

EVOLUTIONARY BIOLOGY

Parallel evolution of Batesian mimicry supergene in two *Papilio* butterflies, *P. polytes* and *P. memnon*

Takuro Iijima,^{1*} Rei Kajitani,^{2*} Shinya Komata,³ Chung-Ping Lin,⁴ Teiji Sota,³ Takehiko Itoh,² Haruhiko Fujiwara^{1†}

Batesian mimicry protects animals from predators when mimics resemble distasteful models. The female-limited Batesian mimicry in *Papilio* butterflies is controlled by a supergene locus switching mimetic and nonmimetic forms. In *Papilio polytes*, recent studies revealed that a highly diversified region (HDR) containing *doublesex* (*dsx*-HDR) constitutes the supergene with dimorphic alleles and is likely maintained by a chromosomal inversion. In the closely related *Papilio memnon*, which exhibits a similar mimicry polymorphism, we performed whole-genome sequence analyses in 11 butterflies, which revealed a nearly identical *dsx*-HDR containing three genes (*dsx*, *Nach-like*, and *UXT*) with dimorphic sequences strictly associated with the mimetic/nonmimetic phenotypes. In addition, expression of these genes, except that of *Nach-like* in female hind wings, showed differences correlated with phenotype. The dimorphic *dsx*-HDR in *P. memnon* is maintained without a chromosomal inversion, suggesting that a separate mechanism causes and maintains allelic divergence in these genes. More abundant accumulation of transposable elements and repetitive sequences in the *dsx*-HDR than in other genomic regions may contribute to the suppression of chromosomal recombination. Gene trees for *Dsx*, *Nach-like*, and *UXT* indicated that mimetic alleles evolved independently in the two *Papilio* species. These results suggest that the genomic region involving the above three genes has repeatedly diverged so that two allelic sequences of this region function as developmental switches for mimicry polymorphism in the two *Papilio* species. The supergene structures revealed here suggest that independent evolutionary processes with different genetic mechanisms have led to parallel evolution of similar female-limited polymorphisms underlying Batesian mimicry in *Papilio* butterflies.

INTRODUCTION

In 1865, Alfred R. Wallace first described a textbook example of a polymorphic, female-limited Batesian mimicry system in two closely related butterflies, *Papilio memnon* and *Papilio polytes* (1, 2). In these butterflies, the mimetic-type females resemble unpalatable models, such as *Atrophaneura polyeuctes* in the case of *P. memnon*, whereas nonmimetic females and males (monomorphic) do not (Fig. 1A). The polymorphic mimicry of these *Papilio* butterflies is known to be controlled by an autosomal “supergene” known as the *H* locus (3–6). In *P. polytes*, recent genomic studies revealed that a 130-kb region containing the *doublesex* (*dsx*) and two other genes constituted the *H* locus with highly diverged mimetic (*H*) and nonmimetic (*h*) alleles (7, 8); the *H*-allele *dsx* functions in switching female phenotypes (8). Notably, the *H* locus corresponded to a region of an inversion polymorphism (8), which would maintain the structural and functional divergence of *dsx* alleles by suppressing recombination, as in many cases of adaptive polymorphisms (9–13). Recently, we found that the dimorphic *dsx* gene is also associated with the mimicry phenotypes (Fig. 1A) in *P. memnon* in Taiwan (14). To understand the evolutionary origin and process for this mimicry polymorphism, we compared the structure of the mimicry locus in *P. memnon* (referred to as the *A* locus with *A/a* alleles in this study) with that of the *H* locus with *H/h* alleles in *P. polytes*. Unexpectedly, we found that the dimorphic genome structure of the *A* locus of *P. memnon* was not associated with a chromosomal inversion and may have evolved independently from the *H* locus in *P. polytes*.

RESULTS

Whole-genome sequencing of mimetic and nonmimetic *P. memnon* samples

To explore the mimicry locus, we first conducted a de novo whole-genome assembly of *P. memnon* using data from a mimetic female (*dsx* genotype AA; ID, mimetic 1) and a nonmimetic female (*aa*; nonmimetic 1) and confirmed the sequences using genome data from additional nine individuals (table S1). The genotypes at the *dsx* locus of these individuals were determined following methods described by Komata *et al.* (14). The total genome size was 232.7 Mb, with a 5.5-Mb scaffold N50 and 11.8-kb contig N50 (table S2). The total number of predicted genes was 12,426, and the mean lengths of coding sequences (CDSs) were 1508 base pairs (bp) (table S3). These values were similar to those of *P. polytes*. The number of ortholog groups among the five *Papilio* species estimated by Proteinortho (identity, >25%; coverage, >50%) is shown in Fig. 1B. The phylogenetic relationships of these species based on whole orthologous genes (Fig. 1C) were consistent with previous results (8, 15). On the basis of the synonymous site divergence (K_s value) data (see Materials and Methods), we also constructed a phylogenetic tree (fig. S1), which shows the same topology as in Fig. 1C.

Association of the *A/a* mimicry alleles with a long highly diversified region near *dsx* in *P. memnon*

Comparing the whole-genome sequences of the mimetic female and nonmimetic female of *P. memnon*, we discovered a highly diversified region (HDR) with less than 90% identity over the 100-kb region near the *dsx* locus on chromosome 25 of *P. memnon* (Fig. 2A), a result similar to that for *P. polytes* (8). The lower percent identity in the HDR (named *dsx*-HDR) indicates structural differences derived from mimetic (*A*) and nonmimetic (*a*) chromosomes. In addition, we found two HDRs on chromosomes 9 and 23 among all autosomes of the mimetic female (mimetic 1). A Haplot plot suggested that chromosomal inversions

Copyright © 2018
The Authors, some
rights reserved;
exclusive licensee
American Association
for the Advancement
of Science. No claim to
original U.S. Government
Works. Distributed
under a Creative
Commons Attribution
NonCommercial
License 4.0 (CC BY-NC).

¹Department of Integrated Biosciences, University of Tokyo, Kashiwa, Japan. ²Department of Biological Information, Tokyo Institute of Technology, Meguro-ku, Japan. ³Department of Zoology, Graduate School of Science, Kyoto University, Kyoto, Japan. ⁴Department of Life Science, National Taiwan Normal University, Taipei, Taiwan.

*These authors contributed equally to this work.

†Corresponding author. Email: haruh@edu.k.u-tokyo.ac.jp

existed for the HDRs on chromosomes 9 and 23 but not for the *dsx*-HDR on chromosome 25 (Fig. 2B). Furthermore, no chromosomal inversion was found around the *dsx*-HDR over 2 Mb (fig. S2, A and B).

To determine the association of heterozygosity in each HDR with the mimicry polymorphism, we conducted read mapping of the whole-genome sequence data against both allelic sequences at the three HDRs from 11 individuals with known *dsx* genotypes [2 AA and 4 Aa mimetic females; 4 nonmimetic (aa) females; 1 AA male] including the two females used for the whole-genome sequence assembly (Fig. 2C and table S1). Before genome sequencing and read mapping, we determined the *dsx* genotype (AA, Aa, and aa) of each sample using a real-time polymerase chain reaction (PCR) method that we have established previously (14). To judge which HDR of the three HDRs was related to the mimicry polymorphism, we conducted read mapping against both allelic sequences at the three HDRs to determine the association of each locus' heterozygous structure with the mimicry phenotype in the 10 individuals (6 mimetic with two AA and four Aa; 4 nonmimetic with aa) (Fig. 2C and table S1). Here, each read from an HDR was mapped to one of the alleles that indicated higher identity, and the resulting coverage depth corresponded to that of specific alleles. The coverage depth for the heterozygous region was assumed to be approximately half of the homozygous peak, and we estimated the zygosity of the HDRs for each individual on the basis of the coverage depth normalized to the homozygous peak.

The coverage depth against the A-type *dsx*-HDR on chromosome 25 corresponded with the *dsx* genotype and was associated with the

mimetic/nonmimetic phenotypes in all 10 females. However, the coverage depth against the inverted-type HDRs on chromosomes 9 and 23 showed an association with the *dsx* genotype in only one (mimetic 1) and two (mimetics 1 and 7) female individuals, respectively (Fig. 2C). For the a-type *dsx*-HDR and non-inverted-type HDRs, similar results were obtained. We further performed read mapping by 1-kb windows in the *dsx*-HDR, excluding repeated and gapped sequences, and found the complete association between coverage depth values and *dsx* genotypes along the entire region for 11 individuals (fig. S3, A and B).

