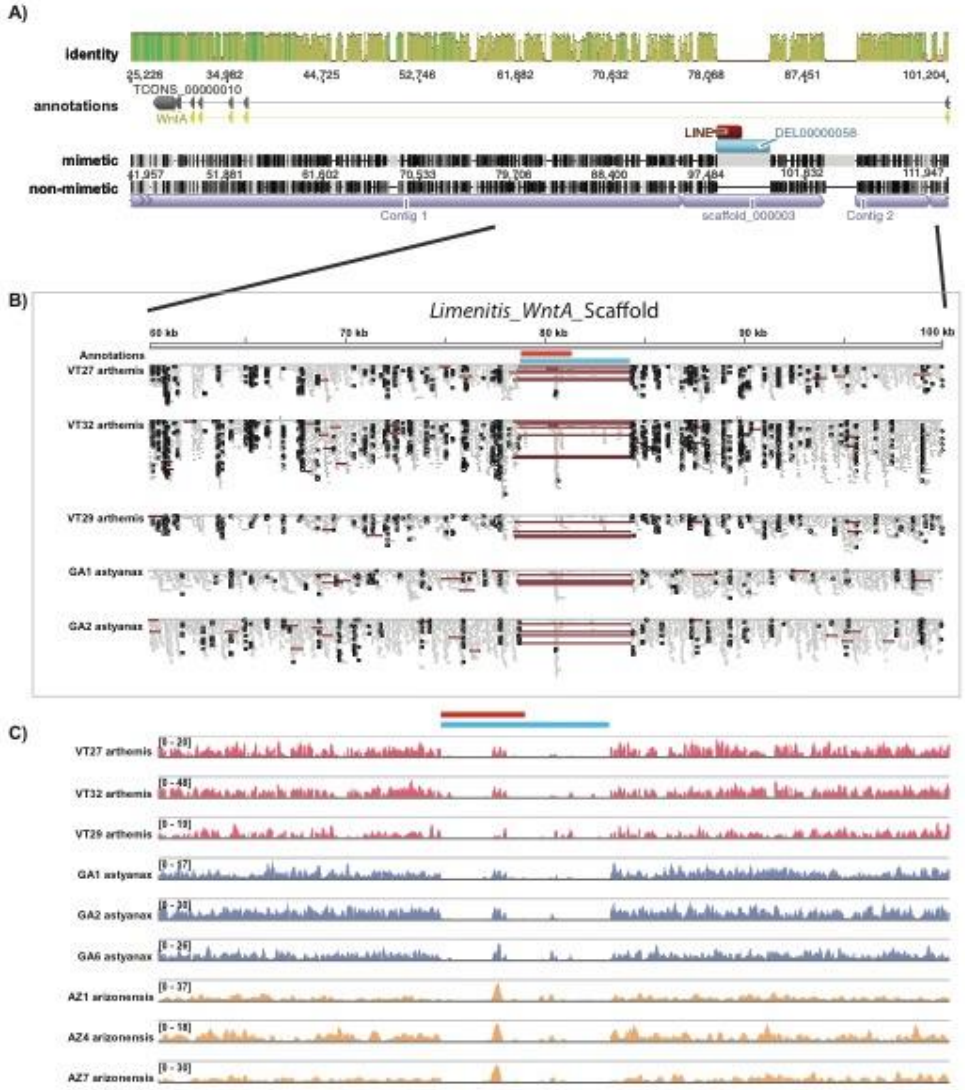

Fig. S17. A) Alignment between *Limenitis_WntA*_scaffold (mimetic) and BAC sequence 60G18 (non-mimetic allele), with the *WntA* mRNA and coding sequence annotations, the LINE element annotation, and the deletion identified by *delly* and *pindel*. **B)** Read pileups of a few sample BAMs across this region. Read pairs highlighted with long red inserts support the presence of the deletion. **C)** Read coverage plots across the region in 3 samples each from *L. a. arthemis*, *L. a. astyanax*, and *L. a. arizonensis*. Annotations follow the same color scheme as in A.

Table S1. Genome sequence data used for the *Limenitis* reference assembly

Insert size	# of libraries	Effective Coverage	Type of Sequencing	Clean Data (GB)
250bp	1	22x	HiSeq PE150	8.8
500bp	1	15x	HiSeq PE150	6.0
800bp	1	12x	HiSeq PE100	4.8
2kb	1	10x	HiSeq PE50	4.0
5kb	1	8x	HiSeq PE50	3.2
10kb	1	5x	HiSeq PE50	2.0
20kb	1	3x	HiSeq PE50	1.2
Total	7	75x		30
BAC Data:	2,124 BACs	0.75x	~20K scaffolds	N50 ~80Kb, max > 1Mb
PacBio Data	~4k mean length	~60X	8.5 million reads	3.8 million corrected reads

Table S2. Summary of *Limenitis* scaffolds mapped and reordered relative to several high quality Lepidopteran reference genomes based on synteny comparisons of reference genome protein-coding genes.

