Part III
Developmental Genetics

Chapter 8
A Practical Guide to CRISPR/Cas9 Genome Editing in Lepidoptera

Linlin Zhang and Robert D. Reed

Abstract CRISPR/Cas9 genome editing has revolutionized functional genetic work in many organisms and is having an especially strong impact in emerging model systems. Here we summarize recent advances in applying CRISPR/Cas9 methods in Lepidoptera, with a focus on providing practical advice on the entire process of genome editing from experimental design through to genotyping. We also describe successful targeted GFP knockins that we have achieved in butterflies. Finally, we provide a complete, detailed protocol for producing targeted long deletions in butterflies.

Keywords Genome editing • Knockin • Butterfly • Transgenic • Transformation • Evo-devo

8.1 Introduction

The order Lepidoptera represents a tenth of the world's described species and includes many taxa of economic and scientific importance. Despite strong interest in this group, however, there has been a frustrating lack of progress in developing routine approaches for manipulative genetic work. While the last two decades have seen examples of transgenesis and targeted knockouts using methods like transposon insertion (Tamura et al. 2000), zinc-finger nucleases (Takasu et al. 2010; Merlin et al. 2013), and TALENs (Takasu et al. 2013; Markert et al. 2016), especially in the silk moth *Bombyx mori*, these approaches have resisted widespread application due to their laborious nature. We see two other main reasons manipulative genetics has failed to become routine in Lepidoptera. The first is that many lepidopterans are sensitive to inbreeding, and in some species it can be difficult to maintain experimental lines without special effort. The second is that lepidopterans appear to have an unusual resistance to RNAi (Terenius et al. 2011; Kolliopoulou and Swevers 2014), a method that has dramatically accelerated work in other groups of insects.

L. Zhang • R.D. Reed (✉)
Department of Ecology and Evolutionary Biology, Cornell University, 215 Tower Rd., Ithaca, NY 14853-7202, USA
e-mail: robertreed@cornell.edu

T. Sekimura, H.F. Nijhout (eds.), *Diversity and Evolution of Butterfly Wing Patterns*, DOI 10.1007/978-981-10-4956-9_8

Given this history of challenges in Lepidoptera, it is with great excitement that over the last few years we have seen an increasing number of studies that demonstrate the high efficiency of CRISPR/Cas9-mediated genome editing in this group. Our own lab began experimenting with genome editing in butterflies in 2014, and we and our collaborators have now successfully edited over 15 loci across six species, generating both targeted deletions and insertions. The purpose of this review is to briefly summarize the current state of this fast-moving field and to provide practical advice for those who would like to use this technology in their own work.

8.2 Published Examples of Cas9-Mediated Genome Editing in Lepidoptera

Between 2013 and early 2017, we identify 22 published studies applying CRISPR/Cas9 methods in Lepidoptera (Table 8.1). The earliest published reports of Cas9-mediated genome editing in Lepidoptera, from 2013 and 2014, all describe work done in *B. mori* (Wang et al. 2013; Ma et al. 2014; Wei et al. 2014) – an experimental system that benefits from a large research community that had already developed efficient methods for injection, rearing, and genotyping. To our knowledge, Wang et al. (2013) represent the first published report of Cas9-mediated genome editing in Lepidoptera and set three important precedents. First, they established the protocol that has been more or less emulated by most following studies, where single-guide RNAs (sgRNAs) are co-injected with Cas9 mRNA into early-stage embryos. Second, they demonstrated that it is possible to co-inject dual sgRNAs to produce long deletions. In this respect, the 3.5 kb deletion they produced was an important early benchmark for demonstrating the possibility of generating long deletions in Lepidoptera. Third, they showed that deletions could occur in the germ line at a high enough frequency to generate stable lines.