To confirm that only the *dsx*-HDR is associated with the mimicry polymorphism in *P. memnon*, we also calculated genetic divergence (F_{st}) between six mimetic (AA or Aa) and four nonmimetic female (aa) individuals throughout the whole genome based on single-nucleotide polymorphisms (SNPs) for 10-kb windows using resequencing reads (Fig. 2D). As a result, 17 contiguous windows (that is, a 170-kbp region) containing the *dsx* gene on chromosome 25 showed the highest F_{st} values, indicating association with the mimicry polymorphism (Fig. 2, D and E). We also examined linkage disequilibrium (LD) analysis and found that average r^2 values between SNPs in the *dsx*-HDR were significantly larger compared to both adjacent regions of the *dsx*-HDR (fig. S4). SNP pairs whose distances were nearly as long as the *dsx*-HDR (168 kb) had larger r^2 , inferred from the rightmost area in fig. S4. This supports the hypothesis that recombination was strongly suppressed in the *dsx*-HDR. These results indicate that the *dsx*-HDR on chromosome 25 is associated with the Batesian mimicry locus of *P. memnon*.

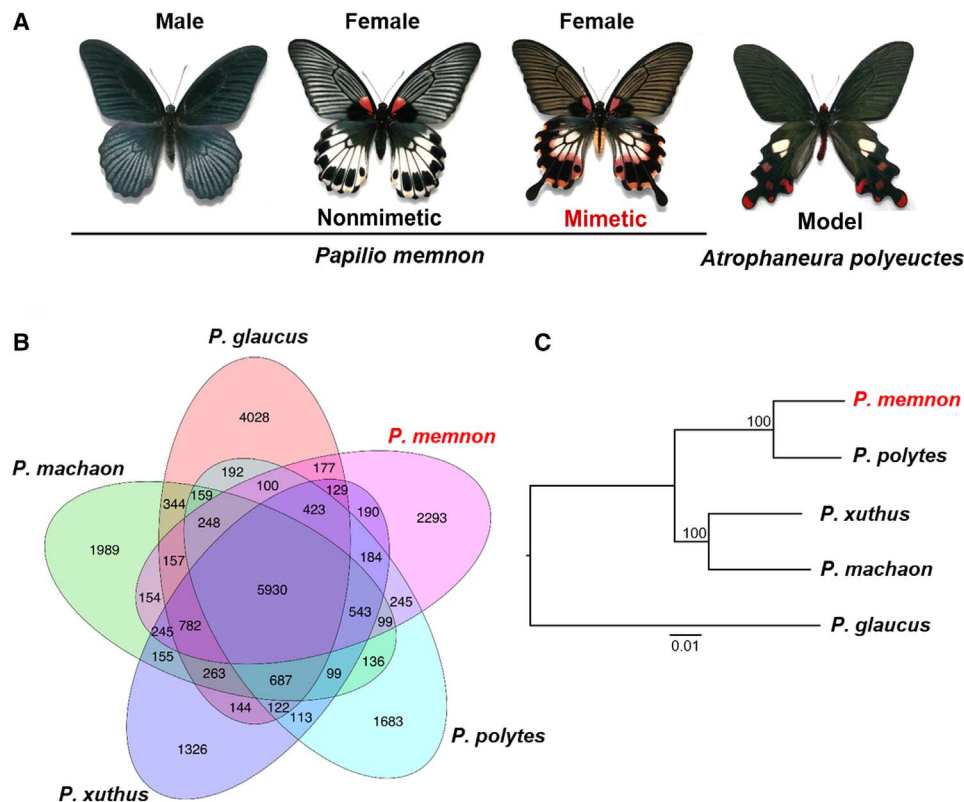


Fig. 1. Phenotypes and genome information for *P. memnon*. (A) *P. memnon* and its model species, *A. polyeuctes*. Males and nonmimetic females (f. *agenor*) of *P. memnon* do not have tails on hind wings, whereas mimetic females (f. *achates*) have tails. (B) Comparison of main genome data among *Papilio* butterflies. (C) Phylogenetic tree of the concatenated protein-coding gene sequences for five *Papilio* butterflies. Multiple alignments of the 4163 ortholog groups were concatenated, and a phylogenetic tree was constructed by using 1,698,871 amino acid sites. Bootstrap values are shown on the branches.

Maintenance of dimorphic *dsx*-HDR in *P. memnon* without chromosomal inversion

The above results indicate that female-limited Batesian mimicry is associated with nearly the same chromosome region in both *Papilio* species, although the structures of these regions differ in terms of the presence/absence of chromosomal inversions (Fig. 3A). The whole length of the *dsx*-HDR for the *A* and *a* alleles was estimated

to be 168 and 150 kb, respectively, both of which were 30 to 40 kb longer than the corresponding regions in *P. polytes* (Fig. 3A and fig. S5). Sequence homology between the chromosomes with *A* and *a* alleles was higher than that between *H* and *h* (Fig. 3A and fig. S6) (8). The *dsx*-HDR of *P. memnon* has boundaries at the specific sites upstream of the newly characterized gene *sodium channel-like* (*Nach-like*) and downstream of *dsx* (Fig. 3, A and B). We refer to these

