Functional involvement of multiple genes as members of the supergene unit in the female-limited Batesian mimicry of *Papilio polytes*

Authors: Shinya Komata[1], Shinichi Yoda[2], Yûsuke KonDo[1], Souta Shinozaki[1], Kouki Tamai[1] and Haruhiko Fujiwara[1]

Affiliations:

[1]Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba, 277-8562, Japan

[2]NIBB Core Research Facilities, National Institute for Basic Biology, Okazaki, 444-8585, Japan.

Abstract

In recent years, many examples have been reported in which complex adaptive traits are regulated by supergene units consisting of multiple genes adjacent to each other on the chromosome. However, few supergene units have been functionally characterized, and it has not been demonstrated that multiple genes in a supergene unit are involved in the adaptive trait. The *H* locus on chromosome 25 has been shown to be responsible for the female-limited Batesian mimicry in *Papilio polytes*. The mimetic *H* locus, about 130 kb, is inverted at both ends from the non-mimetic *h* locus, and the sequences between *H* and *h* locus are diversified (Highly Diversified Region: HDR). The HDR of the *H* locus contains three genes, *doublesex* (*dsx*), *U3X*, and *UXT*, and RNAi analyses have shown that *dsx* is involved in the mimetic pattern formation. Here, we showed that knockdown by RNAi and knockout by Crspr/Cas9 of the transcriptional regulator *UXT* altered the mimetic pattern to the non-mimetic like pattern in hindwings. In addition, it was also shown by RNAi that long non-coding RNA *U3X* was functionally involved in the mimetic pattern formation, and by RNA-seq analysis that *U3X* induced the *dsx* expression and repressed

the *UXT* expression. In addition, the knockdown of *prospero* and *sir2* genes adjacent to the HDR also affected the hindwing pattern formation. These results indicate that not only genes within the HDR but also adjacent genes outside the HDR may be involved in the mimicry supergene function. This suggests that *cis*-elements regulating multiple gene expression exists within the HDR of the *H*-allele, and the need to consider a more extended functional unit of the supergene. This paper shows that multiple genes inside and outside the HDR are functionally involved in the formation of complex adaptive traits, and provides a new perspective on the evolution of the structure and functional units of the supergene.

Introduction

Batesian mimicry is a phenomenon in which a non-toxic species (mimic) escapes predation from predators such as birds (mimic) by mimicking the appearance, colors, shape, and behavior of a toxic and unpalatable species (model) (Bates, 1862). Batesian mimicry is a complex adaptive phenomenon which can be achieved only when multiple traits are combined. For example, the butterfly's wing pattern is composed of various colors and complex patterns, and unless all wing pattern elements are similar to the model species, mimicry cannot be achieved successfully. In the Batesian mimic butterflies, it is also known that not only wing patterns and shapes but also the flying behavior should resemble the model species (Le Roy, 2019). In addition, some *Papilio* species show polymorphic Batesian mimicry, but no intermediate offspring between mimetic and non-mimetic types was observed (Clarke & Sheppard, 1960; 1971; 1972; Clarke et al., 1968). These facts indicate that the multiple sets of traits for the mimicry are inherited in tightly linked manners, which have led to the “supergene” hypothesis (Clarke & Sheppard, 1971). It was originally considered that supergene is composed of multiple flanking genes which are linked in the same chromosomal loci and inherited tightly together (Fisher, 1930). However, it has also been hypothesized that a single gene or a single regulatory element may regulate complex phenotypes such as mimicry by controlling multiple downstream genes (Nijhout, 1994; West-Eberhard, 2003). Thompson and Jiggins (2014) have defined supergene as 'a genetic architecture involving multiple linked functional genetic elements that allows switching between discrete, complex phenotypes maintained in a stable local polymorphism'.

In a swallowtail *Papilio polytes*, only females have the mimetic and non-mimetic phenotypes, and males are monomorphic and non-mimetic (Figure 1a). The mimetic female of *P. polytes* has red spots on the outer edge of the hindwing and pale yellow spots in the center of hindwing, which mimics the unpalatable model butterfly, *Pachliopta aristolochiae* (Euw et al., 1968; Ueugi, 1996). The pale yellow spots of the mimetic and non-mimetic forms differ not only in shape and arrangement, but also in the pigment composition (Nishikawa et al, 2013; Yoda et

al., 2021). Males and non-mimetic females fluoresce under UV fluorescence, whereas mimetic females and model species, *Pachliopta aristolochiae*, do not fluoresce under UV fluorescence (Nishikawa et al, 2013; Yoda et al., 2021). It is also known that the mimetic female also resembles *Pachliopta aristolochiae* in the behavior of flight path (Kitamura & Imafuku, 2015). Previous studies have shown that mimicry is regulated by the *H* locus and that the mimetic female (*H*) is dominant to the non-mimetic female (*h*) according to Mendelian inheritance (Clarke & Sheppard, 1972). Recently, whole genome sequences and genome-wide association studies have shown that the *H* locus is responsible for about 130 kb of chromosome 25, including *doublesex* (*dsx*) (Figure 1b; Kunte et al., 2014; Nishikawa et al., 2015). The direction of this region differs between the *H* allele and the *h* allele due to the inversions at both ends, suggesting that the *H* allele evolved from the *h* allele, given the conserved structure in lepidopteran insects (Kunte et al., 2014; Nishikawa et al., 2015). It is thought that recombination between the two alleles is suppressed by the inversion, and the accumulation of mutations and indels over the years has resulted in a Highly Diversified Region (HDR) with low sequence homology between *H* and *h* (Nishikawa et al., 2015).

Nishikawa et al. (2015) found in *P. polytes*, that knockdown of mimetic (*H*) type *dsx* (*dsx-H*) in the hindwings of mimetic females switched to a wing pattern similar to that of the non-mimetic females using the electroporation mediated RNAi method. Knockdown of the expression of non-mimetic (*h*) type *dsx* (*dsx-h*) did not cause such a switch, indicating that *dsx-H* is important for the formation of mimetic patterns (Nishikawa et al., 2015). On the other hand, the HDR of the *H* allele (HDR-*H*) contains not only *dsx* but also the 5'-untranslated region (UTR) portion of the *Ubiquitously Expressed Transcript* (*UXT*), a transcriptional regulator, and the long non-coding RNA *Untranslated 3 Exons* (*U3X*), present only in HDR-*H* (Figure 1b), but their functions are still unclear (Nishikawa et al., 2015). In *P. memnon*, which is closely related to *P. polytes* and exhibits female-limited Batesian mimicry, the locus responsible for mimicry (*A*) is a *dsx*-containing region of chromosome 25 and consists of two types of HDRs with low homology

between *A*-allele and *a*-allele (Iijima et al., 2018). The mimetic-type HDR (HDR-A) of *P. memnon* also contains the 5'-UTR portion of *UXT* in addition to *dsx* (Iijima et al., 2018). Furthermore, in *P. polytes*, *UXT* and *U3X* are expressed in the hindwing, suggesting that these genes in the HDR may also be involved in the formation of mimetic patterns (Nishikawa et al., 2015).

Many supergene show intraspecific polymorphism due to inversions, but in some cases, such as *P. memenon*, there is no inversion, but two types of HDR structures, mimetic and non-mimetic loci, are maintained (Iijima et al., 2018). However, it has not been clear whether the functional unit of the supergene that regulates complex adaptive traits is limited in the area within the inversion or the region of low homology (i.e., HDR), or whether it extends to neighboring regions. In *P. polytes*, the external gene *prospero*, which is adjacent to the internal gene *UXT* in the HDR, has read-through transcripts only in mimetic females (Nishikawa et al., 2015). This suggests that some *cis*-element in the mimetic HDR may regulate gene expression even in the external region, and that such a gene may be involved in the formation of the mimetic pattern.

In this study, in order to determine the range of functional units in the supergene that controls female-limited Batesian mimicry in *P. polytes*, we examined the expression and function of genes *UXT* and *U3X* in the HDR-*H* region other than *dsx*, as well as *prospero*, *Nach-like*, and *sir2* in close proximity to the outside of the HDR region. For functional analysis, RNAi by *in vivo* electroporation (Ando & Fujiwara, 2013) was performed, and Crispr/Cas9 knockout was also performed for *UXT*. In addition, RNA sequencing (RNA-seq) was performed on the *dsx-H*, *UXT*, and *U3X* gene knockdown wings and the control wings to elucidate the regulatory relationship and downstream genes of the three genes. Finally, we found *rotund* (*rn*) as a common downstream gene of *dsx-H*, *UXT*, and *U3X*, in addition to *Wnt1* and *Wnt6*, which have been reported previously (Iijima et al., 2019), and analyzed its function by RNAi. In addition, we investigated the function of *dsx-h* and the expression and function of the three isoforms in *dsx* in order to investigate the detailed function of *dsx*.

Results

Comparison of the expression levels of genes inside and flanking the HDR

In order to compare the temporal expression levels of internal and external proximal genes of the HDR over time, we prepared mRNA from the hind wings of mimetic females (*Hh*), non-mimetic females (*hh*), and males (*Hh*), and examined the expression levels of *prospero*, *UXT-H* (*UXT* from *H*-allele), *UXT-h* (*UXT* from *h*-allele), *U3X*, *dsx-H*, *dsx-h*, *Nach-like*, and *sir2* by RNA-seq. It is reported and that wing coloration occurs after day 9 of pupation (P9), and that the expression of *dsx-H*, which is thought to play a major role in the formation of mimetic patterns in *P. polytes*, peaked on days 1–3 of pupation (P1–P3), suggesting that the mimetic pattern formation is determined on days 1–3 of pupation (Nishikawa et al., 2013; 2015). Therefore, gene expression in the first half of the pupal stage was considered to be important for the mimetic wing pattern formation. In order to investigate whether each gene in and flanking the HDR-*H* is involved in these process, we examined the expression levels of the genes in the hindwing imaginal discs at the wandering stage (W0) of the last instar larvae, P2 and P5 (Table S1 shows the list of samples used). The results showed that *Nach-like* was not expressed at all in any developmental stage. It is reported that *Nach-like* was not expressed in the hindwings of *P. memnon*, and was only expressed in the testes and ovaries (Iijima et al., 2018). Other genes were expressed in mimetic females, non-mimetic females and males (Figures 1c,1S).