Lepidopteran reference assemblies	Number of mapped *Limenitis* scaffolds	Cumulative size (Mb) of mapped scaffolds
Melitaea cinxia	191 (26%)	286Mb (94%)
Papilio xuthus	203 (26%)	289Mb (94%)
Heliconius melpomene	236 (32%)	294 Mb (96%)

Table S3. List of manually annotated pigmentation genes, including their scaffold position, e-value scores, % identity, top tblastn hit, and Genbank accession number.

Drosophila gene ID	*Limenitis scaffold*	*Scaffold coordinates*	E-value	%Identity	Top Hit (tblastn nr/nt)	GenBank Accession
black	scaffold00082	173102-176137	0.0	87%	AEQ77286.1 putative aspartate decarboxylase [Bicyclus anynana]	MN842725
dopa decarboxylase	scaffold00068	430451-425753	0.0	88%	XM_014514480.1 PREDICTED: Papilio machaon aromatic-L-amino-acid decarboxylase (LOC106719984), transcript variant X1, mRNA	MN842726
dopa decarboxylase-like	scaffold00068	174949-180389	0.0	71%	XM_022271766.1 PREDICTED: Pieris rapae aromatic-L-amino-acid decarboxylase-like (LOC111001764), mRNA	MN842727
ebony	scaffold00058	N/A	0.0	84%	ADU32896.1 ebony [Heliconius melpomene malleti]	MN842728
mfs transporter 1	scaffold00099	1023495-800998	0.0	74%	XP_014365891.1 PREDICTED: uncharacterized protein LOC106716794 [Papilio machaon]	MN842729

mfs transporter 2	scaffold00035	1186912-1163608	0.0	85%	OWR55438.1 monocarboxylate transporter [Danaus plexippus plexippus]	MN842730
mfs transporter 3	scaffold00002	30974194-14053385	0.0	76%	XP_022113441.1 synaptic vesicle glycoprotein 2C-like [Pieris rapae]	MN842731
mfs transporter 4	scaffold00004	44575-47345	0.0	80%	XP_013149032.1 PREDICTED: facilitated trehalose transporter Tret1-like [Papilio polytes]	MN842732
mfs transporter 5	scaffold00021	1282362-1273153	0.0	91%	XP_022131144.1 synaptic vesicle glycoprotein 2B-like [Pieris rapae]	MN842733
mfs transporter 6	scaffold00034	68719-81151	0.0	74%	XP_013167840.1 PREDICTED: synaptic vesicle glycoprotein 2C-like [Papilio xuthus]	MN842734
mfs transporter 7	scaffold00048	1875736-1866542	0.0	82%	OWR48486.1 hypothetical protein KGM_206261 [Danaus plexippus plexippus]	MN842735
pale	scaffold00064	1169549-1177052	0.0	93%	GU063821.1 Heliconius melpomene malleti tyrosine hydroxylase mRNA, complete cds	MN842736
tan	scaffold00110	26802-18196	0.0	89%	GU386341.1 Heliconius melpomene malleti tan mRNA, complete cds	MN842737

yellow	scaffold00035	890753-882089	0.0	84%	GU063822.1Heliconius melpomene yellow mRNA, complete cds	MN842738
yellow-b	scaffold00091	261189-264929	0.0	88%	GU063825.1 Heliconius melpomene yellow-b mRNA, complete cds	MN842739
yellow-c	scaffold00006	3059934-3065112	0.0	87%	GU063827.1 Heliconius erato yellow-c mRNA, complete cds	MN842740
yellow-d	scaffold00155	362595-355796	0.0	73%	GU063831.2 Heliconius melpomene yellow-d mRNA, complete cds	MN842741
yellow-e	scaffold00041	617843-595816	0.0	90%	GU063834.1 Heliconius melpomene yellow-e mRNA, complete cds	MN842742
yellow-f4	no hit	N/A	0.0	70%	GU063836.1 Heliconius melpomene yellow-f4 mRNA, complete cds	MN842743
yellow-h2	scaffold00155	333343-330051	0.0	84%	GU063841.1 Heliconius melpomene yellow-h2 mRNA, complete cds	MN842744
yellow-h3	scaffold00155	326427-323890	0.0	83%	GU063840.1 Heliconius numata yellow-h3 mRNA, complete cds	MN842745
yellow-like	scaffold04833	977-913	0.0	62%	NM_001312559.1 Papilio xuthus protein yellow-like (LOC106126016), mRNA	MN842746
yellow-x	scaffold00070	821271-819946	0.0	82%	GU063844.1 Heliconius melpomene yellow-x mRNA, complete cds	MN842747