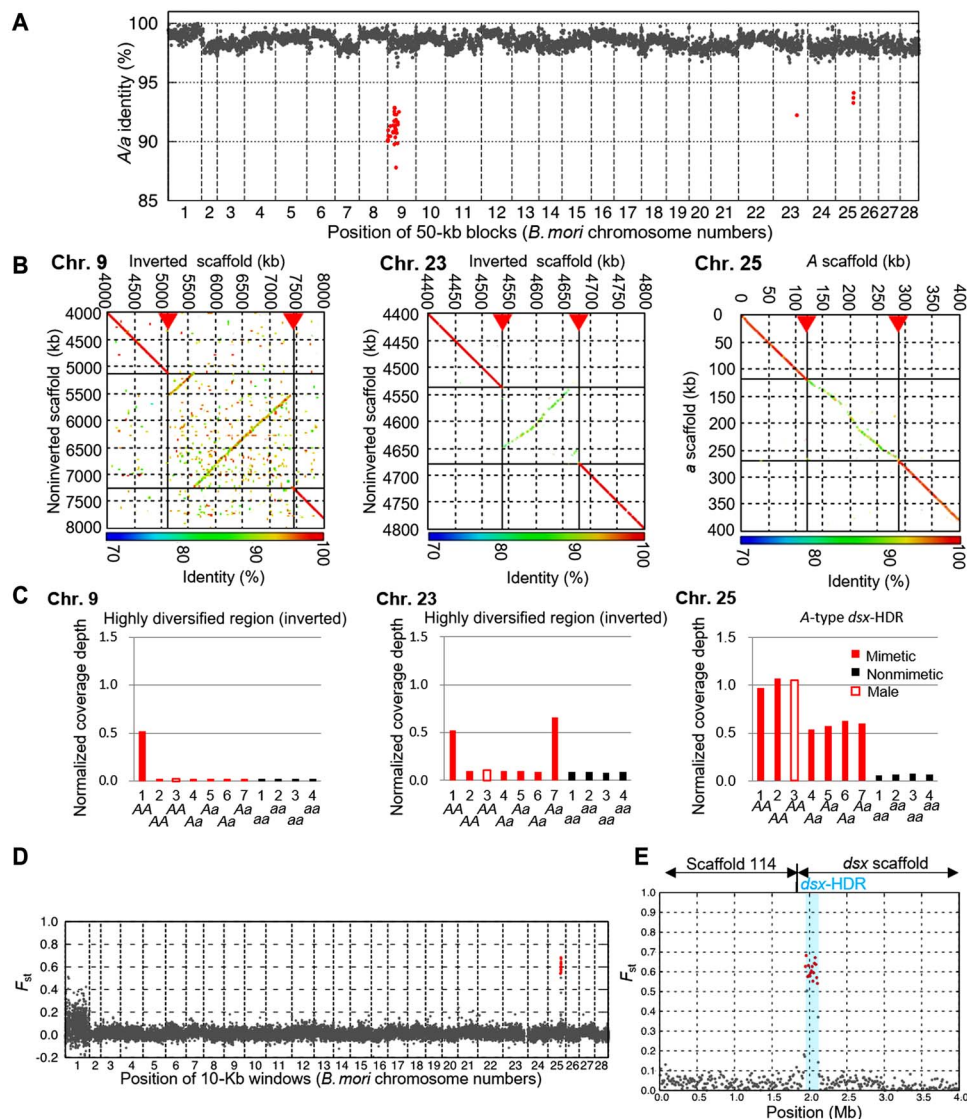


Fig. 2. A long *dsx*-HDR is associated with the mimicry trait. (A) Comparison of the mimetic and nonmimetic genome sequences. For each 50-kb segment in the mimetic draft genome, sequence identity to the nonmimetic genome (vertical axis) and mapped position to the chromosomes of *B. mori* (horizontal axis) were calculated (see Materials and Methods) and plotted as a point. Red points correspond to segments whose identity is <95%. The sequence data were obtained from a nonmimetic female (nonmimetic 1, genotype: *aa*) and a mimetic female (mimetic 1, genotype: *AA*). (B) Harr plot for the three HDRs on chromosomes 9 (scaffold position, 4000 to 8000 kb), 23 (scaffold position, 4400 to 4800 kb), and 25 (scaffold position, 0 to 400 kb). Horizontal axis, allele *a*; vertical axis, allele *A*. The color of each dot indicates the homology level between the sequences of the *a* and *A* alleles as shown at the bottom of each figure. The HDRs on chromosomes 9, 23, and 25 (*dsx*-HDR) are shown with red arrowheads. (C) Estimation of the zygosity of the *A*-type *dsx*-HDR and inverted-type HDRs for 11 butterflies based on the normalized coverage depths. The coverage depth of HDR in chromosomes 9, 23, and 25 of 11 butterflies was normalized by the value corresponding to the homozygous regions (see Materials and Methods). Whole-genome sequencing of 11 butterflies revealed that only the *dsx*-HDR in chromosome 25 was completely associated with mimicry phenotypes. Individuals used: four nonmimetic females (nonmimetic 1 to 4, genotype: *aa*), six mimetic females (mimetic 1 to 2, genotype: *AA* and mimetic 4 to 7, genotype: *Aa*), and one male (mimetic 3, genotype: *AA*). (D) Genetic divergence (F_{st}) between mimetic and nonmimetic populations. Each point corresponds to a 10-kb window in the *P. memnon* draft genome, and windows with $F_{st} > 0.5$ are shown as red points. All windows are mapped on *B. mori* chromosomes. (E) A 4-Mb region on two *P. memnon* scaffolds that contain the highest- F_{st} windows and *dsx* genes. The order of the two scaffolds is determined based on links of mate pairs and synteny with other *Papilio* species.

specific sites at both ends of the *P. memnon* *dsx*-HDR as the “boundary site (BS).” The physical coverage depth around both BSs analyzed by mate-pair mapping showed the same structure as the de novo assembled genome sequence, reinforcing the idea that the *dsx*-HDR structure lacks a chromosomal inversion in *P. memnon* (fig. S7). Furthermore, we used PCR amplification to detect inversions using primer sets for all sequences across BSs of both the *A*-type and *a*-type *dsx*-HDR using 45 butterflies collected in the field. We obtained no evidence for an inverted orientation across both BSs (fig. S8, A to C). From the sequence alignment comparison between *A* and *a* chromosomes, we mapped the left-side BS within the 5′ untranslated region (5′UTR) of *ubiquitously expressed transcript (UXT)* and the right-side BS within the seventh intron of *Nach-like* (Fig. 3B and fig. S9, A and B). Notably, the left BS of the *dsx*-HDR of *P. memnon* (Fig. 3B, turquoise line) coincided precisely with the left breakpoint (BP) of the chromosomal inversion in the *dsx*-HDR of the *H* chromosome of *P. polytes* (Fig. 3B, red line, and fig. S9A). In contrast, the right BS of the *dsx*-HDR

in *P. memnon* was located approximately 8 kb from the right BP in *P. polytes*. A long noncoding RNA gene *U3X* found in the *dsx*-HDR of the *P. polytes* *H* chromosome (8) was not detected in the *P. memnon* genome based on gene annotation and RNA sequencing (RNA-seq) analyses. Transposable elements, especially LINEs (long interspersed nuclear elements) and unclassified repetitive sequences, accumulated more abundantly in the *dsx*-HDR than in other regions (Fig. 3C) or near both BSs (fig. S10). This finding suggests the possibility that large indels and sequence variation caused by repetitive sequences between the *A* and *a* chromosomes reduce recombination in the *dsx*-HDR of *P. memnon*, in contrast to the reduced recombination between the *H* and *h* chromosomes by chromosomal inversion in *P. polytes*.

Dimorphic structure and expression of three genes in *dsx*-HDR

We clarified the sequence divergences at the three genes within the *dsx*-HDR between *A* and *a* in *P. memnon*. There were only four amino

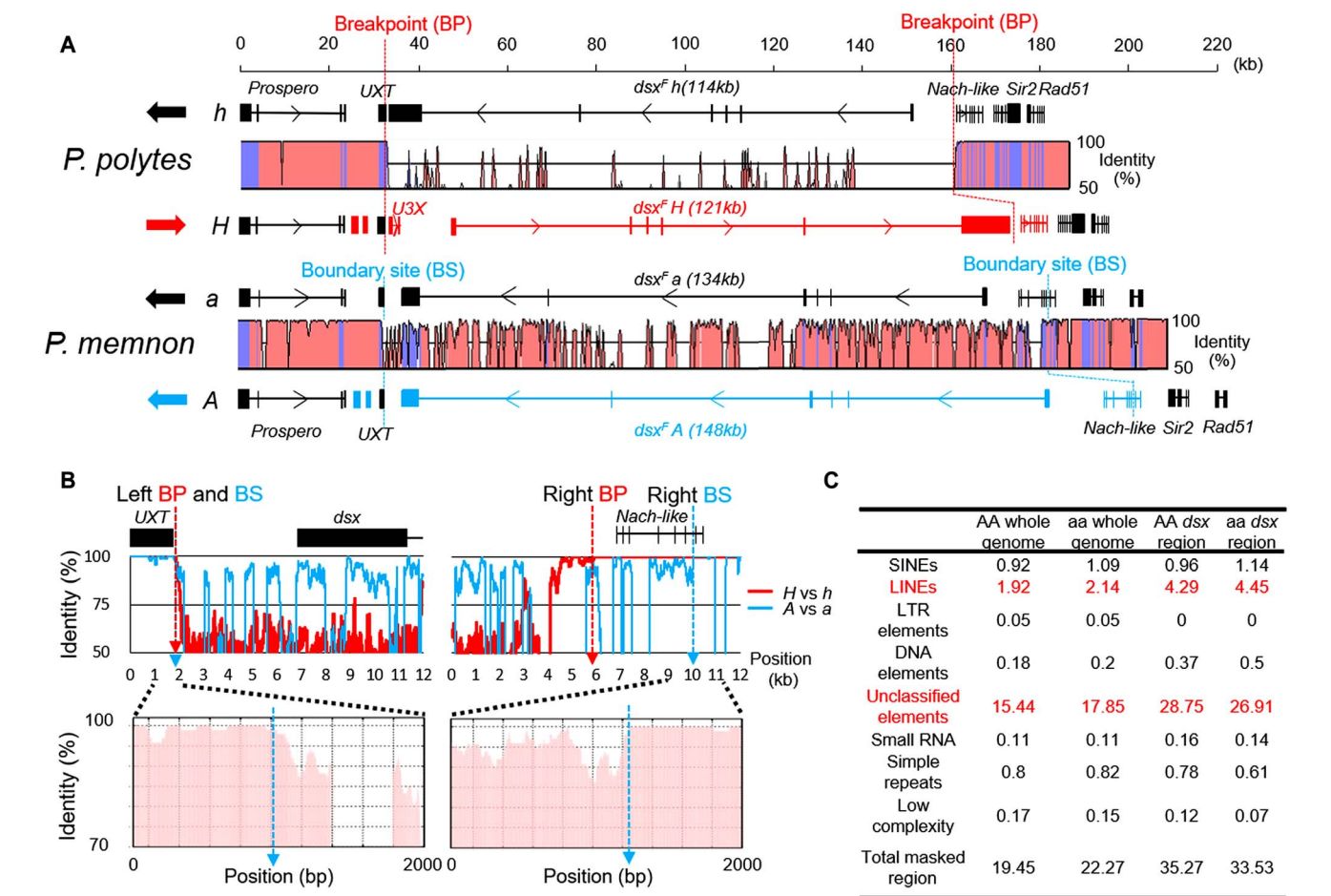


Fig. 3. Detailed structure of *dsx*-HDR in *P. memnon*. (A) Comparison of the detailed structure of *dsx*-HDR on chromosome 25 between *H* and *h* in *P. polytes* and between *A* and *a* in *P. memnon*. The direction of the *dsx*-HDR is reversed between *h* and *H* of *P. polytes*, but is the same between *a* and *A* of *P. memnon*. Putative BPs in *P. polytes* or BSs in *P. memnon* between the heterozygous and homozygous regions are indicated by red or turquoise dotted lines, respectively. Graphical overview of the homology between heterozygous regions is also shown. Mimetic (*H/A*) and nonmimetic (*h/a*) sequences were aligned using the LAGAN program and visualized with VISTA (see Materials and Methods). Exon regions are shown in blue. (B) Marked change in the *A/a* identity across the BS from the homozygous to heterozygous region. The upper panels represent zoom of regions around right and left BPs (*P. polytes*) or BSs (*P. memnon*) in (A). The locations of *UXT*, *dsx*, and *Nach-like* are also shown. Vertical lines (bold or thin) indicate the exon region in each gene, and horizontal lines between exons indicate introns. Red and turquoise dotted lines indicate the BSs in *P. memnon* and the BPs in *P. polytes*, respectively. The lower panels show 2-kb sequence comparisons between *a* and *A* of *P. memnon* in the upper panels. (C) Percentages of the various types of repetitive sequences in the whole genome and within *dsx*-HDR. The proportion of repetitive sequences in scaffolds longer than 500 bp was calculated by using RepeatMasker (see Materials and Methods).