In the RNA-seq data of mimetic females, *dsx-H* and *dsx-h* showed contrasting expression patterns. *dsx-H* showed a peak expression in the P2 stage, while *dsx-h* was highly expressed in the P5 stage (Figure 1c). The expression of *U3X*, which is only present in the *H* locus, tended to be relatively high in P2 while the data was not statistically significant, and to show the constant expression (Figure 1c). There was no significant difference in the expression pattern of *UXT* between *UXT-H* and *UXT-h*, and the expression level of *UXT* was higher in W0 and P2 and significantly lower in P5 (Figure 1c). The expression pattern of *prospero* was significantly larger

in P2 as in *dsx-H*, and that of *sir2* was largest in W0 as in *UXT*, but not statistically significant (Figures 1c, 2f, 2g), In the RNA-seq experiment, we used three or more samples at each stage (W0, P2, P5) for mimetic females, but insufficient numbers of samples for some non-mimetic females and males (Figure S1), and thus we further performed reverse transcription quantitative PCR (RT-qPCR) using P2 and P5 samples for mimetic females (*Hh*), non-mimetic females (*hh*), and males (*Hh*) (Figure 2).

RT-qPCR showed that the *dsx-H* expression was significantly high in P2 of mimic females but low in P5 and males in every stage (Figure 2a). The *dsx-h* expression was significantly high in P5 of non-mimic females compared to other stages, mimic females and males (Figure 2b). It is noteworthy that the expression of *dsx* was low in males at all stages (Figures 2a, b). The highly expression of *dsx-H* in P2 of mimetic females is consistent with RNA-seq results (Figure 1c) and previous studies, which may be related to the mimetic color pattern formation (Nishikawa et al., 2015; Deshmukh et al., 2020). *U3X* was not detected in non-mimetic females (*hh*) because it is present only in the *H* locus, and was expressed in mimetic females and males at P2 and P5 stages. For *UXT-H*, there was no significant difference in expression levels among mimetic females and males in any stages (Figure 2c). The expression of *UXT-h* was significantly greater in non-mimetic females (*hh*), probably due to the fact that they are *h* homozygous, but it is particularly high in non-mimetic females in P5, about five times higher than the expression of *UXT-h* in mimetic females (Figure 2e).The expression of *sir2* and *prospero* in mimetic females was similar to that of RNA-seq results, and their expression was also observed in non-mimetic females and males. The expression of *sir2* was significantly higher in males in P5 than in mimetic or non-mimetic females (Figures 2f, 2g).

Expression and functional roles of *dsx-h* and three isoforms of *dsx*

It is important to clarify the functional roles of *dsx-H* and *dsx-h* in the evolution of mimicry supergene. Although the involvement of *dsx-H* in the formation of mimetic traits has been shown,

the function of *dsx-h* has been unclear (Nishikawa et al., 2015; Iijima et al., 2019; Yoda et al., 2021). The hindwing patterns of non-mimetic females and males are almost the same in the bright field, but there are minute differences when observed under the UV irradiation. In non-mimetic females, the innermost and second innermost pale yellow spots do not fluoresce, whereas in males the innermost one fluoresces slightly and the second one fluoresces completely (see Nontreated in Figures 3a, b). When *dsx* was knocked down in non-mimetic females (*hh*), the second pale yellow spot fluoresced, showing a similar pattern to males (Figures 3a, S3). In males, however, there was no clear change after knockdown (Figures 3b, S3), indicating that *dsx* maintains its original function of sexual differentiation in the hindwing as well as mimetic pattern formation.

In addition, *dsx* has three female isoforms (F1, F2, F3) both in *dsx-H* and *dsx-h*, and one isoform in *dsx-H* and *dsx-h* in males (Kunte et al., 2014; Nishikawa et al., 2015). To investigate the function of the three isoforms in females, we performed expression analysis by RNA-seq and knockdown experiment by RNAi. There was no significant difference in the expression levels of F1, F2, and F3, but F3 tended to have relatively higher expression levels (Figures 4, S2). RNAi by *in vivo* electroporation showed that only the F3-specific knockdown of *dsx-H* changed the mimetic pattern to the similar of the non-mimetic pattern (Figures 5, S4): red spots became smaller and the non-mimetic specific pale yellow spots appeared (Figures 5c, S4). In the RNAi experiment, we confirmed that only the target isoform was down-regulated by RT-qPCR (Figures 5d, 5e, 5f). These results indicate that only isoform 3 of *dsx-H* has an important function in the formation of mimetic traits.

Functional analysis of *UXT* and *U3X* in the HDR

In order to investigate whether *UXT* and *U3X* in HDR-*H* are involved in the formation of mimetic patterns in the hindwing, *in vivo* electroporation mediated RNAi were conducted. We injected siRNA of the target gene into the hindwing immediately after pupation (Day 0 of pupation: P0) and performed electroporation to cover most of the hindwing (Figure 6a), which

induces RNAi only in the target area. On the second day after injection (P2), siRNA un-injected hindwings (control) and injected hindwings (knockdown) were sampled for RNA extraction. The extracted RNA was subjected to RT-qPCR to confirm the decreased expression of the knocked-down gene (Figures 6b, 6c, 6d), and RNA-seq was performed to compare the expression levels in the control and knockdown sides (the list of samples used for RNA-seq is shown in Table S2). In addition, we injected siRNA at P0 and observed the emerged adult to confirm the phenotypic changes caused by knockdown (Figure 6a).

First, knockout of *UXT* by Crispr/Cas9 was also performed. guide RNA was designed to target the functional domain of *UXT*, the prefoldin domain (Figure S5a), and was injected into eggs immediately after egg laying together with Cas9 protein (CP-01, PNA Bio). A total of 294 eggs were injected and 21 adults were obtained (Mimetic female: 8; nonmimetic female: 5; male: 8, Figure S5b). Eight of the mimetic females were subjected to PCR, cloning, Sanger sequencing to confirm the introduction of mutations (Figure S6), and phenotypic observation. The hatching rate (7.1%) of the injected individuals was very low, suggesting that *UXT* may have an important function in survival. Genotyping using DNA extracted from the abdomen, head and wings of emerged individuals yielded five sequences in which mutations were introduced (Figure S6). We observed individuals with the mosaic knockout, in which the pale yellow spots were flattened as in the non-mimetic form, and the red spots were reduced or disappeared (Figure 6a). In only one individual, the pale yellow spots were changed, but the red spots were reduced in four individuals (Figure S7).

Similarly, in the *UXT* knockdown by siRNA injection in the hindwing, the pale yellow spots were reduced and the shape of pale yellow region was flattened like non-mimetic phenotype, and the red spots were reduced or disappeared (Figure 7b, S8). In the *U3X* knockdown, the pale yellow spots were extended downward (Figures 6c, 6d, S9), and the red spots below the innermost pale yellow spots, which were only slightly visible in the control, were enlarged (indicated by red arrow in Figure 6d). These results indicate that *UXT* is involved in mimetic pattern formation,

and knockdown of *U3X* also resulted in phenotypic changes, but not simply a change from mimetic to non-mimetic phenotype. Since *U3X* is a non-coding RNA, the phenotypic changes observed upon knockdown of *U3X* may be due to changes in the expression of other genes that are regulated by *U3X*. The expansion of red and pale yellow spots upon knockdown of *U3X* suggests the existence of genes that play a role in suppressing the excessive appearance of these spots.

The regulatory relationship and downstream genes of the three genes inside the HDR-*H*

Since all three genes in the HDR-*H*, *dsx-H*, *U3X*, and *UXT*, were found to be involved in the formation of mimetic patterns, we decided to examine their mutual regulatory relationships and their regulatory relationships with other genes. To investigate the regulatory relationship with other genes, we next compared gene expression in the knockdown and control sides of *dsx-H*, *UXT*, and *U3X* after the RNAi treatment for each gene. According to the comparative analyses, about 500 to 1500 genes were extracted as genes whose expression was decreased or increased when each gene was knocked down (Figure 8a).

We focused on the transcription factors and signaling factors whose expression is promoted by *dsx-H*, *UXT*, and *U3X* (Figurex 9, 10). As a result, we found that *Wnt1*, *Wnt6*, and *rotund* (*rn*) were commonly down-regulated by knockdown of *dsx-H*, *UXT*, and *U3X* (Figure 8b). *Wnt1* and *Wnt6* have been reported to be involved in mimetic pattern formation (Iijima et al., 2019). When we knocked down *rn*, there was no characteristic change in the mimetic pattern, but there was an overall change in the color of the black, red, and pale yellow regions, which seemed to become lighter (Figures 11, S10). When observed under UV irradiation in the *rn* knockdown wings, the UV fluorescence was observed in the pale yellow spots (Figures 11, S10), where UV fluorescence is not observed usually in the mimetic form. From these observations, we consider that *rn* plays an important role in the unique pigment synthesis of the mimetic phenotype.

We also examined the expression levels of genes inside and flanking the HDR (*dsx-H*,

dsx-h, *UXT*, *U3X*, *prospero*, *sir2*, *rad51*) upon knockdown of *dsx-H*, *UXT*, and *U3X*. When *dsx-H* was knocked down, the expression of *dsx-H* was significantly decreased, but no significant expression changes were observed in other genes (Figure 8c). Similarly, when *UXT* was knocked down, the expression of *UXT* tended to decrease (not statistically significant), but there was no significant effect on the expression of other genes (Figure 8d). On the other hand, notably, when *U3X* was knocked down, in addition to the downward trend of *U3X* expression (not statistically significant), the expression of *dsx-H* was significantly decreased and the expression of *UXT* was significantly increased (Figure 8e).

Functional analysis of *prospero* and *sir2*, two proximal genes outside the HDR-*H*

Next, the functional roles of *prospero* and *sir2* which locate in close proximity to the HDR region but outside the inversion, were analyzed in mimetic females hindwings by *in vivo* electroporation mediated RNAi. In the *prospero* siRNA injected hindwing, there were no obvious changes, but the red spots characteristic of the mimetic form were subtly enlarged (Figures 12a, S11). The decrease in *prospero* expression by RNAi was confirmed by RT-qPCR, and although not statistically significant, there was a tendency for *prospero* expression to decrease in the knockdown side (Figure 12b). In the case of *sir2* RNAi, the pale yellow spots were flattened like non-mimetic phenotype, and the red spots were enlarged under the innermost pale yellow spots (Figures 12c, S12). *sir2* expression was significantly decreased in the knockdown side (Figure 12d). These results suggest that multiple adjacent genes outside the HDR are involved in the formation of mimetic patterns.