ATP-binding cassette subfamily member 4	scaffold00184	188414-223854	0.0	81%	XM_013325974.1 PREDICTED: Papilio xuthus ATP-binding cassette sub-family G member 4 (LOC106127742), mRNA	MN842748
ATP-binding cassette subfamily member 4	scaffold00184	177357-81107	0.0	65%	XM_014505288.1 Select seq XM_014505288.1 PREDICTED: Papilio machaon ATP-binding cassette sub-family G member 4 (LOC106712665), mRNA	MN842749
cardinal	scaffold00001,scaffold00429	4496766-4510638	0.0	70%	XM_022261959.1 PREDICTED: Pieris rapae peroxidase (LOC110994999), mRNA	MN842750
cardinal-like	scaffold00009	2609897-2593099	0.0	91%	XM_022257903.1 PREDICTED: Pieris rapae peroxidase (LOC110992188), mRNA	MN842751
carmine	scaffold00009	2988712-2994766	0.0	98%	XM_013311767.1 PREDICTED: Papilio xuthus AP-3 complex subunit mu-1 (LOC106117444), mRNA	MN842752
carmine-like	scaffold00117	125486-126754	0.0	99%	XM_022269096.1 PREDICTED: Pieris rapae AP-1 complex subunit mu-1 (LOC110999847), mRNA	MN842753

carmine-like	scaffold00091	77296-72454	0.0	100%	XM_022958245.1 PREDICTED: Spodoptera litura AP-2 complex subunit mu (LOC111347862), mRNA	MN842754
carnation	no hit	N/A	0.0	83%	XM_022275362.1 PREDICTED: Pieris rapae vacuolar protein sorting-associated protein 33A (LOC111004360), mRNA	MN842755
cinnabar	scaffold00005	674868-666909	0.0	77%	XM_022262918.1 PREDICTED: Pieris rapae kynurenine 3-monooxygenase (LOC110995656), mRNA	MN842756
deep orange	scaffold00066	469580-475561	0.0	80%	XM_022269462.1 PREDICTED: Pieris rapae vacuolar protein sorting-associated protein 18 homolog (LOC111000113), mRNA	MN842757
garnet	scaffold00088	233713-250091	0.0	74%	XM_022268465.1 Select seq XM_022268465.1 PREDICTED: Pieris rapae AP-3 complex subunit delta (LOC110999428), mRNA	MN842758
henna	scaffold00002	2015877-2020136	0.0	89%	XM_013329500.1 Select seq XM_013329500.1	MN842759

					PREDICTED: Amyelois transitella protein henna (LOC106130608), mRNA	
henna-c	scaffold00218	62321-72039	0	92%	XP_023942067.1 tryptophan 5-hydroxylase 1 [Bicyclus anynana]	MN842760
karmoisin	no hit	N/A	3.00E-173	66%	GQ184571.1 Heliconius melpomene cythera karmoisin (kar) mRNA, partial cds	MN842761
karmoisin-like	no hit	N/A	0.0	62%	XM_013317970.1 PREDICTED: Papilio xuthus monocarboxylate transporter 3 (LOC106122104), transcript variant X3, mRNA	MN842762
kynurenine foramidase	scaffold00009	4168421-4178017	0.0	87%	ACS66705.1 kynurenine formamidase [Heliconius melpomene]	MN842763
light	scaffold00055	1580644-1585196	0.0	81%	XM_022975079.1 PREDICTED: Spodoptera litura vacuolar protein sorting-associated protein 41 homolog (LOC111359501), mRNA	MN842764
optix	scaffold00223	43397-42579	0.0	97%	OWR51623.1 Optix [Danaus plexippus plexippus]	MN842765

orange	scaffold00188	111738-105781	3.00E-136	98%	AK385130.1 Select seq AK385130.1 Bombyx mori mRNA, clone: fcaL52J18_K04259	MN842766
pink	scaffold00023	829270-834680	0.0	67%	XM_022260900.1 PREDICTED: Pieris rapae Hermansky-Pudlak syndrome 5 protein homolog (LOC110994327), transcript variant X1, mRNA	MN842767
ruby	scaffold00057	691062-661244	0.0	80%	XM_022270673.1 Select seq XM_022270673.1 PREDICTED: Pieris rapae AP-3 complex subunit beta-2 (LOC111001009), transcript variant X3, mRNA	MN842768
ruby-like	scaffold00014	345371-356748	0.0	88%	XM_013312195.1 PREDICTED: Papilio xuthus AP-1 complex subunit beta-1 (LOC106117773), transcript variant X2, mRNA	MN842769
scarlet	scaffold00086	855822-890471	0.0	79%	XM_022262816.1 PREDICTED: Pieris rapae protein scarlet (LOC110995590), mRNA	MN842770

scarlet-like	scaffold00017	1756331-1775005	0.0	83%	XM_022259266.1 Select seq XM_022259266.1 PREDICTED: Pieris rapae protein scarlet-like (LOC110993133), transcript variant X1, mRNA	MN842771
sodium-independent sulfate anion transporter-like	scaffold00169	185363-165503	0.0	83%	OWR50240.1 putative Sulfate permease [Danaus plexippus plexippus]	MN842772
vermillion	scaffold00037	245104-224348	0.0	85%	XM_013316171.1 PREDICTED: Papilio xuthus tryptophan 2,3-dioxygenase (LOC106120752), mRNA	MN842773
white	scaffold00086	774623-808987	0.0	88%	XM_014512501.1 PREDICTED: Papilio machaon protein white (LOC106718426), mRNA	MN842774

Table S4. List of *Limenitis* specimens sequenced for population genomic analyses. QC data, alignment rate to the reference assembly, specimen sex if known, and number of SNPs detected is provided for each individual. Note that samples with light grey shading were excluded from downstream analyses due to failure to sequence or poor sequencing quality results.