acid substitutions between the *A* and *a* sequences of *P. memnon* *Dsx*, which are mostly located outside the functional domains (with the exception of one residue in the first position of oligomerization domain). These substitutions were completely different from the 14 amino acid substitutions found in *P. polytes* *Dsx*, suggesting an independent evolution of the dimorphic *Dsx* structures between the two species (Fig. 4A and fig. S11A) (14). The expression level of *dsx* of *A* (*dsx_A*), compared with *dsx_a*, in the hind wings of *Aa* females was higher at the earlier pupal stage (P2) (Fig. 4D) but similar at a later stage (P7) (Fig. 4E). However, expression levels of both *dsx_A* and *dsx_a* in the hind wings of *Aa* males were severely repressed (Fig. 4E). These expression patterns of dimorphic *dsx* were also observed in *P. polytes* (8), suggesting that mimetic *dsx* (*dsx_A* and *dsx_H*) regulates female-limited Batesian mimicry through a similar mechanism in both species.

Comparison of the *Nach-like* sequence between *A* and *a* revealed 10 residue changes in the amino acid sequences (Fig. 4A and fig. S11B). The lepidopteran *Nach-like* gene located near *dsx* is homologous to *Drosophila pickpocket 5* (*ppk5*) (Fig. 4B), and it is reported that *Nach* (*ppk4*) colocalizes with *Dsx* in the genital tract nerve (16, 17). Because two amino acid changes were located within the thumb domain, which is important for *ppk* function (Fig. 4A) (18), the *Nach-like* gene function appeared to have a different function between the *A* and *a* alleles and to be involved in the dimorphic mimetic traits. Unexpectedly, we also found seven amino acid substitutions between *H* and *h* *Nach-like* in *P. polytes*, although the gene is located outside of the inverted region (Fig. 4A). The expression of *Nach-like* was detected in the testes and ovaries but not in the hind wings at least during the pupal stage (Fig. 4D).

UXT is a transcriptional regulator conserved among a wide variety of animals and suggested to be involved in many pathways, such as nuclear factor κ B (NF- κ B) and Notch signaling (19, 20). There was no difference in the amino acid sequence of *UXT* outside the *dsx*-HDR between *A* and *a*, but nucleotide sequence differences between the two alleles did exist in a 50-bp region at the most upstream region of the 5'UTR of *UXT* involved in the *dsx*-HDR (Fig. 3B and fig. S9A). Notably, the expression level of *A-UXT* was higher than that of *a-UXT* in day 2 pupae (P2), as in *P. polytes* *UXT*, but it was similar to that of *a-UXT* in day 7 pupae (P7) (Fig. 4, D and E).

To show that the divergence of the above genes is strictly associated with the *dsx*-HDR, we compared F_{st} for gene regions within or adjacent to the *dsx*-HDR between mimetic and nonmimetic females (fig. S12). We found that gene regions within the *dsx*-HDR (*UXT* 5'UTR, *dsx* CDS, and *Nach-like* CDS) showed considerably elevated F_{st} values compared to gene regions outside the *dsx*-HDR (*UXT* CDS and two others).

To reconstruct the evolutionary process contributing to divergence of the *dsx*-HDR in the two *Papilio* species, we constructed the gene trees of *Dsx*, *UXT*, and *Nach-like* sequences (Fig. 4B). All trees showed a similar topology and indicated that the dimorphic structures (*A/a* or *H/h*) of the three genes have differentiated as a supergene unit after the branching of *P. polytes* and *P. memnon*. The phylogenetic tree for the whole *dsx*-HDR region (Fig. 4C; only reliably aligned portions were included for the analysis) indicated that both allelic sequences of *P. memnon* were similar to the *P. polytes* *h* sequence. These phylogenetic patterns and the absence of shared amino acid substitutions in the mimicry loci between the two species suggest independent divergence of alleles in the two *Papilio* species.

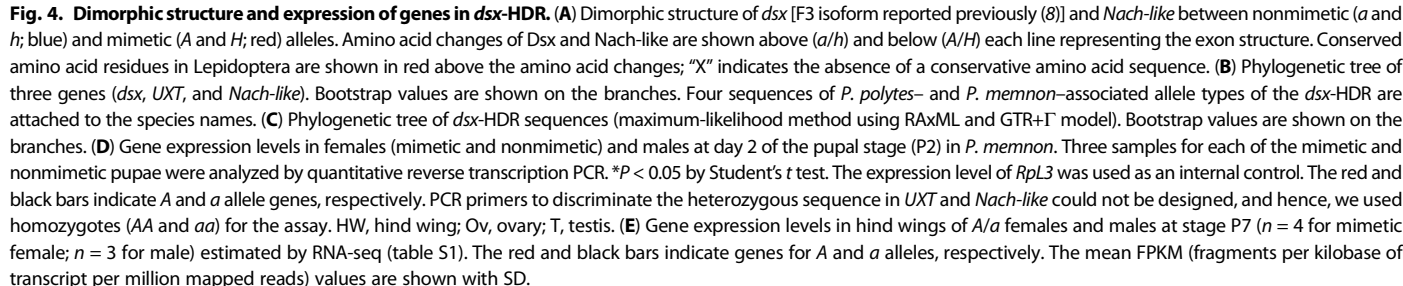
DISCUSSION

Our analyses reveal that the mimicry locus was located at almost the same genomic region in *P. polytes* and *P. memnon*, and contained *dsx*

and the other two genes in both species. However, unlike *P. polytes*, the mimicry locus in *P. memnon* was not contained within a chromosomal inversion. We also found that this *dsx*-HDR showed a strong LD and a sign of recombination suppression. *P. polytes* and *P. memnon* are members of the subgenus *Menelaides* (about 10 species) within the genus *Papilio* (205 species in total) and hence closely related, but are not sister species (15, 21). Because the two species exhibit the same type of mimicry polymorphism, we hypothesized that a chromosomal inversion occurred before the branching of these species, which led to the evolution of the female-limited polymorphism because of recombination repression. As a consequence, this ancestral polymorphism would have been retained and further modified in each species. However, our study revealed that the mimetic and nonmimetic alleles have diverged independently in the two species, although they occurred in the same genomic region. We also demonstrated that two genes (*UXT* and *Nach-like*) showed sequence divergence associated with that of the *dsx* gene and that differential gene expression in pupal hind wings occurred for *UXT* and *dsx*. These observations suggest that these three genes constitute a mimicry supergene. However, we could not detect differential expression of *Nach-like* alleles in pupal hind wings, and the functions of *UXT* and *Nach-like* in determining female phenotypes are unknown. Therefore, further studies examining gene functions are needed to determine whether the *dsx*-HDR actually has a supergene structure containing multiple genes controlling the female polymorphism.

There may be two possible pathways for the independent divergence of mimicry alleles in *P. memnon* without inversion polymorphism. One explanation is that the dimorphic HDR may have evolved independently in *P. memnon* due to factors other than chromosomal inversion that suppressed recombination and promoted sequence divergence in and around the *dsx* gene. This idea may be supported by our observation that many transposable elements or repeated sequences accumulated within the *dsx*-HDR of *P. memnon* (Fig. 3C and fig. S10). Genomic data and in situ hybridization in *Drosophila* species have shown that transposable elements generally accumulate in regions of reduced recombination, polymorphic chromosomal inversions, and neo-Y chromosomes (22–24). In addition, some reports have suggested that transposable elements and repetitive sequences are involved in the recombination suppression (25, 26), although further evidences will be necessary to clarify this idea. Thus, those sequences in the *dsx*-HDR may have reduced the recombination and promoted the divergence of the *dsx* and *Nach-like* sequences between *A* and *a*. In this regard, it is noteworthy that all 14 HDRs other than the *dsx*-HDR on chromosome 25 was identified in the comparison between the Z and W chromosomes of *P. polytes* (8), suggesting the structural similarity between the *dsx*-HDR on chromosome 25 and the sex chromosome. The female-limited mimicry polymorphism in *P. dardanus* is caused by a supergene locus containing eight genes (*dsx* is not one of them) also without inversion (27, 28). However, the accumulation of repeated sequences or transposable elements was not reported for this supergene locus. A second explanation is that chromosomal inversion in *P. memnon* might occur in almost the same position as the *P. polytes* HDR, resulting in divergence of *A* and *a* alleles. A later re-inversion then might have occurred at the same portion to erase the inversion polymorphism. This idea may be supported by our observation that the positions of the left BP of *P. polytes* and the left BS of *P. memnon* appeared to be the same.