After Wang et al. (2013), one of the next most important technical advancements came from Ma et al. (2014), who showed that knockins could be achieved using a donor plasmid to insert a DsRed expression cassette using ~1 kb homology arms. Following this, Zhu et al. showed successful epitope tagging of *BmTUDOR-SN* gene by CRISPR/Cas9-mediated knockin in *Bombyx* cells (Zhu et al. 2015). Unfortunately, to our knowledge, these remain the only two examples of lepidopteran knockins outside of the new data we present below. The first example of Cas9 genome editing in a species besides *B. mori* was described by Li et al. (2015a), who produced deletions in three genes in the swallowtail butterfly *Papilio xuthus*. This was an important case study because it showed that the general approach used by Wang et al. (2013) in *B. mori* could be transferred to other species and still retain the same level of high efficiency. Two more notable technical advancements include the production of an 18 kb deletion in *B. mori* by Zhang et al. (2015) – the longest deletion we know of in Lepidoptera, and much longer than anything in

Table 8.1 Comparison of CRISPR genome editing in Lepidoptera

Species	References	Knockout/ knockin	Knockout strategy	Delivery	Genes targeted	Genotyping	Mosaic (%)	Germ line
Bombyx mori	Wang Y. et al. (2013)	Knockout	Small indels and long deletions	Cas9 mRNA and sgRNA	*BLOS2*	PCR	94–100%	Yes
	Ma S. et al. (2014)	Knockout/ knockin	Small indels	Plasmid of Cas9 and gRNA	*ku70*	T7E1	16.8–30.3%	Yes
	Wei W. et al. (2014)	Knockout	Small indels and long deletions	Cas9 mRNA and sgRNA	*ok/KMO/TH/tan*	Sequencing	16.7–35.0%	Yes
	Liu Y. et al. (2014)	Knockout	Small indels	Plasmid of Cas9 and gRNA	*Th/re/fl/yel-e/ kynu/e*	T7E1	5.7–18.9%	No
	Zhu L. et al. (2015)	Knockout/ knockin	Small indels	All-in-one vector	*Ku70/Ku80/Lig IV/XRCC4/XLF*	T7E1	20–80%	No
	Ling L. et al. (2015)	Knockout	Long deletions	Cas9 mRNA and sgRNA	*awd/fng*	PCR	40–61%	No
	Li Z, et al. (2015b)	Knockout	Long deletions	Cas9 mRNA and sgRNA	*EO*	PCR	60.60%	No
	Xin H. et al. (2015)	Knockout	Small indels	Cas9 mRNA and sgRNA	*sage*	Sequencing	46.67%	Yes
	Zhang Z. et al. (2015)	Knockout	Long deletions	Cas9 mRNA and sgRNA	*wnt1*	T7E1	42.5–90.6%	No
	Zeng B. et al. (2016)	Knockout	Small indels	Cas9 mRNA and sgRNA	*GFP/BLOS2*	T7E1	NA	Yes
Helicoverpa armigera	Khan S. et al. (2017)	Knockout	Small indels	Cas9 mRNA and sgRNA	*w/st/bw/ok*	T7E1	NA	Yes

(continued)

Table 8.1 (continued)

Species	References	Knockout/ knockin	Knockout strategy	Delivery	Genes targeted	Genotyping	Mosaic (%)	Germ line
Papilio xuthus	Li X, et al. (2015a)	Knockout	Small indels and long deletions	Cas9 mRNA and sgRNA	*Abd-B*/*e*/*fz*	T7E1	18.33–90.85%	No
	Zhang L. et al. (2017)	Knockout	Long deletions	Cas9 protein and sgRNA	*y*	NA	83.33%	No
	Perry M. et al. (2016)	Knockout	Long deletions	Cas9 protein and sgRNA	*y*/*ss*	Sequencing	NA	No
Vanessa cardui	Zhang L. and Reed R. (2016)	Knockout	Long deletions	Cas9 protein and sgRNA	*Ddc*/*Dll*/*spalt*	PCR and sequencing	51.7–56%	No
	Zhang L. et al. (2017)	Knockout	Long deletions	Cas9 protein and sgRNA	*y*/*b*/*e*/*yel-d*	PCR	42–80%	No
	Perry M. et al. (2016)	Knockout	Long deletions		*y*/*ss*	Sequencing	NA	No
Junonia coenia	Zhang L. and Reed R. (2016)	Knockout	Long deletions	Cas9 protein and sgRNA	*Dll*/*spalt*	PCR and sequencing	33–41%	No
	Zhang L. et al. (2017)	Knockout	Long deletions	Cas9 protein and sgRNA	*Ple*/*e*	NA	27–42%	No
Bicyclus anynana	Zhang L. et al. (2017)	Knockout	Long deletions	Cas9 mRNA/protein and sgRNA	*y*/*e*	T7E1	31–72%	No
	Beldade P. and Peralta C. (2017)	Knockout	NA	NA	*Ddc*/*e*	NA	NA	No
Danaus plexippus	Markert M. et al. (2016)	Knockout	Small and long deletions	Cas9 mRNA and sgRNA	*cry2*/*clk*	Cas9 in vitro cleavage	57.9–61.5%	Yes