Species	Sample ID	GPS Coordinates	Filtered R1 reads	Filtered R2 reads	Alignment rate	Sex	SNP sites (UnifiedGenotyper)	Mean depth (UnifiedGenotyper)
L. archippus	IV1	N42.296.36°; W76.230.6°	13209567	13209567	86.79%	F	36507915	9.98
L. archippus	IV2	N42.296.36°; W76.230.6°	17443333	17443333	84.92%	F	36660843	11.25
L. archippus	V1	N25.5719°; W81.2122.2°	23569922	23569922	85.40%	M	37901507	17.88
L. archippus	V2	N25.5719°; W81.2122.2°	31611618	31611618	84.71%	Unknown	38400720	22.89
L. archippus	V3	N25.5719°; W81.2122.2°	14376540	14376540	85.63%	M	34997116	7.07
L. archippus	V4	N28.550.7°; W82.1841.9°	14827201	14827201	83.88%	F	36715721	10.93
L. archippus	V5	N28.550.7°; W82.1841.9°	11731410	11731410	83.42%	M	34923784	7.99

L. archippus	V6	N28.550.7°; W82.1841.9°	15779127	15779127	81.61%	F	37201275	11.69
L. archippus	V9	N28.550.7°; W82.1841.9°	22315749	22315749	84.50%	M	37805632	16.29
L. a. arizonensis	AZ10	N33.5137.6°; W111.4252.8°	15312838	15312838	86.15%	M	38317398	9.54
L. a. arizonensis	AZ11	N33.5137.6°; W111.4252.8°	33434885	33434885	88.86%	M	41632569	21.26
L. a. arizonensis	AZ12	N33.5137.6°; W111.4252.8°	6325868	6325868	88.03%	M	30909219	4.20
L. a. arizonensis	AZ13	N33.5137.6°; W111.4252.8°	14945097	14945097	82.17%	M	39776003	8.91
L. a. arizonensis	AZ1	N34.4155.3°; W112.819.9°	14560113	14560113	89.24%	M	39142814	8.47
L. a. arizonensis	AZ4	N34.4155.3°; W112.819.9°	12621427	12621427	89.38%	M	37358842	7.28
L. a. arizonensis	AZ5	N34.4155.3°; W112.819.9°	16800579	16800579	90.86%	M	26202976	6.50

L. a. arizonensis	AZ6	N34.4155.3°; W112.819.9°	12719306	12719306	86.59%	M	36038455	6.56
L. a. arizonensis	AZ7	N34.4155.3°; W112.819.9°	12197884	12197884	85.20%	M	37824556	7.14
L. a. arizonensis	AZ8	N34.4155.3°; W112.819.9°	12133667	12133667	86.15%	Unknown	38772106	7.40
L. a. arizonensis	AZ9	N34.4155.3°; W112.819.9°	17502456	17502456	85.84%	M	40775494	11.32
L. a. arthemis	VT27	N44.258.2°; W72.5736.0°	18862833	18862833	89.02%	M	41674034	8.63
L. a. arthemis	VT29	N44.258.2°; W72.5736.0°	12082229	12082229	85.95%	M	40582179	5.54
L. a. arthemis	VT32	N44.258.2°; W72.5736.0°	43575855	43575855	88.93%	M	42443349	19.09
L. a. arthemis	VT33	N44.258.2°; W72.5736.0°	8683821	8683821	88.68%	M	37497104	4.25
L. a. arthemis	VT36	N44.258.2°; W72.5736.0°	10733095	10733095	88.72%	M	39390336	5.07
L. a. arthemis	VT38	N44.258.2°; W72.5736.0°				M		

L. a. arthemis	VT44	N44.258.2°; W72.5736.0°	61719923	61719923	87.75%	M	42691967	23.57
L. a. arthemis	VT48	N44.258.2°; W72.5736.0°	26592378	26592378	87.86%	M	41162995	10.46
L. a. arthemis	VT51	N44.258.2°; W72.5736.0°				M		
L. a. arthemis	VT53	N44.258.2°; W72.5736.0°	12302939	12302939	21.02%		17148740	2.28
L. a. arthemis	VT54	N44.258.2°; W72.5736.0°”	25014737	25014737	88.91%	M	40881404	9.88
L. a. arthemis	VT59	N44.258.2°; W72.5736.0°	31069368	31069368	87.41%	M	41277609	11.25
L. a. arthemis	VT63	N44.258.2°; W72.5736.0°	41377214	41377214	88.20%	M	42133253	16.81
L. a. astyanax	GA17	N38.5248°; W83.28538°	12211071	12211071	2.12%	F	3638171	1.31
L. a. astyanax	GA18	N38.5248°; W83.28538°	13426587	13426587	0.20%	F	158496	1.55
L. a. astyanax	GA19	N38.5248°; W83.28538°						
L. a. astyanax	GA1	N38.5248°;	10858801	10858801	89.13%	F	40307415	5.17