Among the five *Papilio* species most closely related to *P. polytes*, only three show female-limited mimicry, and only *P. polytes* and *P. memnon* exhibit the mimicry polymorphism (15). To understand the apparently complex evolutionary process underlying this polymorphism, further



6 of 10

Genetic divergence (F_{st}) between mimetic and nonmimetic populations

To detect SNPs, paired-end reads from seven mimetic and four nonmimetic individuals were mapped onto the mimetic (AA) draft genome using BWA (version 0.7.12-r1044) (29) with default parameters. For two HDRs homologous to silkworm chromosomes 9 and 23, only non-inverted haplotype sequences in the draft genome were used as a reference sequence. Note that the *dsx*-HDR sequence in the draft genome was a mimetic (A) type in this analysis. Using GATK (Genome Analysis Toolkit) (30), mapped reads were realigned (“RealignerTargetCreator” and “IndelRealigner” commands), and genotype information was inferred and stored in GVCF format (“HaplotypeCaller” command) for each individual. After all GVCF files were combined (“GenotypeGVCFs” command), SNP information was extracted through “SelectVariants” and “VariantFiltration” commands under the following parameters to filter out low-quality reads: “QD < 2.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0.”

On the basis of the identified SNPs, genetic divergence (F_{st}) between mimetic and nonmimetic populations was calculated using VCFtools (31). We masked sites of the draft genome according to the following two steps: (i) For each site, individuals whose coverage depths were less than 0.25× mean coverage depth or more than 2.0× mean coverage depth were excluded, and (ii) sites were masked if more than half of the individuals were excluded in both the mimetic and nonmimetic populations. Note that coverage depths were calculated based on the “DP” values in GVCF files. The masking information was input into VCFtools as a FASTA-like format using “-mask” option, and mean F_{st} values between mimetic and nonmimetic populations were calculated for each nonoverlapping 10-kb windows (“-fst-window-size 10000 -fst-window-step 10000” options). In the result of VCFtools, the windows that had 10 or more variants (“N_VARIANTS” value ≥ 10) were extracted, and corresponding F_{st} values were used as the final output. The method of mapping for the 10-kb windows to *Bombyx mori* chromosomes is described in the next section.

Mapping *P. memnon* genomic segments to *B. mori* chromosomes

Scaffolds of *P. memnon* were repeat-masked (see the “Masking of repetitive sequences” section) and aligned to the repeat-masked genome of *B. mori*, which was downloaded from KAIKObase (<http://sgp.dna.affrc.go.jp/KAIKObase/>), using Blastn (version 2.2.26+) with the following options: “-task blastn -evalue 1e-5.” Each scaffold was assigned to a certain *B. mori* chromosome on the basis of the sum of alignment scores, and the position on the chromosome was determined on the basis of the average position of the alignments. Finally, the position of a fixed-length segment in the scaffolds was determined by almost the same way of scaffolds above, limiting target of Blastn to a chromosome assigned to its parental scaffold. If a segment had no hit whose *e* value was $\leq 10^{-5}$, then its position was inferred according to the position of its parental scaffold.

Detection of HDRs

To detect highly differentiated mimetic genomic regions from the nonmimetic ones, each raw mimetic scaffold was divided into nonoverlapping 50-kb segments, and these segments were aligned to the raw *Platanus* scaffolds of the nonmimetic (*aa*) individual, using Blat (version 36) (32) with the following options: “-minIdentity=70 -maxIntron=50000.” From raw output of Blat (PSL format), for each segment, the maximum matches hit, whose alignment length was $\geq 25\%$ of its non-N bases, was extracted, and average sequence identity was calculated using the script

“pslScore.pl” downloaded from the University of California, Santa Cruz website (<https://genome.ucsc.edu/FAQ/FAQblat.html>). Next, mimetic 50-kb segments were mapped to *B. mori* chromosomes (see the “Mapping of *P. memnon* genomic segments to *B. mori* chromosomes” section). Finally, chromosome positions and identities against nonmimetic *P. memnon* scaffolds were plotted (Fig. 2A), and low-identity ($\leq 95\%$) segments were detected on chromosomes 9, 23, and 25.

Estimation of the zygosity of HDRs for whole-genome sequencing data

Genomic reads from each sample were mapped to the draft genome, which included both types of HDRs. The coverage depth values corresponding to homologous regions (homo-peak depth) were determined. The average coverage depth was calculated for each HDR, excluding gaps (Ns), repeat-masked regions, and sites whose coverage depth was $\geq 2\times$ homo-peak depth. The average values were divided by the homo-peak depth (Fig. 2C). Normalized depths of 0, 0.5, and 1 corresponded to missing, heterologous, and homologous regions, respectively.

Homology between heterozygous regions of *a*, *A*, *h*, *H*, and *Papilio xuthus*

A graphical overview of homology between heterozygous regions [mimetic (*H/A*), nonmimetic (*h/a*), and *Papilio xuthus* sequences] was constructed using the LAGAN (33) program and visualized with VISTA (34). The graph is a plot of nucleotide identity with a 70% identity threshold and a 100-bp sliding window centered at a given position. Annotated genes are shown in blue lines.

PCR validation for right and left BS sequences in *dsx*-HDR of *P. memnon*

We designed allele-specific PCR primer sets to amplify the left and right BS junctions. For example, to amplify the left BS for the *A* allele, we designed three primer sets (forward and reverse) for the first PCR (fig. S8A), second PCR (fig. S8B), and third PCR (fig. S8C). Using the PCR sample in which the DNA band was amplified by the first PCR, we further performed the second PCR with the different primer set to amplify the inner region within the first PCR-amplified region. Using this dual checking by two PCR analyses, we could verify the junction direction and sequence for the BS. In addition, we performed the third PCR with a different primer set to amplify the hypothetical inverted *A* region. Similarly, we designed two primer sets at the other three BSs (left BS of *a*, right BS of *A*, and right BS of *a*). Thus, we confirmed that all DNA samples (45 individuals) have the anticipated junction sites and direction.

The following PCR primers and conditions were used for PCR validation: first PCR: 35 cycles of 98°C for 10 s, 60°C for 15 s, and 68°C for 30 s, using primers for *a*-type (left BS of *a**F, left BS of *a**R, right BS of *a**F, and right BS of *a**R) and *A*-type (left BS of *A**F, left BS of *A**R, right BS of *A**F, and right BS of *A**R) alleles. Left BS of *a**F, 5'-TTGATCCTGCACACGAAGAC-3'; left BS of *a**R, 5'-AGATGACGCTGCAAAGACAA-3'; right BS of *a**F, 5'-AGCTGGCATAGTGCACCTAAA-3'; right BS of *a**R, 5'-ACAAGGTAGCGACCGTCATC-3'; left BS of *A**F, 5'-TTTGAAACCATCCTGATGAGC-3'; left BS of *A**R, 5'-TGGATAACCAAAACGAAATGA-3'; right BS of *A**F, 5'-GCCTAGGGTTAAATTTTCAGG-3'; right BS of *A**R, 5'-ACAAGGTAGCGACCGTCATC-3'. Second PCR: 35 cycles of 98°C for 10 s, 60°C for 15 s, and 68°C for 60 s, using primers for *a*-type (left BS of *a**2F, left BS of *a**2R, right BS of *a**2F, and right BS of *a**2R) and *A*-type (left BS of

A*2F, left BS of A*2R, right BS of A*2F, and right BS of A*2R) alleles. Left BS of *a**2F, 5'-ACTGCAGCCTCCATGAAAAA-3'; left BS of *a**2R, 5'-TTTCATTCACCTGCCGTTTTG-3'; right BS of *a**2, 5'-CT-CTTCCGCGACGAACATA-3'; right BS of *a**2R, 5'-GTTCGATTG-CTTTCGTGCTT-3'; left BS of A*2F, 5'-CTGACCGATGCCAAATAA-CA-3'; left BS of A*2R, 5'-GCGAACACAATCAATGAGGA-3'; right BS of A*2F, 5'-CCGCGACGAATATAAGAAGG-3'; right BS of A*2R, 5'-GTCGATTGCTTTTCGTGCTT-3'. Third PCR: 35 cycles of 98°C for 10 s, 60°C for 15 s, and 68°C for 60 s, using primers for hypothetical inverted A-type (left BS of A*3F, left BS of A*3R, right BS of A*3F, and right BS of A*3R) allele. Left BS of A*3F, 5'-TTTGAAACCATCCTGATGAGC-3'; left BS of A*3R, 5'-TCATTTCGTTTTTGGTTATCCA-3'; right BS of A*3F, 5'-CCTTCTTATATTCGTGCGCG-3'; right BS of A*3R, 5'-ACAAGGTAGCGACCGTCATC-3'.