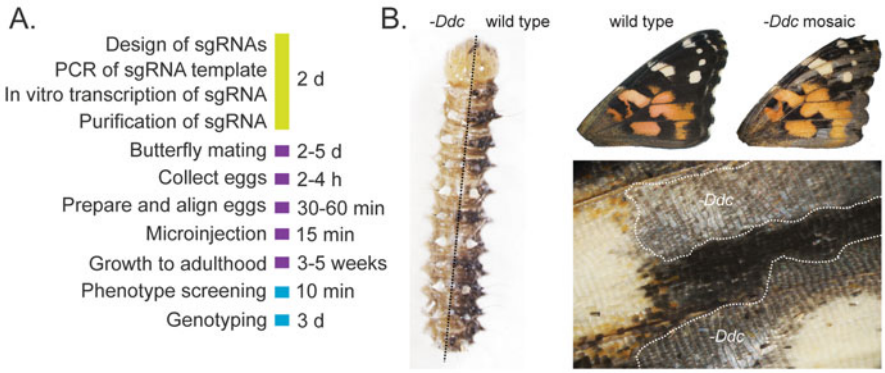

Fig. 8.1 Timeline and example outcome of G_0 CRISPR/Cas9 mosaic knockout experiments in butterflies. (**a**) Overview and timeline of mutant generation by CRISPR/Cas9 injection to butterfly embryos. (**b**) Example of larval and adult wing somatic mosaic phenotypes resulting from knockout of the melanin pigmentation gene *Ddc*

Drosophila reports we have seen – and the direct injection of recombinant Cas9 protein instead of Cas9 mRNA (Zhang and Reed 2016; Perry et al. 2016), which was an important improvement to the protocol that significantly simplifies the genome editing workflow.

Through our lab's research on butterfly wing pattern development, we have tried most of the methods described in the studies cited above, and we have gained significant experience in porting these protocols across species. We now perform targeted long deletions routinely and with a fairly high throughput. As of the end of 2016, we and our colleagues have successfully applied this general approach in six butterfly and two moth species (*Vanessa cardui*, *Junonia coenia*, *Bicyclus anynana*, *Papilio xuthus*, *Heliconius erato*, *Agraulis vanillae*, *B. mori*, and *Plodia interpunctella*), with each species requiring only minor modifications to physical aspects of egg injection protocol. As we describe below, we have also successfully achieved protein coding knockins similar to Zhu et al. (2015), although our efficiency levels remain similarly low. Below we outline the approach that we have found to be the most time- and cost-efficient and transferable between species (Fig. 8.1a).

8.3 Experimental Design

Deletions Loss-of-function deletion mutations can be generated by nonhomologous end joining (NHEJ) following double-strand breaks (DSBs). Both small indel (single cleavage) and long deletion knockout strategies (co-injection of two sgRNAs) have been employed in Lepidoptera (Table 8.1). Our lab currently favors long deletions using dual sgRNAs because it facilitates rapid screening and genotyping of mutants using PCR and regular agarose gel electrophoresis. Small indels produced by single cleavages are too small to detect

easily using normal agarose gels. Dual sgRNA deletions, however, can be tens, hundreds, or thousands of base pairs long and are easy to identify in gels. sgRNA target sites can be easily identified simply by scanning the target region for $GGN_{18}NGG$ or $N_{20}NGG$ motifs on either strand using the CasBLASTR web tool (http://www.casblastr.org/). In our experience, the relative strandedness of sgRNAs does not appear to have a significant effect on the efficiency of double sgRNA long deletion experiments. If a reference genome is available, candidate sgRNA sequences should be used for a blast search to confirm there are not multiple binding sites that may produce off-target effects. The injection mix we typically use is 200 ng/μl Cas9 and 100 ng/μl of each sgRNA – this will tend to give larger effects and is suitable for less potentially lethal loci. For targets that may result in more deleterious effects, we recommend decreasing the amount of Cas9/sgRNA mix and injecting later in embryonic development to induce fewer and smaller clones. We have been able to induce mosaic mutants (e.g., Fig. 8.1b) using as low as 20 ng/μl Cas9 and 50 ng/μl of each sgRNA in different butterfly species.