Discussion

In this paper, using the *in vivo* electroporation mediated RNAi method, we show that not only *dsx*, but also *UXT* and *U3X* in the inversion region for the *H*-allele, furthermore even outside

flanking genes *prospero* and *sir2* are involved in the mimetic wing pattern formation in *P. polytes* (Figures 7, 12).

In recent years, many examples have been reported of supergenes in which complex adaptive traits showing intraspecific polymorphism are regulated throughout a certain region of the chromosome (Gutiérrez-Valencia et al., 2021; Villoutreix et al., 2021). In many of these cases, chromosomal inversions are observed, and the structural diversity of multiple alleles is thought to be fixed by the inversions. However, most supergene by an inversion contain more than a few dozen genes (some large supergenes contain more than 100 genes), and very few functional analyses have shown that specific genes are involved in adaptive traits. Historically, it was assumed that multiple genes work together to produce more complex traits and to prevent recombination by placing genes adjacent to each other on the chromosome to avoid intermediate forms in the next generation, and such regions were defined as supergene (Fisher, 1930; Thompson & Jiggins, 2014). However, there has been no evidence that multiple genes belonging to the supergene are involved in complex adaptive traits. In the first place, it should be discussed whether the supergene should be limited to the internal region fixed by inversion or not. In the case of the female-limited polymorphic Batesian mimicry of *P. polytes* and its close relative, *P. memnon*, the whole genome sequence and GWAS showed that the causative region of the mimicry was a 150-kb region including *dsx* on chromosome 25 (Iijima et al., 2018). Both species have two types of low homology sequences (HDRs) corresponding to mimetic and non-mimetic alleles, but there is an inversion between the two alleles in *P. polytes*, but not in *P. memnon* (Iijima et al., 2018). It is not clear how sequence diversity arose and was maintained in *P. memnon*, but at least in *P. memnon*, the supergene cannot be defined in terms of the internal region of inversion. Rather, it may be possible to define a supergene inside an HDR with low sequence homology. However, is it possible to define a supergene including outer regions adjacent to an HDR with low sequence homology? This is an important question for understanding how we should think about the unit of the supergene and how the supergene has evolved.

The fact that the mimicry supergene of *P. polytes* is only 130 kb in size and contains only three genes in the inversion region makes it more suitable than other supergenes for answering the above questions. In addition, it is a great advantage to be able to discuss it in comparison with the supergene of a related species, *P. memnon*, which does not have an inversion. On the other hand, it has been shown that *dsx-H* in the supergene induces the mimetic pattern of female hindwings by analysis of electroporation mediated RNAi (Nishikawa et al., 2015). In this study, *dsx-h*, which has been considered to have no specific function, was shown to induce non-mimetic patterns in females (Figure 3). This clearly indicates that the non-mimetic pattern in males is a default trait, and if *dsx-h* is expressed there, it becomes non-mimetic female, and if *dsx-H* is expressed there, it becomes mimetic female. This result suggests that *dsx-h*, which was originally involved in the regulation of sexual dimorphism in wing pattern, underwent recombination suppression (i.e. inversion), resulting in *dsx-H* and the evolution of female-limited Batesian mimicry. In butterflies, the evolution of female-limited polymorphism based on sexual dimorphism has been frequently hypothesized from evolutionary studies (Kunte, 2009; Timmermans et al., 2017). Furthermore, functional analysis suggests that only isoform 3 of *dsx-H*, induces a mimetic pattern among the three female isoforms in this study (Figure 5). The expression levels of each isoform were not significantly different, suggesting that the function of isoform 3 as a protein is important for the induction of mimetic traits, rather than the relative quantities of its expression. Dsx is a transcription factor involved in sexual differentiation, and each isoform binds to a different response sequence, suggesting that the downstream gene network may change among three isoforms. Iijima et al. (2019) explored the downstream gene network of *dsx-H* as a whole, and it may be necessary to explore the downstream genes specific to isoform 3 of *dsx-H* for clarifying the mimicry mechanism in the future.

The transcription factor *dsx* has been thought to function as a mimetic supergene as a single gene because it induces downstream genes to express the trait (Kunte et al., 2014; Baral et al., 2019). However, the present experiments indicate that internal genes such as *U3X* and *UXT*

and external genes such as *prospero* and *sir2* are also involved in pattern formation. These genes are not downstream genes of *dsx-H*, but are likely to function as members of the supergene. On the other hand, it is also important to note that we found that the expression of *U3X* may induce the expression of *dsx-H*. *U3X* is a long non-coding RNA not found in the *h*-allele, other genomic regions and is thought to have arisen specifically in HDR-*H* during evolution. In addition, *U3X* is located upstream of the transcription start site of *dsx-H*, suggesting that *U3X* cis-regulates *dsx-H* expression. In *Daphnia magna*, long non-coding RNAs are also present upstream of *dsx* and regulate *dsx* function (Kato et al., 2018), which indicates that further investigating more details of the regulatory mechanism of *U3X* expression are necessary.

Knockdown of *UXT* switches the pattern to resemble the non-mimetic phenotype, including flattening of the upper part of the pale yellow spots in the center of the hindwing (Figure 7b). Importantly, knockdown of *UXT* in Crispr/Cas9 resulted in a similar phenotypic change (Figure 7a), and the results were consistent between the two completely different experimental methods. The knockdown of *U3X* did not necessarily cause a change from the mimetic to the non-mimetic patterns, but the RNA-seq results showed that *U3X* also repressed the expression of *UXT*, suggesting that the knockdown of *U3X* may have had the effect of increasing the gene expression of *UXT* (Figure 8e). The knockdown of *prospero* and *sir2* also did not have the effect of simply changing the mimetic pattern to a non-mimetic pattern (Figure 12), but these genes have never been reported to be involved in pattern formation, so it is highly likely that they are involved in mimetic pattern formation. In addition to *dsx*, other genes may be modifiers that are fine-tuned to improve mimicry (hypothesized by Charlesworth (2016)).

On the other hand, the results of expression analysis of each gene do not clearly indicate whether each gene is involved in the control of mimicry pattern formation (Figures 1c, 2). In this study, mRNAs were prepared from the entire hindwing for comparison, and thus if a gene is expressed in a specific region (e.g., red spot region), it may not be possible to clearly judge the functional involvement of the gene in a mimetic pattern from the expression level. The only way

to solve this problem is to compare the expression of each gene by *in situ* hybridization. On the other hand, we compared gene expression levels at only three developmental timing: W0 (the first stage of the prepupa), P2, and P5. In order to obtain clear results, it is necessary to continuously compare gene expression levels at a wider range of time points. In addition, due to technical limitations, electroporation-mediated RNAi (siRNA injection) in the wing can only be performed immediately after pupation, which may not necessarily correspond to the time when each gene is functioning. In the case of *Dsx-H* knockdown, it is noteworthy that the mimetic pattern is switched to the non-mimetic pattern even if RNAi is performed immediately after pupation (Nishikawa, 2015), suggesting that the fate of pattern formation is carried over at least to the early pupal stage. If RNAi can be applied to other stages of development, the functional role of each gene can be clarified.

Our present results suggest that the unit of the mimicry supergene should be defined to include at least the external neighboring genes. The results of comparative transcriptome analysis after knockdown of *dsx-H*, *U3X*, and *UXT* did not show that the expressions of *sir2* and *prospero* were not affected as downstream genes of the 3 genes, suggesting that *sir2* and *prospero* expression is likely to be regulated by some *cis*-elements within HDR-*H*. The existence and location of *cis*-regulatory elements need to be investigated in the future, including possible epigenetic regulation of multiple genes in HDR-*H*. The significance of having related genes adjacent to each other on the chromosome may also be re-considered from the perspective of such expression regulation.

Even if outside the recombination suppression region (i.e., within the recombination region between *H* and *h*), *prospero* and *sir2* of *P. polytes* may have basic functions involved in mimicry pattern formation. For example, in the past, recombination sites may have been located further out and HDR dimorphism may have been more widespread, including adjacent *sir2* and *prospero*. In the process of evolution, *sir2* and *prospero* acquired functions involved in the pattern formation in addition to their original functions, and if these genes are able to function outside of

recombination, they may not necessarily be in the recombination repressor region. However, the regulation of their expression needs to be affected inside the H-HDR.

Methods

Butterfly rearing

We purchased wild-caught *P. polytes* from Chokan-kabira (Okinawa, Japan) and Mr. Itsuzai Aoki (Okinawa, Japan), and obtained eggs and used for the experiment. The larvae were fed the leaves of *Citrus hassaku* (Rutaceae) or on an artificial diet, and were kept at 25 °C under long-day conditions (light:dark = 16:8 h). Adults were fed a sports drinks (Calpis, Asahi. Japan).

Analysis of expression levels of genes in and flanking the HDR

In this study, we used the entire hindwing of *P. polytes* to analyze the expression levels of internal (*U3X*, *UXT*, *dsx*) and external flanking genes (*Nach-like*, sir2, prospero) in the HDR by RNA-seq and RT-qPCR. In addition to the published RNA-seq read data of *P. polytes* (BioProject ID: PRJDB2955; Nishikawa et al. 2015), newly sampled RNA was used for the analysis. The sample list used in the experiment is shown in Table S1. The newly added RNA-seq reads were obtained by the following procedure. The entire hindwing was sampled for RNA extraction on pupal day 2 (P2) and pupal day 5 (P5), and RNA extraction was performed using TRI reagent (Singma) in the same manner as Nishikawa et al. (2015) and Iijima et al. (2019). The extracted and DNase I (TaKaRa, Japan) treated RNA was sent to Macrogen Japan Corporation for library preparation by TruSeq stranded mRNA (paired-end, 101 bp) and sequenced by Illumina platform. The obtained RNA-seq reads were quality-checked by FastQC (version 0.11.9; Andrews, 2010), mapped by Bowtie 2 (version 2.4.4; Langmean, 2012), and the number of reads was counted using SAMtools- (version 1.14; Li et al., 2009). Based on the number of reads, FPKM (Fragments Per Kilobase of transcript per Million mapped reads) was calculated (FPKM=number of mapped reads/gene length(bp)/total number of reads*10^9). For mapping, full-length mRNA sequences including

UTRs were used for *prospero*, *UXT-H*, *UXT-h*, *U3X*, *sir2*, and *Nach-like*, and ORF region sequences were used for *dsx-H* and *dsx-h*. *dsx-H* and *dsx-h* were mapped to three female isoforms for female individuals and one male isoform for male individuals. For *dsx-H* and *dsx-h*, three female isoforms were used for female individuals, and male isoforms were used for male individuals. Sequence information for each gene was obtained from Nishikawa et al. (2015) and Iijima et al. (2019).