		W83.28538°						
L. a. astyanax	GA20	N38.5248°; W83.28538°	40863585	40863585	0.56%		1513352	1.62
L. a. astyanax	GA2	N38.5248°; W83.28538°	29045479	29045479	88.79%	F	41859568	12.75
L. a. astyanax	GA3	N38.5248°; W83.28538°”	12989981	12989981	88.67%	F	39548602	5.83
L. a. astyanax	GA4	N38.5248°; W83.28538°	40565570	40565570	87.66%	F	41039770	13.98
L. a. astyanax	GA5	N38.5248°; W83.28538°				F		
L. a. astyanax	GA6	N38.5248°; W83.28538°	20730951	20730951	88.78%	F	40636452	8.68
L. a. astyanax	GA7	N38.5248°; W83.28538°	11581590	11581590	86.84%	F	33185606	4.77
L. lorquini	RIH2088	N38.17193°; W122.47571°	24153400	24153400	89.80%	M	40970376	16.30
L. lorquini	RIH2089	N38.17193°; W122.47571°	16119903	16119903	88.39%	M	40312870	10.97
L. lorquini	RIH2090	N37.53816°; W121.83892°	22394013	22394013	88.70%	M	41059496	14.64

L. lorquini	RIH2091	N38.15047°; W120.8194°	21956756	21956756	88.44%	M	40511154	14.34
L. lorquini	RIH2092	N38.15047°; W120.8194°	19493893	19493893	88.39%	M	40373644	12.55
L. lorquini	RIH2093	N38.15047°; W120.8194°	41744174	41744174	89.36%	M	42255388	24.64
L. lorquini	RIH2094	N38.15047°; W120.8194°	28107114	28107114	88.93%	M	41331126	17.83
L. lorquini	RIH2095	N38.15047°; W120.8194°	17296996	17296996	88.77%	F	40408302	11.04
L. lorquini	RIH2287	N38.15047°; W120.8194°	21314613	21314613	64.69%	F	40166064	10.2944
L. lorquini	RIH2399	N38.05358°; W119.12797°	14742215	14742215	0.83%	F	953580	1.41298
L. weidemeyerii	RIH2106	N38.05358°; W119.12797°	14523078	14523078	89.34%	M	39969200	9.44177
L. weidemeyerii	RIH2107	N38.11665°; W119.07725	16234914	16234914	88.50%	M	40598100	10.7828
L. weidemeyerii	RIH2108	N38.11665°; W119.07725	21745514	21745514	87.77%	M	40383732	11.6942

L. weidemeyerii	RIH2109	N38.11665°; W119.07725	18342481	18342481	87.81%	M	40439246	11.244
L. weidemeyerii	RIH2110	N38.11665°; W119.07725	26560347	26560347	67.00%	F	40704581	12.412
L. weidemeyerii	RIH2113	N38.11928°; W119.084°	11870098	11870098	88.46%	M	39345271	7.8302
L. weidemeyerii	RIH2114	N38.11928°; W119.084°	13209440	13209440	88.26%	M	39002236	8.1954
L. weidemeyerii	RIH2115	N38.11928°; W119.084°	22382578	22382578	88.38%	M	40792224	13.5487
L. weidemeyerii	RIH2125	N38.11928°; W119.084°	22210040	22210040	88.90%	F	41259840	14.8571
L. weidemeyerii	RIH2214	N38.11928°; W119.084°	24858366	24858366	67.05%	F	40186894	11.2267

Table S5. G-PhoCS data filters

Filter name	Filter description
Scaffold size filtering	scaffold size > 2.16 Mb (N50)
RepeatMasker filtering	masking repetitive elements using RepeatMasker
Tandem Repeats Finder filtering	masking repetitive elements using Tandem Repeats Finder
Phastcons filtering	excluding conserved non-coding and 100 bp flanking regions by blasting against UCSC phastCons elements (phastcons score > 0.8, size > 50 bp) in the 27way alignment for *Drosophila melanogaster*
Genes filtering	excluding exons and 10 kb flanking regions based on the annotation of *Limenitis* v1.0
Read depth filtering	excluding missing calls and calls with read depth twice as high as mean depth in each sample
Non-overlapping filtering	selecting 1 kb blocks at least 50 kb apart

Table S6. Population divergence times and effective population sizes estimated in G-PhoCS analysis (with no migration band).