De novo assembly of the *P. memnon* genome

The two draft genomes of *P. memnon* (mimetic and nonmimetic individuals) were assembled separately from Illumina reads using Platanus (version 1.2.4) (35). Before Platanus assembly, all reads were preprocessed to exclude adaptor sequences and low-quality bases by Platanus_trim (version 1.0.7; <http://platanus.bio.titech.ac.jp/>). Short-insert reads (estimated insert size $\leq 0.5 \times$ nominal size) and PCR duplicates in mate-pair libraries were removed based on mapping of reads to the assembly results (scaffolds), which were constructed only from paired-end libraries, using an in-house program. All of the above tools were executed with default settings. As a result, the total size, scaffold N50, and contig N50 were 232.1 and 5.5 Mb and 11.8 kb for the mimetic individual, whereas those were 240.7 and 4.6 Mb and 9.2 kb for the nonmimetic individual, respectively (statistics were calculated for scaffolds whose lengths were ≥ 500 bp). The two draft genomes were analyzed to detect the highly differentiated regions between mimetic and nonmimetic individuals (see the “Detection of HDRs” section).

We further constructed the representative draft genome of *P. memnon* on the basis of the mimetic-type draft genome with the following additional procedures:

- 1) The complete mitochondrial genome was assembled by Platanus with the high-coverage data setting (Platanus assemble -n 100 -k 100), and the resulting complete genome replaced the partial mitochondrial fragments in the initial assembly results.

- 2) Short scaffolds (length, <500 bp) and contamination candidates that matched bacterial or viral genomes (database, National Center for Biotechnology Information Bacteria and RefSeq viral genomes; tool, Blastn; threshold identity, $\geq 90\%$; and query coverage, $\geq 50\%$) were excluded.

- 3) HDRs that mapped to chromosomes 9 and 23 of the silkworm genome (see the “Detection of HDRs” section) were assembled independently with Platanus using a setting to assemble heterozygous haplotype sequences (assemble -u 0, scaffold -u 0), and the scaffolds were extended manually. We constructed haplotype sequences, which contained inversions, for both regions of chromosomes 9 and 23. The resulting scaffolds of those haplotypes replaced the partial sequences corresponding to the HDRs.

The final draft genome (table S2) was used for gene annotation and downstream analysis.

Masking of repetitive sequences

To construct the repetitive sequence library for *P. memnon*, the mimetic draft genome was input into RepeatModeler (version 1.0.7; www.repeatmasker.org/RepeatModeler.html). RepeatMasker (version

4.0.5; www.repeatmasker.org/) was used to mask repetitive sequences in the draft genome with the resulting library from RepeatModeler.

Annotation of protein-coding genes

The annotation procedure followed methods from a previous study of *P. polytes* and *P. xuthus* genomes (8), and protein-coding genes were annotated on the basis of the final draft genome of *P. memnon*. Briefly, we first applied MAKER (version 2.31) (36) annotation pipeline, which uses the homology-based, transcript-based, and ab initio gene predictions. Its result was further corrected by using RNA-seq information. For homology-based prediction, protein sequences from the following species were used: *P. polytes*, *P. xuthus*, *B. mori*, *Danaus plexippus*, *Heliconius melpomene*, *Drosophila melanogaster*, *Apis mellifera*, and *Tribolium castaneum*. RNA-seq-based gene prediction was performed using TopHat (version 2.1.0) (37), Cufflinks (version 2.2.1) (38), Trinity (version 2.2.0) (39), GMAP (version 2016-06-30) (40), and TransDecoder (version 3.0.0; <http://transdecoder.github.io/>). A MAKER gene was replaced if its exons were covered by an RNA-seq-based gene covered MAKER exons (coverage, $\geq 50\%$), and its CDS was longer than that of the MAKER gene.

Orthologs among *Papilio* species

We classified all the proteins from five *Papilio* species (*P. memnon*, *P. polytes*, *P. xuthus*, *P. machaon*, and *P. glaucus*) into orthologous groups using Proteinortho (41). This tool constructs groups on the basis of an all-against-all alignment of Blastp. The definition of an orthologous relationship between proteins was as follows: *e* value, ≤ 10 to 5; identity, $\geq 25\%$; and alignment coverage, $\geq 50\%$ for both sequences. Protein sequences of *P. polytes* and *P. xuthus* were obtained from PapilioBase (<http://papilio.bio.titech.ac.jp/>), and those of *P. machaon* and *P. glaucus* were obtained from RefSeq database.

Construction of a phylogenetic tree of concatenated protein-coding gene sequences

We determined orthologous groups of proteins from five *Papilio* species (see the “Orthologs among *Papilio* species” section) and extracted 4163 groups that had a one-to-one relationship across all species. For each group, multiple alignments were performed by MAFFT (42) and sites containing gaps (‘-’) or ambiguous characters (‘X’) were excluded. All alignments were concatenated, and 1,698,871 amino acid sites were used for phylogenetic analysis. The phylogenetic tree was constructed with RAxML (version 7.0.4) (43). Here, we applied the JTT substitution matrix with a gamma model of rate heterogeneity (-m PROTGAMMAJTT) and the number of 1000 replicates for bootstrap analysis.

Construction of a phylogenetic tree of *dsx*-HDRs including intergenic regions

In addition to *P. memnon*, four draft genomes of *P. polytes*, *P. xuthus*, *P. machaon*, and *P. glaucus* were downloaded from GenBank. Genomic sequence homologous to a region between BPs of the *dsx* locus inversion of *P. polytes* was extracted for each species. The positions of the regions were determined by manual investigations using NUCmer and mummerplot in the MUMmer package (44). For *P. polytes* and *P. memnon*, both mimetic and nonmimetic genomic sequences were obtained. For each extracted sequence, repetitive regions were excluded using RepeatModeler (www.repeatmasker.org/RepeatModeler.html) and RepeatMasker (www.repeatmasker.org/). Next, the repeat-excluded sequences were multiple-aligned by MAFFT (version 7.157b) (42), and

sites containing gaps (‘-’) were further excluded, resulting in 30,683 nucleotide sites. Finally, with the resultant alignment, a maximum-likelihood tree reconstruction was performed using RAxML (version 7.0.4) (43) with the general time-reversible substitution model with a gamma model for rate heterogeneity (GTR+G) and 1000 bootstrap replicates.

Construction of *dsx*, *UXT*, and *Nach-like* gene trees

Nucleotide sequences of three genes (*dsx*, *UXT*, and *Nach-like*) were obtained from KAIKObase (<http://sgp.dna.affrc.go.jp/KAIKObase/>) for *B. mori*, MonarchBase (<http://monarchbase.umassmed.edu>) for *D. plexippus*, Manduca Base (<http://agripestbase.org/manduca/>) for *Manduca sexta*, Butterfly Genome Database for *H. melpomene*, PapilioBase (<http://papilio.bio.titech.ac.jp/>) for *P. polytes* and *P. xuthus*, and RefSeq database for *P. machaon* and *P. glaucus*. In addition, the *ppk* sequences (*ppk1*–6) from *D. melanogaster* (homologous to *Nach-like*) were obtained from FlyBase (<http://flybase.org>). For each gene, multiple alignment was performed using MAFFT (42), resulting in sequence matrices with 816, 635, and 2162 aligned sites for *dsx*, *UXT*, and *Nach-like*, respectively. Maximum-likelihood analysis was performed for each gene using RAxML version 8.2.3 (43) with RAPID ML analysis with 1000 bootstrapping analyses. The general time-reversible substitution model with a gamma model for rate heterogeneity (GTR+G) was used in this analysis.