The expression levels of *U3X*, *UXT-H*, *UXT-h*, *dsx-H*, *dsx-h*, *sir2*, and *prospero* were also analyzed by RT-qPCR. RNA obtained by the above method was subjected to cDNA synthesis using Verso cDNA synthesis kit (Thermo Fisher Scientific). The qPCR was performed using Power SYBR® Green Master Mix (Thermo Fisher Scientific) by The StepOneTM Real-Time PCR System (ABI). The detailed method was followed by Iijima et al. (2019). A total of 18 whole hindwing samples from 18 individuals were used, including three each of mimetic females (*Hh*), non-mimetic females (*hh*), and males (*Hh*) of P2 and P5. *RpL3* was used as an internal standard and the primers used are shown in Table S2.

Knockout of *UXT* by Crispr/Cas9

A single guide RNA (sgRNA) was used to generate deletions and frameshifts within the prefoldin domain of *UXT* (Figure S5). A sgRNA was designed using CRISPRdirect (https://crispr.dbcls.jp), and the specificity of the sequence of sgRNA was assessed using BLAST to ensure that there were no multiple binding sites. The target sequence is shown in Figure S5a. The sgRNA template was generated by PCR amplification with forward primers encoding the T7 polymerase binding site and the sgRNA target site (Pp_UXT_F1modi, GAAATTAATACGACTCACTATAGGCCGACCAGAAGCTTCATCGTTTAAGAGCTATGCTGGAAACAGCATAGC), and reverse primers encoding the remainder of the sgRNA sequence (sgRNA_Rmodi, AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTAAA

CTTGCTATGCTGTTTCCAGCATA), using Phusion DNA polymerase (M0530, New England Biolabs, Ipswich, MA, USA) (Zang & Reed, 2017). In vitro transcription was performed using the Megascript T7 Kit (Thermo Fisher Scientific) and sgRNA was purified with the MEGAclear Transcription Clean-Up Kit (AM1908, Thermo Fisher Scientific). To collect eggs for injection, host plants were provided to female butterflies and allowed to lay eggs for 1 hour. The obtained eggs were aligned on a glass slide and fixed with an instant glue Aron Alpha (Toagosei Company, Japan). The fixed eggs were disinfected with formalin for 3 min, the tip of the glass capillary was cut with a razor at an angle of 30-40°, perforated with a tungsten needle, and the capillary was injected with an injection mixture containing sgRNA (500 ng/ul) and Cas9 protein (CP-01, PNA Bio; 500 ng/ul) (injection pressure Pi 100 Pa, steady pressure Pc 40-80 Pa). Finally, the holes were sealed with Arone Alpha, placed in a Petri dish, and stored in a plastic case along with a well-moistened comfort towel. The hatched larvae were reared in the same manner as described above. The emerged adults were observed for phenotype, and parts of the head, abdomen and wings were taken for genotyping. DNA was extracted using a phenol-chloroform protocol and PCR amplified across the target sites (primers, Pp_UXT_cr_F1, ttcgtgttcaggatcaacag; Pp_UXT_cr_R1, tatttgttaactgcccgatg). PCR products were used to perform TA cloning, Sanger sequencing, and genotyping.

Functional analysis by RNAi using *in viro* electroporation

siDirect (http://sidirect2.rnai.jp/) was used to design the siRNAs. The target sequences were blasted against the predicted gene sequence (BioProject: PRJDB2954) and the genome sequence (BioProject: PRJDB2954) in *P. polytes* to confirm that the sequences were highly specific, especially for the target genes. The designed siRNA was synthesized by FASMAC Co., Ltd. (Kanagawa, Japan). The RNA powder received was dissolved in Nuclease-Free Water (Thermo Fisher, Ambion), adjusted to 500 μM, and stored at -20°C. The sequence information of the siRNA used is listed in Table S3. A glass capillary (Narishige, GD-1 Model, Glass Capillary with

Filament) was processed into a needle shape by heating it at HEATER LEVEL 66.6 using a puller (Narishige, PP-830 Model). The capillary was filled with siRNA. siRNA was adjusted to 250 μM when only one type of siRNA was used for one target gene (*dsx-H*, *dsx-h&H*, *UXT*, *U3X*, *rn*), and 500 μM siRNA solution was mixed in equal amounts when two types of siRNA were mixed for one target gene (*prospero*, *sir2*). The capillary was filled with siRNA and 4 μl of siRNA was injected into the left hindwing under a stereomicroscope using a microinjector (FemtoJet, eppendorf). Then, siRNA was introduced into only the positive pole side of the electrode by applying voltage (5 square pulses of 7.5 V, 280 ms width) using an electroporator (Cellproduce, electrical pulse generator CureGine). A PBS gel (20×PBS: NaCl 80g, Na2HPO4 11g, KCl 2g, K2HPO4 2g, DDW 500ml; 1% agarose) was placed on the dorsal side of the hindwing and a drop of PBS was placed on the ventral side of the hindwing. The detailed method follows that described in the previous paper (Ando and Fujiwara, 2013). The pictures of all the individuals who performed the function analysis are described collectively as Supplementary figures.

Regulatory relationship of *dsx-H*, *UXT*, and *U3X* expression by RNAi and downstream gene screening

After sampling the hindwings of individuals with *dsx-H*, *UXT*, and *U3X* knockdown by RNAi in the P2 stage with the siRNA-injected side as knockdown and the non-injected side as control, total RNA was extracted and DNase I treated RNA was sent to Macrogen Japan Corporation. Libraries were prepared using TruSeq stranded mRNA (paired-end, 101 bp) and sequenced using the Illumina platform. Sample and read information are shown in Table 4. dsx-H_Control_2 and dsx-H_knockdown_2 are the read data used in a previous study (Iijima et al., 2019). We first performed quality check using FastQC (Version 0.11.9; Andrews, 2010), and the reads were mapped to the transcript sequences of *P. polytes* to calculate the expression levels. The transcript sequence was obtained from NCBI, GCF_000836215.1_Ppol_1.0_rna.fna (BioProjects: PRJNA291535, PRJDB2954). Because the transcript sequence information of the genes around

H locus described in GCF_000836215.1_Ppol_1.0_rna.fna was incomplete (*H* and *h* derived isoforms of *dsx* and *UXT* were confused), read mapping to the genes around *H* locus (*prospero*, *UXT*, *U3X*, *dsx-H*, *dsx-h*, *sir2*, *rad51*) was performed separately: the full-length mRNA sequences including UTRs were used for *prospero*, *UXT*, *U3X*, *sir2*, and *rad51*, and the ORF region sequences for *dsx-H* and *dsx-h* were used. The mapping software used was Bowtie2 (version 2.4.4; Langmead & Salzberg, 2012), and the number of reads counted was calculated using SAMtools (version 1.14; Li et al., 2009). Based on the number of reads, FPKM (Fragments Per Kilobase of transcript per Million mapped reads) was calculated (FPKM=number of mapped reads/gene length (bp)/total number of reads*10^9).

In addition, R software was used to extract genes with variable expression by statistical analysis of read data, and comparison between two groups with correspondence using Wald-test of DESeq2 (version 3.14; Love et al., 2014) was performed. The transcription factors and signaling factors were extracted using the GOterms of the top hit amino acid sequences by Blastx against the Uniprot protein database. “transcription factor activity” [GO:0003700], “DNA-binding transcription factor activity, RNA polymerase II-specific” [GO:0000981] as transcription factors and “signaling receptor binding” [GO:0005102], “DNA-binding transcription factor activity” [GO:0003700] as signaling factor.

References

Ando T, Fujiwara H 2013 Electroporation-mediated somatic transgenesis for rapid functional analysis in insects. *Development* **140,** 454–458.

Andrews S. 2010 *FastQC: A Quality Control Tool for High Throughput Sequence Data.* [http://www.bioinformatics.babraham.ac.uk/projects/fastqc/]

Baral S, Arumugam G, Deshmukh, R, Kunte K. 2019 Genetic architecture and sex-specific selection govern modular, male-biased evolution of *doublesex*. *Sci*. *Adv*. **5,** eaau3753.

Bates HW. 1862 Contributions to an insect fauna of the Amazon Valley (Lepidoptera: Heliconidae). *Trans Linn Soc Lond* **23,** 495–556.

Charlesworth D. 2016 The status of supergenes in the 21st century: recombination suppression in Batesian mimicry and sex chromosomes and other complex adaptations. *Evol. Applica*. **9,** 74–90.

Clarke CA, Sheppard PM. 1960 The genetics of *Papilio dardanus* Brown. II. Races *dardanus*, *polytrophus*, *meseres*, and *tibullus*. *Genetics* **45,** 439–456.

Clarke CA, Sheppard PM. 1971 Further studies on the genetics of the mimetic butterfly *Papilio memnon* L. *Phil. Trans. R. Soc. Lond, B* **263,** 35–70.

Clarke CA, Sheppard PM. 1972 The genetics of the mimetic butterfly *Papilio polytes* L. *Phil. Trans. R. Soc. Lond, B* **26,** 431–458.

Clarke CA, Sheppard PM, Thornton IWB. 1968 The genetics of the mimetic butterfly *Papilio memnon* L. *Phil. Trans. R. Soc. Lond, B* **254,** 37–89.

Deshmukh R, Lakhe D, Kunte K. 2020 Tissue-specific developmental regulation and isoform usage underlie the role of *doublesex* in sex differentiation and mimicry in *Papilio* swallowtails. *R. Soc. Open Sci.* **7,** 200792.

Euw JV, Reichstein T, Rothschild M. 1968 Aristolochic acid-I in the swallowtail butterfly *Pachlioptera aristolochiae* (Fabr.) (Papilionidae). *Isr. J. Chem*. **6,** 659–670.

Fisher RA. 1930 The Genetical Theory of Natural Selection. (Clarendon Press, Oxford)

Gutiérrez-Valencia J, Hughes PW, Berdan EL, Slotte T. 2021 The genomic architecture and evolutionary fates of supergenes. *Genom. Biol. Evol.* **13,** evab057.

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

Iijima T, Yoda S, Fujiwara H. 2019 The mimetic wing pattern of *Papilio polytes* butterflies is regulated by a *doublesex*-orchestrated gene network. *Commun. Biol.* **2,** 1–10.