Population size	Raw estimates x 10_4	Calibrated
*Ne*VT	1.947 (0.2428-4.2557)	16,225 (2,023-35,464)
*Ne*GA	1.4248(0.1737-3.2005)	11,873 (1,448-26,671)
*Ne*AZ	18.433 (17.2402-19.6737)	153,608 (143,668-163,948)
*Ne*lor	21.8078 (12.357-30.8416)	181,732(102,975-257,013)
*Ne*wei	10.5597 (5.5339-15.7064)	87,998 (46,116-130,887)
*Ne*V	66.1956 (61.9482-70.4797)	551,630(516,235-587,331)
*Ne*anc-VT-GA	111.623 (104.4964-118.9812)	930,192(870,803-991,510)
*Ne*anc-VT-GA-AZ	90.2404 (83.0448-97.7229)	752,003(692,040-814358)
*Ne*anc-lor-wei	156.3444 (148.0664-164.8728)	1,302,870(1,233,887-1,373,940)
*Ne*anc-VT-GA-AZ-lor-wei	3.1354 (0.7582-6.5953)	26,128(6,318-54,961)
*Ne*root	184.6762 (178.9891-190.4926)	1,538,968(1,491,576-1,587,438)
Divergence time	Raw estimates x 10_4	Calibrated (yr)
*T*anc-VT-GA	0.0561(0.00727-0.1298)	468(61-1082)
*T*anc-VT-GA-AZ	15.6906(14.8961-16.521)	130,755(124,134-137,675)
*T*anc-lor-wei	1.5651(0.6766-2.4657)	13,043(5,638-20,548)
*T*anc-VT-GA-AZ-lor-wei	30.1344(29.3453-30.9206)	251,120(244,544-257,672)
*T*root	30.1947(29.4008-30.9953)	251,623(245,007-258,294)

GA = *L. a. astyanax*

AZ = *L. a. arizonensis*

Lor = *L. lorquini*

wei = *L. weidemeyerii*

V = *L. archippus*

Table S7. Potential migration bands tested in 16 separate G-PhoCS analyses.

Analysis ID	Labelled migration bands
1	A to B, B to A, B to C, C to B, anc(A, B) to C, C to anc(A, B)
2	C to D, D to C, D to E, E to D, anc(A, B, C) to F, F to anc(A, B, C)
3	anc(A, B) to F, F to anc(A, B), A to C, C to A, B to D, D to B
4	C to E, E to C, D to F, F to D, anc(A, B, C) to anc(D, E), anc(D, E) to anc(A, B, C)
5	A to D, D to A, B to E, E to B, anc(D, E) to F, F to anc(D, E)
6	C to F, F to C, A to E, E to A
7	B to F, F to B, A to F, F to A
8	anc(A, B, C, D, E) to F, F to anc(A, B, C, D, E), A to B, B to A
9	anc(A, B, C, D, E) to F, F to anc(A, B, C, D, E), B to C, C to B
10	anc(A, B, C) to D, D to anc(A, B, C), anc(A, B, C) to E, E to anc(A, B, C), anc(A, B) to F, F to anc(A, B)
11	anc(A, B, C) to D, D to anc(A, B, C), anc(A, B, C) to E, E to anc(A, B, C), C to D, D to C
12	anc(D, E) to F, F to anc(D, E), D to E, E to D, A to C, C to A
13	anc(A, B, C) to anc(D, E), anc(D, E) to anc(A, B, C), B to D, D to B, A to D, D to A
14	C to E, E to C, D to F, F to D, anc(A, B) to C, C to anc(A, B)
15	B to E, E to B, C to F, F to C, A to E, E to A
16	B to F, F to B, A to F, F to A, anc(A, B, C) to F, F to anc(A, B, C)

Candidate migration bands are highlighted in red with a total migration rate above 0.001.
Candidate migration bands are highlighted in blue with a total migration rate between 0.0001 and 0.001.
A denotes *L.a. arthemis*
B denotes *L. a. astyanax*
C denotes *L. a. arizonensis*
D denotes *L. lorquini*
E denotes *L. weidemeyerii*
F denotes *L. archippus*

Table S8. Population divergence times, effective population sizes and migration rates estimated in G-PhoCS full model tests