Reverse transcription PCR

Total RNA was extracted from hind wings, ovary, and testis from each individual using the TRI reagent (Sigma-Aldrich) and the RNeasy Mini Kit (Qiagen) following the manufacturer’s protocol. The samples were treated with 2.5 U of deoxyribonuclease I/1 µg of RNA (Takara) for 15 min at 37°C to remove genomic DNA. After phenol-chloroform extraction, reverse transcription was performed using the Verso cDNA Synthesis Kit (Thermo Fisher Scientific). A StepOne Real-Time PCR System (Applied Biosystems Inc.) with the Power SYBR Green PCR Master Mix (Applied Biosystems) was used for gene quantification. PCR cycling conditions were as follows: 95°C for 10 min, followed by 40 cycles of 95°C for 15 s, and 60°C for 60 s. Standard curves were generated for all primer pairs. The following quantitative PCR primers were used for the gene quantification: *dsx* a*F, 5′-AAAGTGGCGGTT-GAACCAAA-3′; *dsx* a*R, 5′-TTCAATTGTTGCTGAGACCAT-3′; *dsx* A*F, 5′-GAGTTTAGCCGTGTTATAACTCATA-3′; *dsx* A*R, 5′-TTCAAATGTTGCGCCGAGACAAT-3′; *UXT**F, 5′-GATGAAGC-TTCTGGTCGTTATACAC-3′; *UXT**R, 5′-GTCTTCGTGTTTCAG-GATCAACAGTT-3′; *Nach-like**F, 5′-GCCCCAAGTGCGCAAA-3′; *Nach-like**R, 5′-CGTCGCGGAAGAGACATTTTC-3′; *rpL3**F, 5′-CTCACCGTATGGGCAGAACATATGTC-3′; *rpL3**R, 5′-TCT-TGCTGGCTTTGGTAAATGCCTTCTTC-3′. Expression levels of these genes were normalized to that of *ribosomal protein L3* (*RpL3*), which was used as an internal control. Statistical analysis was performed on data from individuals in each pupal stage. Statistical significance was determined using Student’s *t* test.

Calculation of synonymous site divergence (K_s value) for each pair of the five *Papilio* species

The K_s value for each pair of the five *Papilio* species was calculated based on the ortholog groups used for the phylogenetic analysis in the main text (Fig. 1C). For each ortholog group, genes were included one by one for all species, and the K_s value was calculated as follows:

- 1) Protein sequences from all the species were aligned by MAFFT (version 7.157b) (42).
- 2) Aligned protein sequences were converted to nucleotide sequences.

3) Sites that contained gaps (‘-’) were excluded.

- 4) Pairwise K_s values were calculated by KaKs_Calculator (version 1.2) (45) according to the Nei-Gojobori method (-m NG) (46).

K_s values between two species were averaged, excluding ortholog groups for which KaKs_Calculator could not calculate the value (‘NA’ output) due to lack of informative sites. We constructed a neighbor-joining tree method using T-REX web server (47, 48), inputting the average K_s data as a distance matrix.

LD analysis

To investigate LD in the *dsx*-HDR, we calculated r^2 for all pairs of biallelic SNPs in the *dsx*-HDR. The SNPs used were the same as used for F_{st} [see the “Genetic divergence (F_{st}) between mimetic and nonmimetic populations” section], and pairwise r^2 between SNPs was calculated with VCFtools (version 0.1.14) (31) by inputting a VCF file and using the option “-geno-r2.” The resulting r^2 values were binned according to distances of SNP pair (bin size, 1 kb), and average values in bins were plotted in fig. S4. For comparison, adjacent regions of the *dsx*-HDR, whose lengths equaled that of the *dsx*-HDR (168 kb), were also analyzed in the same way.

SUPPLEMENTARY MATERIALS

Supplementary material for this article is available at <http://advances.sciencemag.org/cgi/content/full/4/4/eaao5416/DC1>

- fig. S1. Neighbor-joining tree based on average K_s value matrix.
fig. S2. Harr plot view of the 2-Mb region around *dsx*-HDRs on chromosome 25 between *P. memnon* and *P. polytes*.
fig. S3. The normalized coverage depth of each region (1-kb window) in the *dsx*-HDR.
fig. S4. Relation between LD and genomic distance for *dsx*-HDR.
fig. S5. Comparison of detailed and dimorphic structures of *dsx*-HDR between the *A/a* locus in *P. memnon* and the *H/h* locus in *P. polytes*.
fig. S6. Sequence homology of *dsx*-HDR among *A*, *a*, *H*, *h*, and the *P. xuthus* corresponding regions.
fig. S7. Verification of the BS junction without inversion by mate-pair mapping.
fig. S8. PCR-based verification of junction sites for the left and right BSs in *dsx*-HDR of *P. memnon*.
fig. S9. Nucleotide sequence comparison of the left and right BSs of *dsx*-HDR between the *A/a* locus of *P. memnon* and the *H/h* locus of *P. polytes*.
fig. S10. Repetitive sequences around BPs or BSs in the *A/a* locus.
fig. S11. Amino acid sequence substitutions in *Dsx* and *Nach-like*.
fig. S12. F_{st} between mimetic and nonmimetic *P. memnon* at CDSs or UTRs related to *dsx*-HDR.
table S1. Statistics of sequenced read data.
table S2. Statistics of draft genome.
table S3. Statistics of protein-coding genes.

REFERENCES AND NOTES

1. A. R. Wallace, On the phenomena of variation and geographical distribution as illustrated by the Papilionidae of the Malayan Region. *Trans. Linn. Soc.* **25**, 1–71 (1865).
2. H. W. Bates, Contributions to an insect fauna of the Amazon valley (Lepidoptera: Heliconidae). *Trans. Linn. Soc.* **23**, 495–556 (1862).
3. C. A. Clarke, P. M. Sheppard, The genetics of the mimetic butterfly *Papilio polytes* L. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **263**, 431–458 (1972).
4. C. A. Clarke, P. M. Sheppard, Further studies on the genetics of the mimetic butterfly *Papilio memnon* L. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **263**, 35–70 (1971).
5. C. A. Clarke, P. M. Sheppard, I. W. B. Thornton, The genetics of the mimetic butterfly *Papilio memnon* L. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **254**, 37–89 (1968).
6. C. A. Clarke, P. M. Sheppard, Super-genes and mimicry. *Heredity* **14**, 175–185 (1960).
7. K. Kunte, W. Zhang, A. Tenger-Trolander, D. H. Palmer, A. Martin, R. D. Reed, S. P. Mullen, M. R. Kronforst, *doublesex* is a mimicry supergene. *Nature* **507**, 229–232 (2014).
8. H. Nishikawa, T. Iijima, R. Kajitani, J. Yamaguchi, T. Ando, Y. Suzuki, S. Sugano, A. Fujiyama, S. Kosugi, H. Hirakawa, S. Tabata, K. Ozaki, H. Morimoto, K. Ihara, M. Obara, H. Hori, T. Itoh, H. Fujiwara, A genetic mechanism for female-limited Batesian mimicry in *Papilio* butterfly. *Nat. Genet.* **47**, 405–409 (2015).