Kato Y, Perez CAG, Ishak NSM, Nong QD, Sudo Y, Matsuura T, Wada T, Watanabe H. 2018 A 5′ UTR-overlapping LncRNA activates the male-determining gene *doublesex1* in the crustacean *Daphnia magna*. *Curr. Biol*. **28,** 1811–1817.

Kitamura T, Imafuku M. 2015 Behavioural mimicry in flight path of Batesian intraspecific polymorphic butterfly *Papilio polytes*. *Proc. R. Soc. B* **282,** 20150483.

Kunte K. 2009 The diversity and evolution of Batesian mimicry in *Papilio* swallowtail butterflies. *Evolution* **63,** 2707–2716.

Kunte K, Zhang W, Tenger-Trolander A, Palmer DH, Martin A, Reed RD, Mullen SP, Kronforst MR. 2014 *doublesex* is a mimicry supergene. *Nature* **507,** 229–232.

Langmead B, Salzberg SL. 2012 Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9,** 357–359.

Le Roy C, Debat V, Llaurens V. 2019 Adaptive evolution of butterfly wing shape: from morphology to behaviour. *Biol. Rev.* **94,** 1261–1281.

Li H et al. 2009 The sequence alignment/map format and SAMtools. *Bioinformatics* **25,** 2078-2079.

Love MI, Huber W, Anders S. 2014 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. **15,** 550. doi: 10.1186/s13059-014-0550-8.

Nijhout HF. 1994. Developmental perspectives on evolution of butterfly mimicry. *Bioscience* **44,** 148–157.

Nishikawa H, Iga M, Yamaguchi J, Saito K, Kataoka H, Suzuki Y, Sugano S, Fujiwara H. 2013 Molecular Basis of the wing coloration in a Batesian mimic butterfly, *Papilio polytes*. *Sci. Rep*. **3,** e3184.

Nishikawa H *et al*. 2015 A genetic mechanism for female-limited Batesian mimicry in *Papilio* butterfly. *Nat. Genet.* **47,** 405–409.

Thompson MJ, Jiggins CD. 2014 Supergenes and their role in evolution. *Heredity* **113,** 1–8.

Timmermans MJ, Thompson MJ, Collins S, Vogler AP. 2017 Independent evolution of sexual

dimorphism and female-limited mimicry in swallowtail butterflies (*Papilio dardanus* and *Papilio phorcas*). *Mol. Ecol.* **26,** 1273–1284.

Uesugi K. 1996 The adaptive significance of Batesian mimicry in the swallowtail butterfly, *Papilio polytes* (lnsecta, Papilionidae): associative learning in a predator. *Ethology* **102,** 762–775.

Villoutreix R, Ayala D, Joron M, Gompert Z, Feder JL, Nosil P. 2021 Inversion breakpoints and the evolution of supergenes. *Mol. Ecol.* **30,** 2738–2755.

West-Eberhard MJ. 2003 Developmental Plasticity and Evolution. (Oxford University Press, New York).

Yoda S *et al*. 2021 Genetic switch in UV response of mimicry-related pale-yellow colors in Batesian mimic butterfly, *Papilio polytes*. *Sci. Adv.* **7,** eabd6475.

Zhang L, Reed RD. 2017 *Chapter8: A Practical Guide to CRISPR/Cas9 Genome Editing in Lepidoptera*, In: *Diversity and Evolution of Butterfly Wing Patterns* (eds: T. Sekimura & H. F. Nijhout), pp. 155-172.

Acknowledgments

We thank Drs. T. Kojima and T. Iijima for helpful comments and experimental supports. This work was supported by Ministry of Education, Culture, Sports, Science and Technology/ Japan Society for the Promotion of Science KAKENHI (20017007, 22128005, 15H05778, 18H04880, 20H04918, 20H00474 to H.F.; 19J00715 to S. K).

Author contributions

H.F. conceived the study; S.K., S.Y., Y.K., S.S. and K.T. conducted experiments; S.K. and H.F. wrote the paper. H. F. supervised this project. All authors reviewed the manuscript.

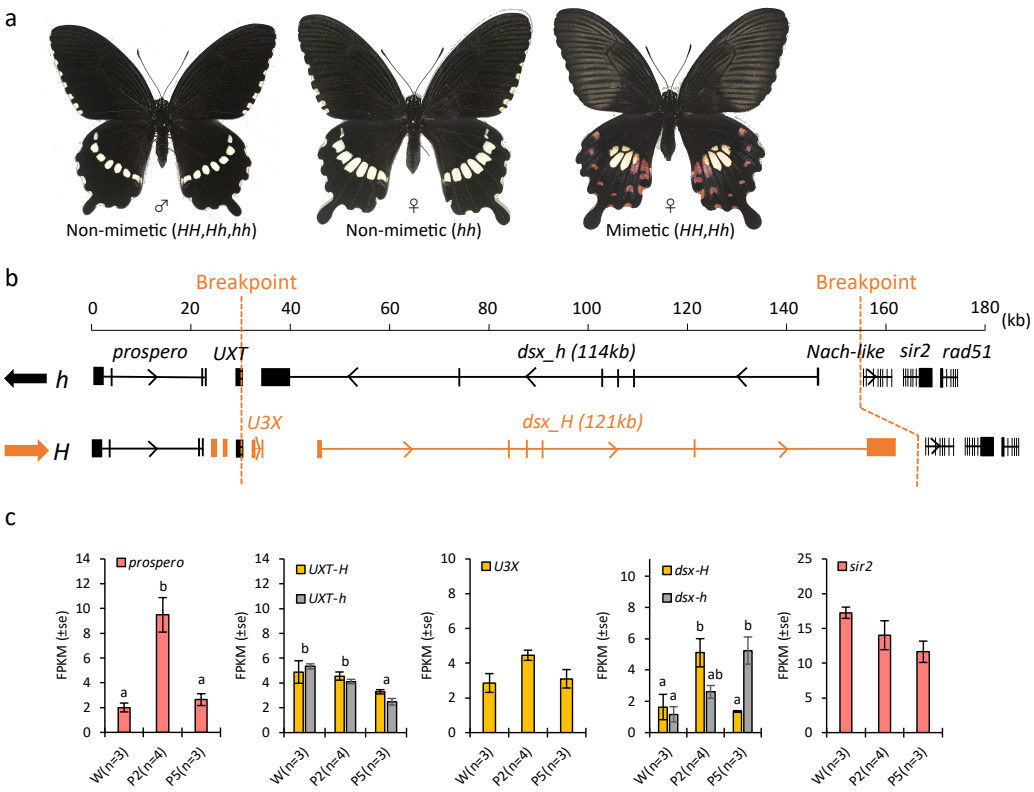


Figure 1. (a) Wing patterns of adult male and non-mimetic and mimetic females of *Papilio polytes*. Mimicry is regulated by *H* locus and mimetic allele (*H*) is dominant over the non-mimetic allele (*h*). Males are non-mimetic and monomorphism, but females have non-mimetic and mimetic phenotypes. (b) Detailed structure of mimicry highly diversified region (HDR) in *P. polytes* (Nishikawa et al., 2015; Iijima et al., 2018). The direction of the HDR is reversed between *h* and *H* (i.e., inversion). Putative breakpoints of the HDRs are indicated by orange dotted lines. The breakpoint on the left side is located inside the 5'-untranslated region (UTR) of the *ubiquitously expressed transcript* (*UXT*) gene, and that on the right side is located just on the outer side of *doublesex* (*dsx*).

(c) Expression levels of genes within and flanking the HDR in hindwings of mimetic (*Hh*) females at the wandering stage (W) of the late last instar larvae, 2 day after pupation (P2) and 5 day after pupation (P5). The mean fragment per kilobase of transcript per million mapped reads (FPKM) values by RNA sequencing are shown with SE. Different letters indicate significant differences (Tukey post hoc test, $P<0.05$).

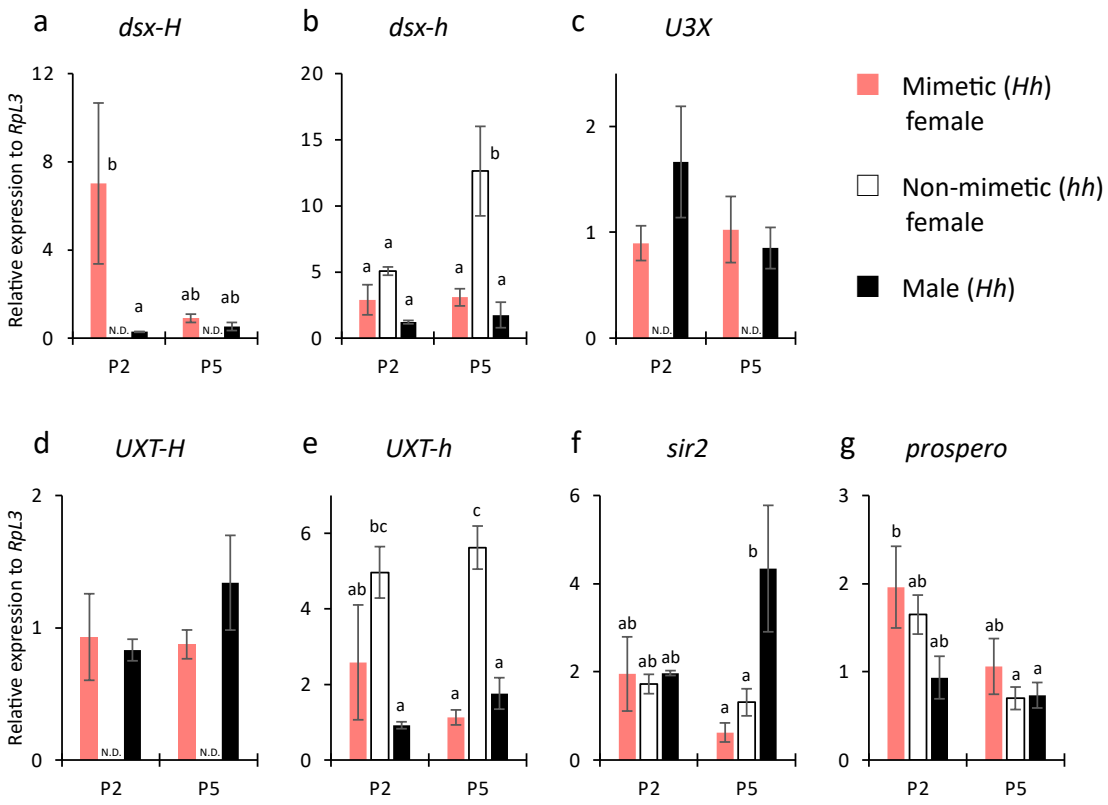


Figure 2. Gene expression levels in mimetic (*Hh*) and non-mimetic (*hh*) females and males (*Hh*) at 2 day after pupation (P2) and 5 day after pupation (P5) in *Papilio polytes*. The expression levels of *dsx-H* (a), *dsx-h* (b), *U3X* (c), *UXT-H* (d), *UXT-h* (e), *sir2* (f) and *prospero* (g) were estimated by RT-qPCR using *RpL3* as the internal control. Error bars show standard error. Different letters indicate significant differences (Tukey post hoc test, P<0.05).