	Raw estimates x 10_4	Calibrated
*Ne*VT	51.5325 (44.0331-59.1466)	429,438 (366,943-492,888)
*Ne*GA	59.3042 (53.0975-65.0003)	494,202 (442,479-541,669)
*Ne*AZ	20.7625 (19.5545-21.9462)	173,021 (162,954-182,885)
*Ne*Lor	92.8631 (82.8797-102.8867)	773,859 (690,664-857,389)
*Ne*Wei	22.2382 (18.74-25.8287)	185,318 (156,167-215,239)
*Ne*V	65.1493 (61.0082-69.6602)	542,911 (508,402-580,502)
*Ne*anc-VT-GA	1.9928 (0.0908-4.6015)	16,607 (757-38,346)
*Ne*anc-VT-GA-AZ	33.9548 (30.9851-36.967)	282,957 (258,209-308,058)
*Ne*anc-Lor-Wei	2.0929 (0.0761-5.034)	17,441 (634-41,950)
*Ne*anc-VT-GA-AZ-Lor-Wei	2.1575 (0.0733-5.0186)	17,979 (611-41,822)
*Ne*root	198.8724 (192.6313-205.1908)	1,657,270 (1,605,261-1,709,923)
*T*anc-VT-GA	24.1706 (23.5713-24.7199)	201,422 (196,428-205,999)
*T*anc-VT-GA-AZ	24.1717 (23.577-24.7252)	201,431 (196,475-206,043)
*T*anc-Lor-Wei	29.9987 (29.0377-30.8917)	249,989 (241,981-257,431)
*T*anc-VT-GA-AZ-Lor-Wei	30.0154 (29.0408-30.8992)	250,128 (242,007-257,493)
*T*root	30.0166 (29.039-30.8976)	250,138 (241,992-257,480)

migration rate	
VT to GA	27.71% (16.49% - 37.80%)
GA to VT	178.12% (154.53% - 204.42%)
anc-VT-GA-AZ to Lor	0.42% (0.00% - 2.46%)
AZ to Lor	20.98% (18.10% - 23.85%)
Lor to Wei	239.15% (193.25% - 294.64%)
anc-VT-GA-AZ to V	0.04% (0.00% - 0.27%)
AZ to V	11.99% (9.77% - 13.83%)
anc-VT-GA to V	0.02% (0.00% - 0.09%)

GA = *L. a. astyanax*
AZ = *L. a. arizonensis*
Lor = *L. lorquini*
Wei =*L. weidemeyerii*
V = *L. archippus*

Table S9. Pairwise absolute divergence for *Limenitis* taxa. Differences in genome-wide vs. WntA scaffold levels of d_{XY} were evaluated with Wilcoxon signed rank tests (test statistic = W).

		Genome-wide		**WntA Scaffold**			
Pop 1	**Pop 2**	Mean	Standard Error	Mean	Standard Error	W	P-value
arthemis	*astyanax*	0.0116	0.0001	0.0196	0.0019	29320	1.34E-05
arthemis	*arizonensis*	0.0158	0.0002	0.0214	0.0018	27344	7.03E-04
arthemis	*lorquini*	0.0215	0.0002	0.0179	0.0013	14814	6.42E-03
arthemis	*weidemeyerii*	0.0215	0.0002	0.0180	0.0013	15295	1.27E-02
astyanax	*arizonensis*	0.0153	0.0002	0.0137	0.0013	16688	7.01E-02
astyanax	*lorquini*	0.0213	0.0002	0.0245	0.0020	24386	5.18E-02
astyanax	*weidemeyerii*	0.0213	0.0002	0.0245	0.0020	24572	4.18E-02
arizonensis	*lorquini*	0.0224	0.0002	0.0245	0.0019	23102	1.87E-01
arizonensis	*weidemeyerii*	0.0213	0.0002	0.0243	0.0019	24217	6.25E-02
lorquini	*weidemeyerii*	0.0184	0.0002	0.0139	0.0009	11758	2.47E-05

Table S10. Comparison of polymorphism and divergence among *L. arthemis* subspecies revealed 32 SNPs across the *WntA* scaffold (vs. 4 elsewhere in the entire genome) that were fixed in both mimetic subspecies (*L. a. astyanax* and *L. a. arizonensis*) and at a frequency of <0.3 in *L. a. arthemis*. Cells shaded in grey demarcate the approximate location of the associated haplotype region identified by Gallant et al. 2014. Two clusters of fixed SNPs near the start and end of the associated region (highlighted in bold) are the targets of future functional work.

Position on *WntA* scaffold	**Allele in *L.a. astyanax/ arizonensis***	**Allele in *L. a. arthemis***	**Freq(*astyanx* allele in *arthemis*)**
19716	C	A	0.05
53720	A	G	0.05
57119	A	T	0.111
60310	G	A	0.125
60315	**A**	**T**	**0**
60332	**A**	**T**	**0**
60350	**G**	**A**	**0**
60378	**T**	**A**	**0**
60387	**T**	**C**	**0**
60397	**T**	**C**	**0**
60410	**A**	**G**	**0**
64890	A	C	0
64991	T	C	0
68082	C	T	0
68086	T	A	0
69442	T	C	0
69443	T	G	0
74035	C	A	0
74060	A	T	0
74100	A	T	0.1
77280	C	T	0
77285	T	C	0
77299	G	A	0.227
77316	T	G	0
85455	**T**	**A**	**0**
85493	**A**	**G**	**0**
86210	**G**	**A**	**0**
86759	**A**	**T**	**0**
94752	A	C	0.2
96431	T	C	0.15
96455	A	G,T	0
102352	A	T	0.278

References

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, et al. 2000. The genome sequence of Drosophila melanogaster. Science 287:2185–2195.