9. C. Küpper, M. Stocks, J. E. Risse, N. dos Remedios, L. L. Farrell, S. B. McRae, T. C. Morgan, N. Karlionova, P. Pinchuk, Y. I. Verkuil, A. S. Kitaysky, J. C. Wingfield, T. Piersma, K. Zeng, J. Slate, M. Blaxter, D. B. Lank, T. Burke, A supergene determines highly divergent male reproductive morphs in the ruff. *Nat. Genet.* **48**, 79–83 (2016).
10. S. Lamichhaney, G. Fan, F. Widemo, U. Gunnarsson, D. S. Thalmann, M. P. Hoepfner, S. Kerje, U. Gustafson, C. Shi, H. Zhang, W. Chen, X. Liang, L. Huang, J. Wang, E. Liang, Q. Wu, S. M.-Y. Lee, X. Xu, J. Höglund, X. Liu, L. Andersson, Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nat. Genet.* **48**, 84–88 (2016).
11. M. Joron, L. Frezal, R. T. Jones, N. L. Chamberlain, S. F. Lee, C. R. Haag, A. Whibley, M. Becuwe, S. W. Baxter, L. Ferguson, P. A. Wilkinson, C. Salazar, C. Davidson, R. Clark, M. A. Quail, H. Beasley, R. Glithero, C. Lloyd, S. Sims, M. C. Jones, J. Rogers, C. D. Jiggins, R. H. French-Constant, Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature* **477**, 203–206 (2011).
12. K. Dasmahapatra, J. R. Walters, A. D. Briscoe, J. W. Davey, A. Whibley, N. J. Nadeau, A. V. Zimin, D. S. T. Hughes, L. C. Ferguson, S. H. Martin, C. Salazar, J. J. Lewis, S. Adler, S. J. Ahn, D. A. Baker, S. W. Baxter, N. L. Chamberlain, R. Chauhan, B. A. Counterman, T. Dalmay, L. E. Gilbert, K. Gordon, D. G. Heckel, H. M. Hines, K. J. Hoff, P. W. H. Holland, E. J. Joly, F. M. Jiggins, R. T. Jones, D. D. Kapan, P. Kersey, G. Lamas, D. Lawson, D. Mapleson, L. S. Maroja, A. Martin, S. Moxon, W. J. Palmer, R. Papa, A. Papanicolaou, Y. Pauchet, D. A. Ray, N. Rosser, S. L. Salzberg, M. A. Supple, A. Surridge, A. T. Trolander, H. Vogel, P. A. Wilkinson, D. Wilson, J. A. Yorke, F. Yuan, A. L. Balmuth, C. Eland, K. Gharbi, M. Thomson, R. A. Gibbs, Y. Han, J. C. Jayaseelan, C. Kovar, T. Mathew, D. M. Muzny, F. Onger, L. L. Pu, J. Qu, R. L. Thornton, K. C. Worley, Y. Q. Wu, M. Linares, M. L. Blaxter, R. H. French-Constant, M. Joron, M. R. Kronforst, S. P. Mullen, R. D. Reed, S. E. Scherer, S. Richards, J. Mallet, W. O. McMillan, C. D. Jiggins, Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. *Nature* **487**, 94–98 (2012).
13. T. Schwander, R. Libbrecht, L. Keller, (2014) Supergenes and complex phenotypes. *Curr. Biol.* **24**, R288–R294 (2014).
14. S. Komata, C.-P. Lin, T. Iijima, H. Fujiwara, T. Sota, Identification of *doublesex* alleles associated with the female-limited Batesian mimicry polymorphism in *Papilio memnon*. *Sci. Rep.* **6**, 34782 (2016).
15. E. V. Zakharov, M. S. Caterino, F. A. H. Sperling, Molecular phylogeny, historical biogeography, and divergence time estimates for swallowtail butterflies of the genus *Papilio* (Lepidoptera: Papilionidae). *Syst. Biol.* **53**, 193–215 (2004).
16. C. Rezával, H. H. Pavlou, A. J. Dorman, Y.-B. Chan, E. A. Kravitz, Neural circuitry underlying *Drosophila* female postmating behavioral responses. *Curr. Biol.* **22**, 1155–1165 (2012).
17. C. Rezával, T. Nojima, M. C. Neville, A. C. Lin, S. F. Goodwin, Sexually dimorphic octopaminergic neurons modulate female postmating behaviors in *Drosophila*. *Curr. Biol.* **24**, 725–730 (2014).
18. K. M. Zelle, B. Lu, S. C. Pyfrom, Y. B. Shahar, The genetic architecture of degeneration/epithelial sodium channels in *Drosophila*. *G3* **3**, 441–450 (2013).
19. Y. Zhou, R. Ge, R. Wang, F. Liu, Y. Huang, H. Liu, Y. Hao, Q. Zhou, C. Wang, UXT potentiates angiogenesis by attenuating Notch signaling. *Development* **142**, 774–786 (2015).
20. S.-K. Su, C.-Y. Li, P.-J. Lei, X. Wang, Q.-Y. Zhao, Y. Cai, Z. Wang, L. Li, M. Wu, The EZH1-SUZ12 complex positively regulates the transcription of NF- κ B target genes through interaction with UXT. *J. Cell Sci.* **129**, 2343–2353 (2016).
21. K. Kunte, The diversity and evolution of Batesian mimicry in *Papilio* swallowtail butterflies. *Evolution* **63**, 2707–2716 (2009).
22. P. D. Sniegowski, B. Charlesworth, Transposable element numbers in cosmopolitan inversions from a natural population of *Drosophila melanogaster*. *Genetics* **137**, 815–827 (1994).
23. D. Bachtrog, Accumulation of *Spock* and *Worf*, two novel non-LTR retrotransposons, on the neo-Y chromosome of *Drosophila miranda*. *Mol. Biol. Evol.* **20**, 173–181 (2003).
24. P. Fontanillas, D. L. Hartl, M. Reuter, Genome organization and gene expression shape the transposable element distribution in the *Drosophila melanogaster* euchromatin. *PLOS Genet.* **3**, 2256–2267 (2007).
25. R. Gorelick, Transposable elements suppress recombination in all meiotic eukaryotes, including autotetraploid asexuals: A reply to Schön and Martens. *J. Nat. Hist.* **37**, 903–909 (2003).
26. S.-F. Li, G.-J. Zhang, J.-H. Yuan, C.-L. Deng, W.-J. Gao, Repetitive sequences and epigenetic modification: Inseparable partners play important roles in the evolution of plant sex chromosomes. *Planta* **243**, 1083–1095 (2016).
27. M. J. T. N. Timmermans, S. W. Baxter, R. Clark, D. G. Heckel, H. Vogel, S. Collins, A. Papanicolaou, I. Fukova, M. Joron, M. J. Thompson, C. D. Jiggins, R. H. French-Constant, A. P. Vogler, Comparative genomics of the mimicry switch in *Papilio dardanus*. *Proc. Biol. Sci.* **22**, 281 (2014).
28. H. F. Nijhout, Polymorphic mimicry in *Papilio dardanus*: Mosaic dominance, big effects, and origins. *Evol. Dev.* **5**, 579–592 (2003).
29. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
30. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
31. P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, R. Durbin; 1000 Genomes Project Analysis Group, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
32. W. J. Kent, BLAT—The BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
33. M. Brudno, C. B. Do, G. M. Cooper, M. F. Kim, E. Davydov; NISC Comparative Sequencing Program, E. D. Green, A. Sidow, S. Batzoglou, LAGAN and multi-LAGAN: Efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res.* **13**, 721–731 (2003).
34. C. Mayor, M. Brudno, J. R. Schwartz, A. Poliakov, E. M. Rubin, K. A. Frazer, L. S. Pachter, I. Dubchak, VISTA: Visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* **16**, 1046–1047 (2000).
35. R. Kajitani, K. Toshimoto, H. Noguchi, A. Toyoda, Y. Ogura, M. Okuno, M. Yabana, M. Harada, E. Nagayasu, H. Maruyama, Y. Kohara, A. Fujiyama, T. Hayashi, T. Itoh, Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. *Genome Res.* **24**, 1384–1395 (2014).
36. C. Holt, M. Yandell, MAKER2: An annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491 (2011).
37. D. Kim, G. Pertea, C. Trapnell, H. Pimentel, R. Kelley, S. L. Salzberg, TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
38. C. Trapnell, B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. van Baren, S. L. Salzberg, B. J. Wold, L. Pachter, Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
39. M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
40. T. D. Wu, C. K. Watanabe, GMAP: A genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859–1875 (2005).
41. M. Lechner, S. Findeiß, L. Steiner, M. Marz, P. F. Stadler, S. J. Prohaska, Proteinortho: Detection of (co-)orthologs in large-scale analysis. *BMC Bioinformatics* **12**, 124 (2011).
42. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
43. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
44. S. Kurtz, A. Phillippy, A. L. Delcher, M. Smoot, M. Shumway, C. Antonescu, S. L. Salzberg, Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).
45. Z. Zhang, J. Li, X. Q. Zhao, J. Wang, G. K.-S. Wong, J. Yu, KaKs_Calculator: Calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics* **4**, 259–263 (2006).
46. M. Nei, T. Gojobori, Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**, 418–426 (1986).
47. N. Saitou, M. Nei, The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425 (1987).
48. A. Boc, A. B. Diallo, V. Makarenkov, T-REX: A web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic Acids Res.* **40**, W573–W579 (2012).

Acknowledgments: We thank Y. KonDo, T. Kitamura, and S. Kawamura for helpful suggestions on our research. We also thank T. Kojima for helpful comments on the manuscript. **Funding:** This work was supported by Ministry of Education, Culture, Sports, Science and Technology/ Japan Society for the Promotion of Science KAKENHI (20017007, 22128005, and 15H05778 to H.F.; 15K14603 to T.S.; 16J05495 to T. Iijima) and Fujiwara Natural History Public Interest Incorporated Foundation (to S.K.). **Author contributions:** T. Iijima and S.K. provided organisms and isolated the RNA and DNA; R.K. and T. Itoh designed and performed genome sequencing, assembly, and annotation; S.K. and C.-P.L. collected butterflies in Taiwan; T. Iijima and S.K. designed and performed the experiments; T. Iijima, R.K., T.S., T. Itoh, and H.F. wrote the manuscript; T.S., T. Itoh, and H.F. supervised this project. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** Raw sequencing data have been deposited in the DNA data bank of Japan. Accession information: short-read archive for the *P. memnon* genomes, BioProject ID: PRJDB5519; transcriptome sequence accession ID, SAMD00074277 to SAMD00074282 and SAMD00052594 to SAMD00052595 [previous study (14)].

Submitted 2 August 2017

Accepted 6 March 2018

Published 18 April 2018

10.1126/sciadv.aao5416

Citation: T. Iijima, R. Kajitani, S. Komata, C.-P. Lin, T. Sota, T. Itoh, H. Fujiwara, Parallel evolution of Batesian mimicry supergene in two *Papilio* butterflies, *P. polytes* and *P. memnon*. *Sci. Adv.* **4**, eaao5416 (2018).