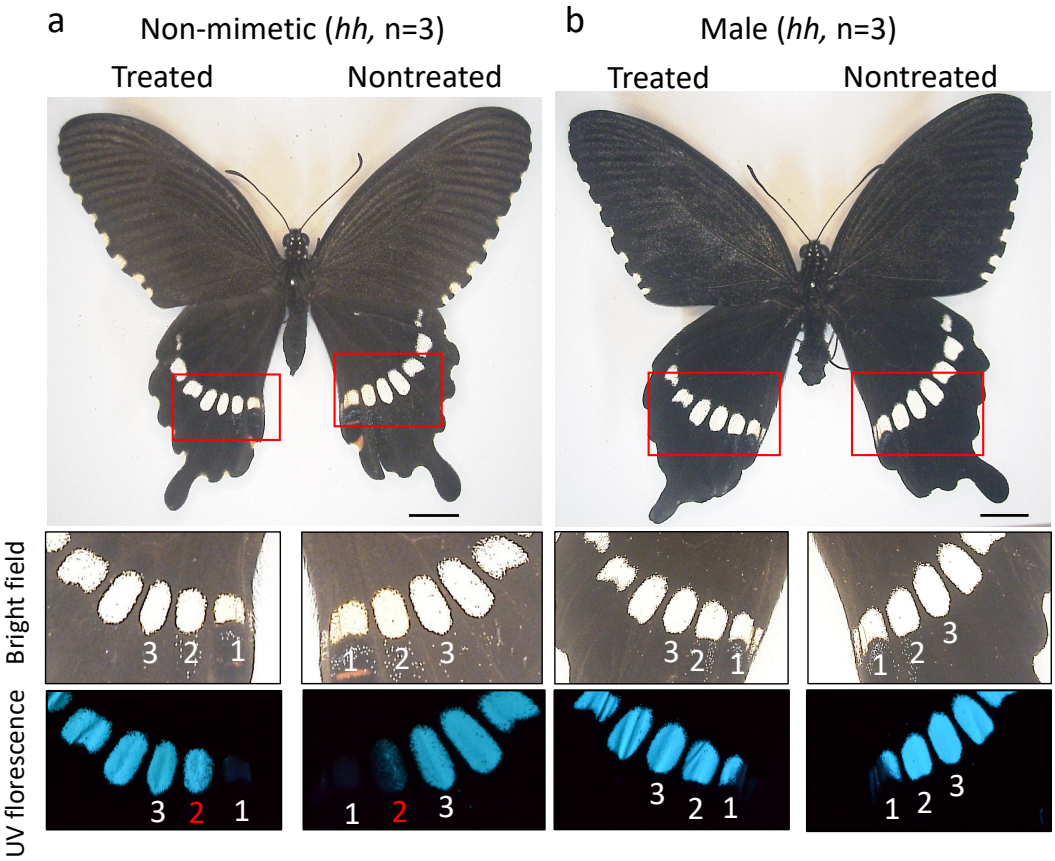


Figure 3. Knockdown of *dsx* in the left hindwings of non-mimetic (*hh*) female (a) and male (b) of *Papilio polytes*. siRNA was injected into the left pupal hindwing immediately after pupation and electroporated into the dorsal side. Pale yellow spots are numbered starting from the inside. The numbered spots in red indicate those whose UV response was altered by knockdown. In non-mimetic females (a), knockdown of *dsx* changed the second spot to produce UV fluorescence like males. Scale bars, 1cm. Supplementary Figure S3 show other replicates.

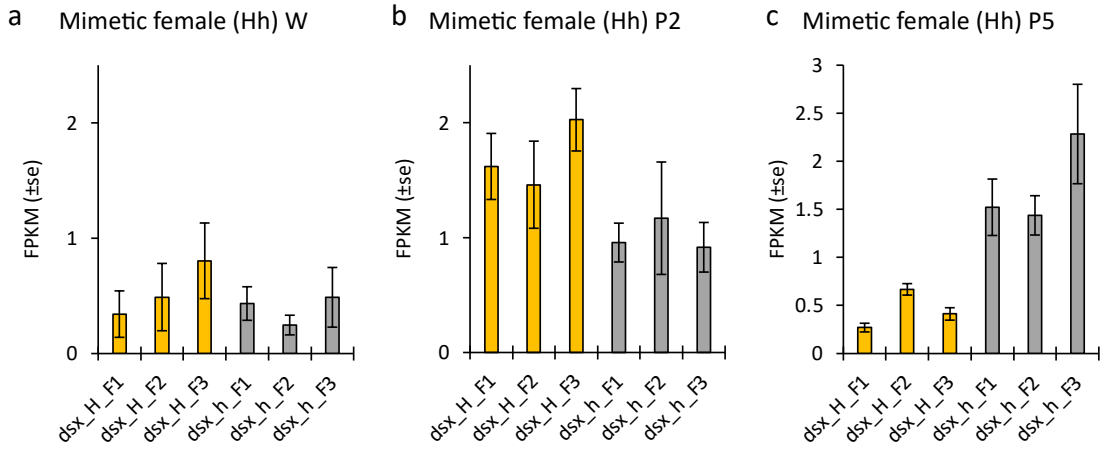


Figure 4. Gene expression levels of each *dsx* isoforms in mimetic (*Hh*) females of *Papilio polytes*. The mean fragment per kilobase of transcript per million mapped reads (FPKM) values by RNA sequencing at the wandering stage (W) of the late last instar larvae is (a), at 2 day after pupation (P2) is (b) and at 5 day after pupation (P5) is (c). Orange bars indicate the expression levels of *dsx* isoforms from mimetic (*H*) allele and gray bars indicate from non-mimetic (*h*) allele. F1, F2 and F3 means female isoform 1, 2 and 3, respectively. There was no statistically significant difference among isoforms.

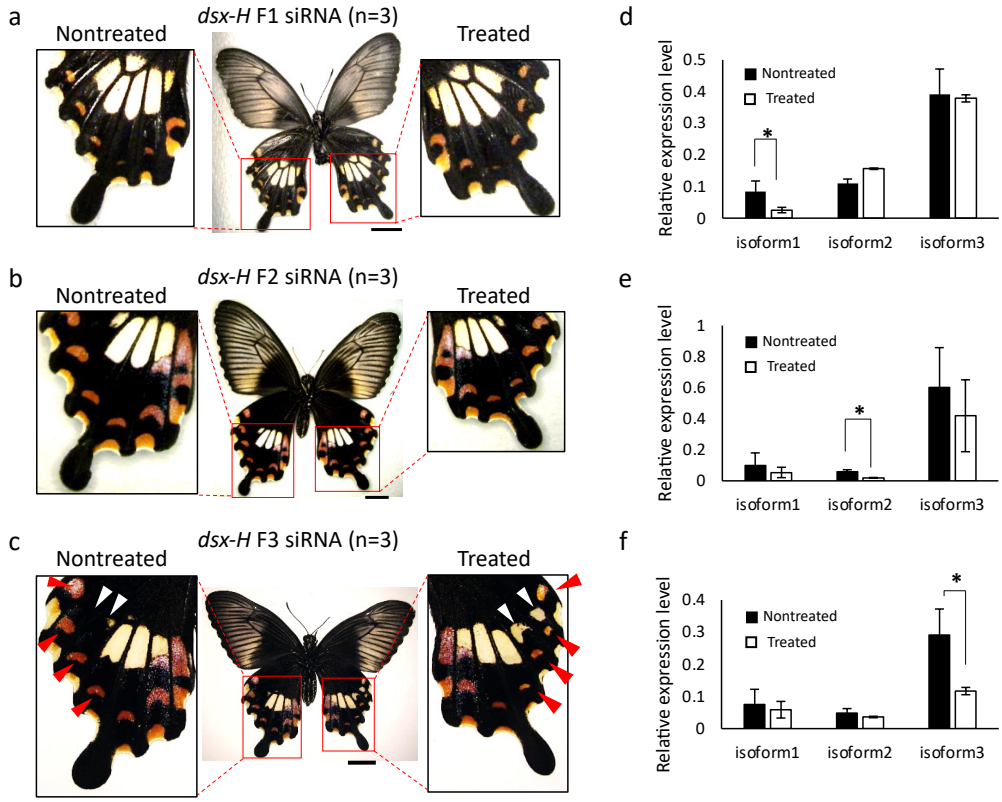


Figure 5. Knockdown of *dsx* female isoform 1 (a), 2 (b) and 3 (c) in the hindwings of mimetic (*Hh*) females of *Papilio polytes*. siRNA was injected into the left pupal hindwing immediately after pupation and electroporated into the ventral side. No phenotypic changes were observed by knockdowns of *dsx* female isoform 1 (F1) and female isoform 2 (F2) (a,b), but knockdown of *dsx* female isoform 3 (F3) changed the mimetic pattern to the similar of the non-mimetic pattern (c). Red and white arrowheads represent the changed red and pale yellow regions, respectively. Scale bars, 1 cm. Supplementary Figure S4 show other replicates. (d–f) Gene expression levels of each isoform in the knockdown wings of mimetic (*Hh*) females at 2 days after pupation. When isoform1 (F1) was knocked down, only isoform 1 (F1) was down-regulated (d), when isoform 2 (F2) was knocked down, only isoform 2 (F2) was down-regulated (e), and when isoform 3 (F3) was knocked down, only isoform 3 (F3) was down-regulated (f). Black and white bars shown the expression in nontreated and treated hindwings, respectively. We estimated the gene expression levels by RT-qPCR using *RpL3* as the internal control. Error bars show standard error of three biological replicates. *$P < 0.05$ for Student's t-test.