Ané C, Larget B, Baum DA, Smith SD, Rokas A. 2006. Bayesian estimation of concordance among gene trees. Molecular Biology and Evolution 24: 412-426.

Ahola V, Lehtonen R, Somervuo P, Salmela L, Koskinen P, Rastas P, Välimäki N, Paulin L, Kvist J, Wahlberg N, et al. 2014. The Glanville fritillary genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera. Nature communications 5:4737.

Campbell MS, Holt C, Moore B, Yandell M. 2014. Genome annotation and curation using MAKER and MAKER-P. Current Protocols in Bioinformatics 48:4–11.

Consortium ISG, others. 2008. The genome of a lepidopteran model insect, the silkworm Bombyx mori. Insect biochemistry and molecular biology 38:1036–1045.

Davey JW, Chouteau M, Barker SL, Maroja L, Baxter SW, Simpson F, Merrill RM, Joron M, Mallet J, Dasmahapatra KK, et al. 2016. Major improvements to the Heliconius melpomene genome assembly used to confirm 10 chromosome fusion events in 6 million years of butterfly evolution. G3: Genes, Genomes, Genetics 6:695–708.

Durand EY, Patterson N, Reich D, Slatkin M. 2011. Testing for ancient admixture between closely related populations. Molecular Biology and Evolution 28: 2239-2252.

Freedman AH, Gronau I, Schweizer RM, Ortega-Del Vecchyo D, Han E, Silva PM, Galaverni M, Fan Z, Marx P, Lorente-Galdos B, et al. 2014. Genome sequencing highlights the dynamic early history of dogs. PLoS genetics 10:e1004016.

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. Nature communications 5:4817.

Guerrero RF, Hahn MW. 2018. Quantifying the risk of hemiplasy in phylogenetic inference. Proceedings of the National Academy of Sciences 115:12787-12792.

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. Nature biotechnology 29:644.

Gronau I, Hubisz MJ, Gulko B, Danko CG, Siepel A. 2011. Bayesian inference of ancient human demography from individual genome sequences. Nature genetics 43:1031.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nature protocols 8:1494.

Kajitani R, Toshimoto K, Noguchi H, Toyoda A, Ogura Y, Okuno M, Yabana M, Harada M, Nagayasu E, Maruyama H, et al. 2014. Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome research 24:1384–1395.

Kent WJ. 2002. BLAT—the BLAST-like alignment tool. Genome research 12:656–664.

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome research 27:722–736.

Korf I. 2004. Gene finding in novel genomes. BMC bioinformatics 5:59.

Kriventseva EV, Kuznetsov D, Tegenfeldt F, Manni M, Dias R, Simão FA, Zdobnov EM. 2018. OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs. Nucleic acids research 47:D807–D811.

Larget BR, Kotha SK, Dewey CN, Ané C. 2010. BUCKy: gene tree/species tree reconciliation with Bayesian concordance analysis. Bioinformatics 26: 2910-2911.

Liu KJ, Dai J, Truong K, Song Y, Kohn MH, Nakhleh L. 2014. An HMM-based comparative genomic framework for detecting introgression in eukaryotes. PLoS computational biology 10:e1003649.

Martin SH, Van Belleghem, SM. 2017. Exploring evolutionary relationships across the genome using topology weighting. Genetics 206:429-438.

Martin SH, Davey JW, Jiggins CD. 2014. Evaluating the use of ABBA–BABA statistics to locate introgressed loci. *M*olecular Biology and Evolution 32: 244-257.

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. Nature genetics 47:405.

Pryszcz LP, Gabaldón T. 2016. Redundans: an assembly pipeline for highly heterozygous genomes. Nucleic acids research 44:e113–e113.

Rodriguez F, Oliver JL, Marin A, Medina JR s. 1990. The general stochastic model of nucleotide substitution. Journal of theoretical biology 142:485–501.

Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, et al. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61: 539-542.

Shen J, Cong Q, Kinch LN, Borek D, Otwinowski Z, Grishin NV. 2016. Complete genome of Pieris rapae, a resilient alien, a cabbage pest, and a source of anti-cancer proteins. F1000Research 5.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24:637–644.

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 research 18:1979–1990.

Than C, Ruths D, Nakhleh L. 2008. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. BMC bioinformatics 9:322.

Vicoso B, Kaiser VB, Bachtrog D. 2013. Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. Proceedings of the National Academy of Sciences 110:6453–6458.

Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2017. BUSCO applications from quality assessments to gene prediction and phylogenomics. Molecular biology and evolution 35:543–548.

Wen D, Yu Y, Zhu J, Nakhleh L. 2018. Inferring phylogenetic networks using PhyloNet. Systematic Biology 67:735–740.

Zhan S, Merlin C, Boore JL, Reppert SM. 2011. The monarch butterfly genome yields insights into long-distance migration. Cell 147:1171–1185.

Zhang W, Dasmahapatra KK, Mallet J, Moreira GR, Kronforst MR. 2016. Genome-wide introgression among distantly related Heliconius butterfly species. Genome biology 17:25.