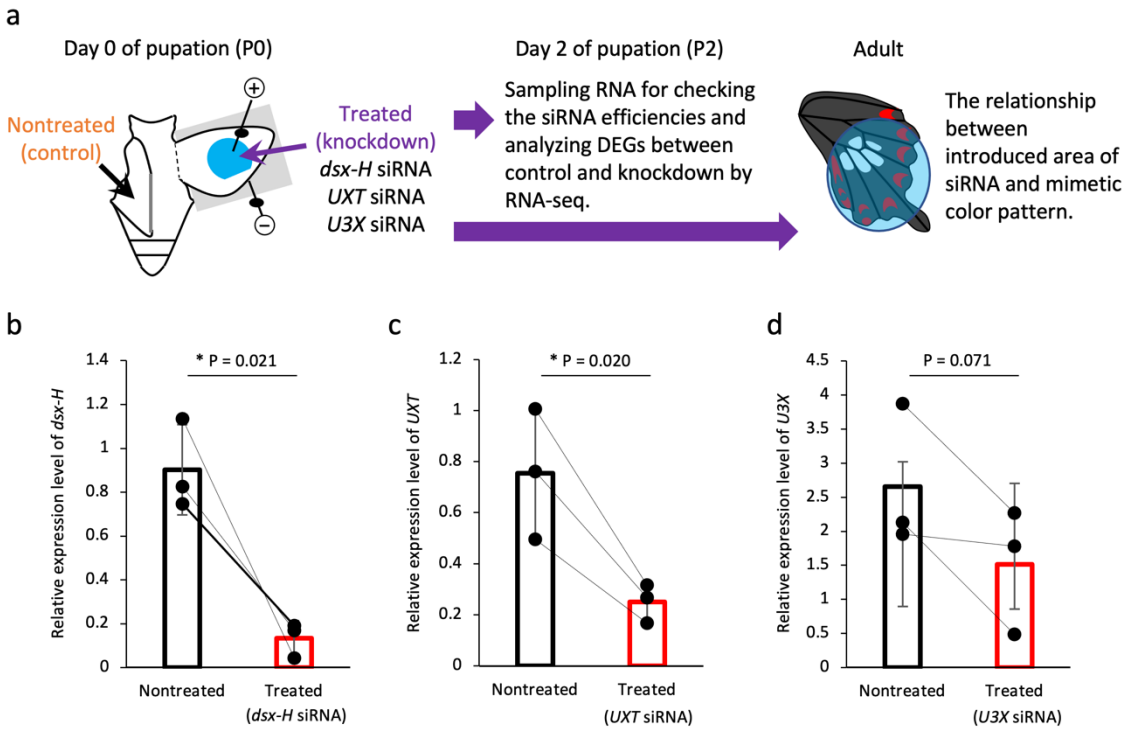


Figure 6.

(a) Scheme of functional analysis and sampling for RNA sequencing samples. *dsx-H*, *UXT* and *U3X* siRNAs were injected each into the left pupal hindwings immediately after pupation (day 0 of pupation: P0) and electroporated into the ventral side. At 2 days after pupation (P2), we sampled RNAs from nontreated (control) and treated (knockdown) hindwings.

(b–d) Measurement of knockdown effect using RT-qPCR. We compared the expression levels of *dsx-H* (b), *UXT* (c) and *U3X* (d) between untreated (white bar) and treated hindwings (red bar) by RT-qPCR using *RpL3* as an internal control. Values and error bars denote the mean and standard deviation of three biological replicates. $*P < 0.05$ for paired t-test.

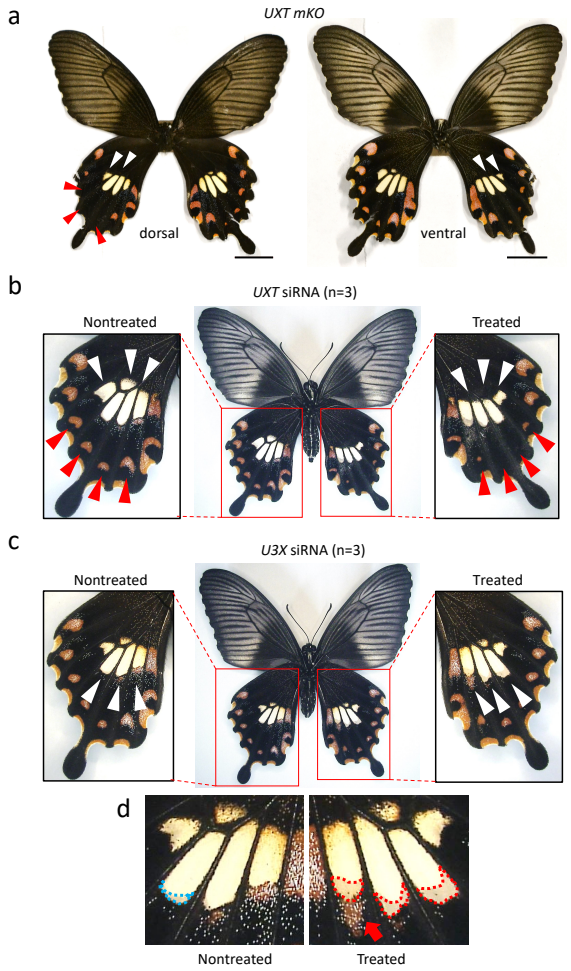


Figure 7. Functional analysis of *UXT* and *U3X* by Crispr/Cas9 and RNAi in *Papilio polytes*

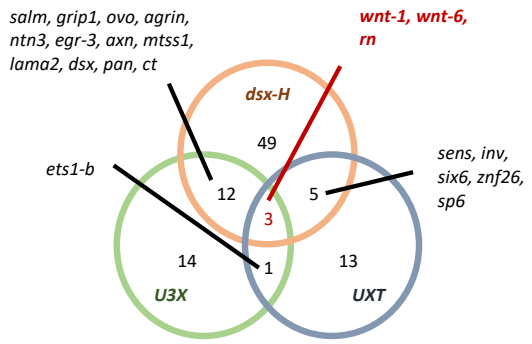
(a) Mosaic knockout of *UXT* by Crispr/Cas9. Dorsal and ventral views of one representative of the eight individuals observed are shown. Red and white arrowheads represent the changed red and pale yellow regions, respectively. In this individual, phenotypic changes were observed mainly on the left hindwing in dorsal view. Scale bars, 1 cm. Supplementary Figure S7 shows other replicates.

(b–d) Knockdown of *UXT* (b) and *U3X* (c,d) in the hindwings of mimetic (*Hh*) females. siRNA was injected into the left pupal hindwing immediately after pupation and electroporated into the ventral side. Red and white arrowheads represent the changed red and pale yellow regions, respectively. A magnified view of the pale yellow regions of *U3X* knockdown is shown in (d). The red arrow indicates the area where the red spot has expanded, and the red dotted line indicates the area where the pale yellow spot has extended. Supplementary Figures S8–9 show other replicates.

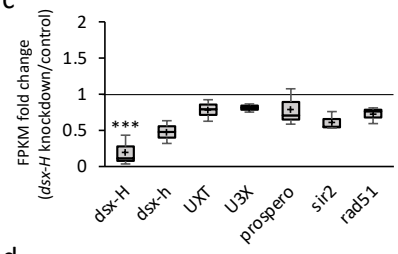
a

	dsx-H	*UXT*	*U3X*
DEGs down-regulated by knockdown	978	544	612
Transcription factor	101	24	27
Signaling ligand/receptor	29	4	10
DEGs up-regulated by knockdown	1543	1429	1010
Transcription factor	15	23	10
Signaling ligand/receptor	10	22	11

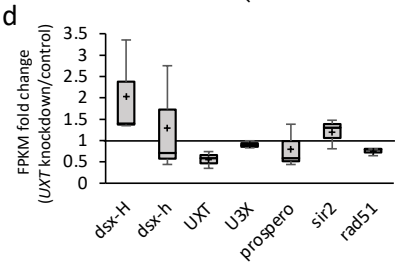
b



c



d



e

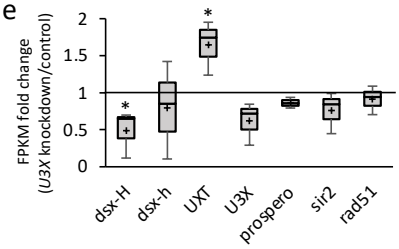
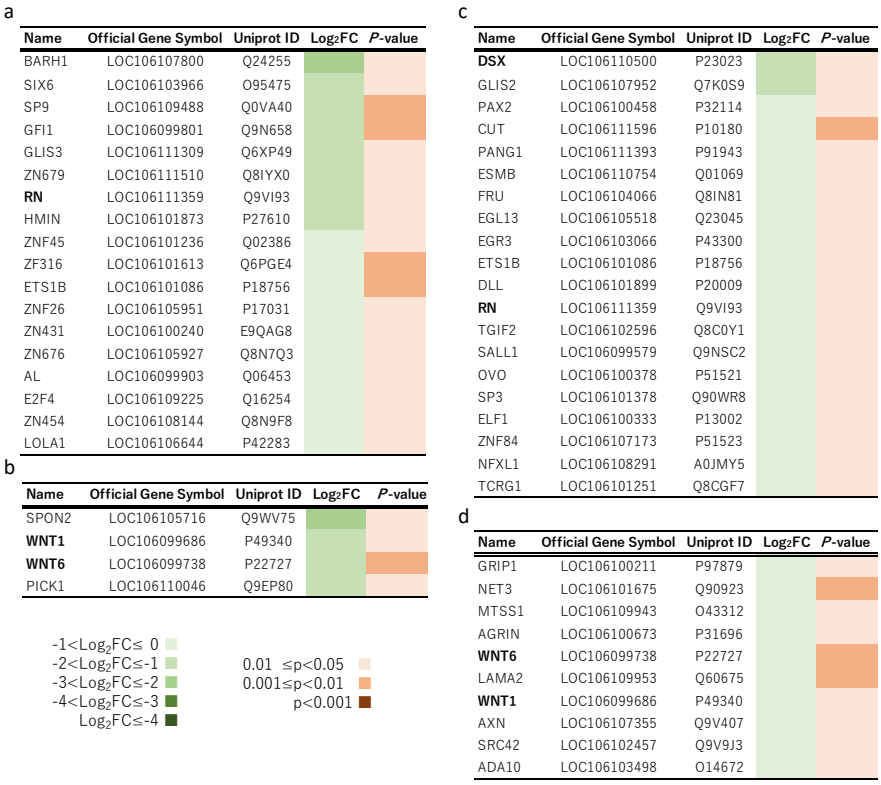


Figure 8. Comparison of gene expression levels between knockdown of *dsx-H*, *UXT* and *U3X*, and control hindwings by RNA sequencing. (a) The number of differentially expressed genes (DEGs) identified by *dsx-H*, *UXT* and *U3X* knockdowns. (b) Venn diagram depicting the abundance of DEGs ($P < 0.05$) for each comparison between three genes by untreated and siRNA-treated samples, and shows only the number of transcription factors and signal factors. (c–e) Fold Change of FPKM values (knockdown/control sides) of genes around *H* locus during knockdowns of *dsx-H* (c), *UXT* (d) and *U3X* (e). FPKM fold changes of *dsx-H*, *dsx-h*, *UXT*, *U3X*, *prospero*, *sir2*, and *rad51* are shown. The value is 1 when the FPKM values of the siRNA treated and untreated sides are equal. Wald-test, paired; ***P<0.001, *P<0.05.

Figure 9.

DEGs of transcription factors (a) and signaling factors (b) down-regulated by *dsx-H* knockdown. Sequences of DEGs obtained by RNA-seq were blasted with Uniprot, and transcription factors and signal factors were extracted using GO terms of the top hit Uniprot ID. In case multiple isoforms of the same gene were included, only the isoform with the lowest Log_2FC (Log2-fold change value = Log2 siRNA-treated FPKM–Log2 untreated FPKM) is shown, and Name is the protein name of the top hit Uniprot ID.



a

Name	Official Gene Symbol	Uniprot ID	Log2FC	*P*-value
BARH1	LOC106107800	Q24255		
SIX6	LOC106103966	O95475		
SP9	LOC106109488	Q0VA40		
GFI1	LOC106099801	Q9N658		
GLIS3	LOC106111309	Q6XP49		
ZN679	LOC106111510	Q8IYX0		
RN	LOC106111359	Q9VI93		
HMIN	LOC106101873	P27610		
ZNF45	LOC106101236	Q02386		
ZF316	LOC106101613	Q6PGE4		
ETS1B	LOC106101086	P18756		
ZNF26	LOC106105951	P17031		
ZN431	LOC106100240	E9QAG8		
ZN676	LOC106105927	Q8N7Q3		
AL	LOC106099903	Q06453		
E2F4	LOC106109225	Q16254		
ZN454	LOC106108144	Q8N9F8		
LOLA1	LOC106106644	P42283		

b

Name	Official Gene Symbol	Uniprot ID	Log2FC	*P*-value
SPON2	LOC106105716	Q9WV75		
WNT1	LOC106099686	P49340		
WNT6	LOC106099738	P22727		
PICK1	LOC106110046	Q9EP80		

c

Name	Official Gene Symbol	Uniprot ID	Log2FC	*P*-value
DSX	LOC106110500	P23023		
GLIS2	LOC106107952	Q7K0S9		
PAX2	LOC106100458	P32114		
CUT	LOC106111596	P10180		
PANG1	LOC106111393	P91943		
ESMB	LOC106110754	Q01069		
FRU	LOC106104066	Q8IN81		
EGL13	LOC106105518	Q23045		
EGR3	LOC106103066	P43300		
ETS1B	LOC106101086	P18756		
DLL	LOC106101899	P20009		
RN	LOC106111359	Q9VI93		
TGIF2	LOC106102596	Q8C0Y1		
SALL1	LOC106099579	Q9NSC2		
OVO	LOC106100378	P51521		
SP3	LOC106101378	Q90WR8		
ELF1	LOC106100333	P13002		
ZNF84	LOC106107173	P51523		
NFXL1	LOC106108291	A0JMY5		
TCRG1	LOC106101251	Q8CGF7		

d

Name	Official Gene Symbol	Uniprot ID	Log2FC	*P*-value
GRIP1	LOC106100211	P97879		
NET3	LOC106101675	Q90923		
MTSS1	LOC106109943	O43312		
AGRIN	LOC106100673	P31696		
WNT6	LOC106099738	P22727		
LAMA2	LOC106109953	Q60675		
WNT1	LOC106099686	P49340		
AXN	LOC106107355	Q9V407		
SRC42	LOC106102457	Q9V9J3		
ADA10	LOC106103498	O14672		

Figure 10.

DEGs of transcription factors and signaling factors down-regulated by *UXT* (a,b) and *U3X* (c,d) knockdown. Among the DEGs obtained by knockdown of *UXT*, transcription factors are shown in (a) and signal factors are shown in (b). Similarly, the transcription factor and signal factor of *U3X* are shown in (c) and (d), respectively.

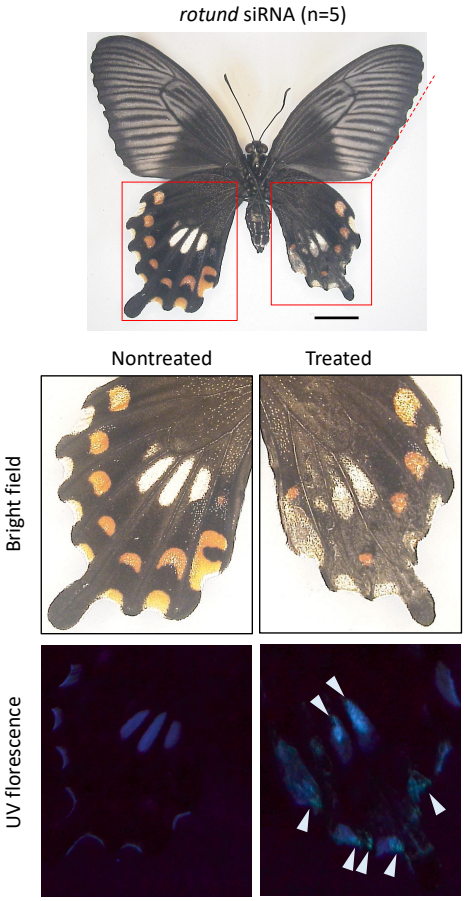


Figure 11.

Knockdown of *rotund* (*rn*) in the hindwings of mimetic (*Hh*) females of *Papilio polytes*. siRNA was injected into the left pupal hindwing immediately after pupation and electroporated into the ventral side. In UV fluorescence, white arrowheads represent the changed pale yellow regions by knockdown. Knockdown of *rn* changed the pale yellow spots to produce UV fluorescence like males. Scale bars, 1cm.Supplementary Figure S10 show other replicates.

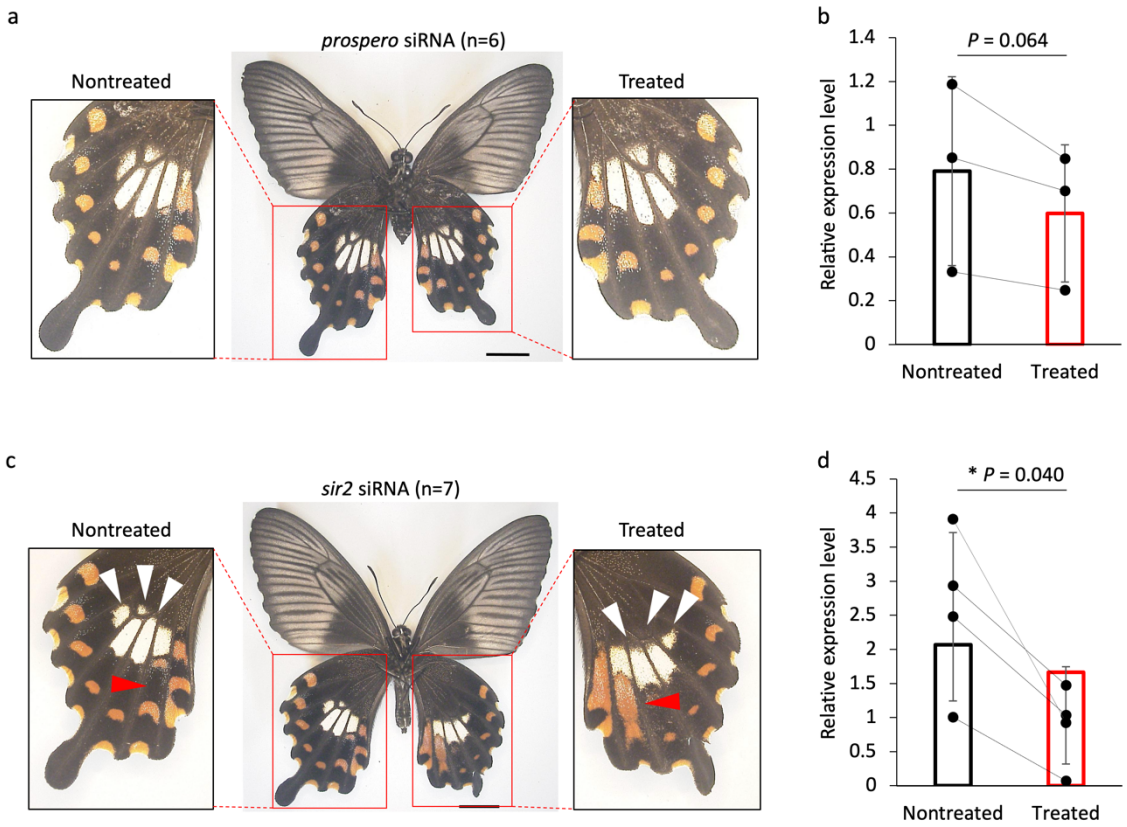


Figure 12.

Knockdown of *prospero* (a) and *sir2* (c) in the hindwings of mimetic (*Hh*) females of *Papilio polytes*. siRNA was injected into the left pupal hindwing immediately after pupation and electroporated into the ventral side. In (c), red and white arrowheads represent the changed red and pale yellow regions, respectively. (b,d) Measurement of knockdown effect using RT-qPCR. We compared the expression levels of *prospero* (d) and *sir2* (d) between untreated (white bar) and treated hindwings (red bar) by RT-qPCR using *RpL3* as an internal control. Values and error bars denote the mean and standard deviation of three biological replicates. $*P < 0.05$ for paired t-test. Supplementary Figures S11–12 show other replicates.