Insertions CRISPR/Cas9-induced-site-specific DSBs can be precisely repaired by homology-directed recombination repair (HDR). The HDR pathway can replace an endogenous genome segment with a homologous donor sequence and can thus be used for knockin of foreign DNA into a selected genomic locus. To our knowledge, there are only two published examples of this approach in Lepidoptera, both of which

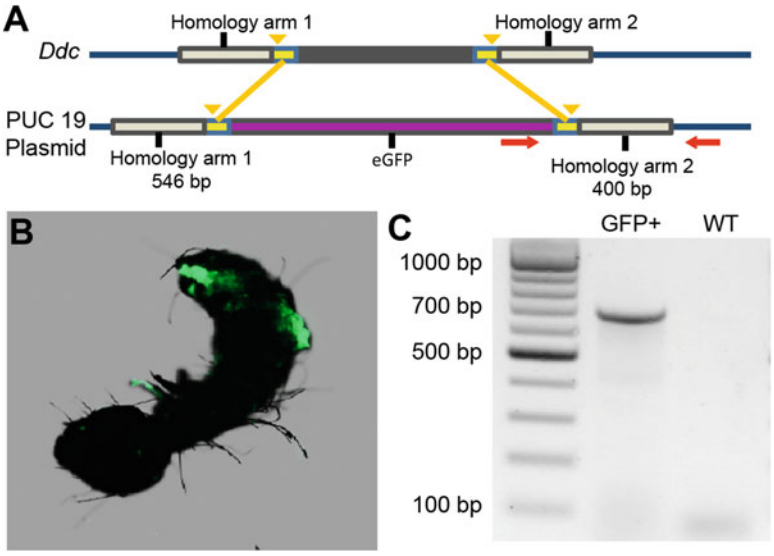

Fig. 8.2 Knockin tagging of the *Ddc* gene in *V. cardui*. (**a**) Schematic overview of the *Ddc* locus and donor construct consisting of homology arms, EGFP coding region, and genotyping primers. PAM regions are marked by *yellow*, cut sites are marked by *yellow arrowhead*, and genotyping primers are marked by *red arrows*. (**b**) Strong mosaic EGFP expression in knockin caterpillars visualized by fluorescent microscopy. (**c**) PCR analysis demonstrates using the primers in (**a**) showing the insertion of EGFP into the *Ddc* coding region

are in *B. mori* (Ma et al. 2014; Zhu et al. 2015). To test the feasibility of this approach in butterflies, we sought to insert an in-frame EGFP coding sequence into to the *V. cardui dopa decarboxylase* (*Ddc*) locus using a donor plasmid containing the EGFP coding sequences and homologous arms matching endogenous sequences flanking the Cas9 cut sites (Fig. 8.2a). As shown in Fig. 8.2b, EGFP fluorescence was detected in clones in the mutant caterpillars. In addition, PCR analysis with primers flanking the 5′ and 3′ junctions of the integration shows a clear band in mutants, but not in wild type (Fig. 8.2c). Our results show that donor DNA with ~500 bp homology arms is sufficient for precise in-frame knockins. Compared to NHEJ-mediated high efficiency knockouts (69% in the case of *V. cardui Ddc* deletion knockouts (Zhang and Reed 2016)), the rate of HDR-mediated targeted integration is low, at ~3% in our most recent trials. It has been shown that knocking out factors in the NHEJ pathway can enhance the HDR pathway and increase gene targeting efficiency in *Bombyx* (Ma et al. 2014; Zhu et al. 2015). Some Cas9-mediated homology-independent knockin approaches have shown higher efficiency rates in zebrafish (Auer et al. 2014) and human cell lines (He et al. 2016), suggesting NHEJ repair may provide an alternate strategy to improve incorporation of donor DNA in Lepidoptera.

8.4 Embryo Injection

When adapting CRISPR/Cas-9 genome editing to a new species, the greatest technical challenges we face typically lie in optimizing the injection protocol. The main reason for this is that the eggs of different species can be quite different in terms of how difficult they are to puncture with a glass needle and how they react to mechanical injection, especially in terms of internal pressure and postinjection backflow.

Injection Needles Proper needle shape is critical for achieving successful egg injections in Lepidoptera. In our experience some taxa like *Heliconius* spp. have very soft, easy-to-inject eggs that present very few problems and are relatively robust to variation in needle shape. Many lepidopterans, however, have difficult-to-puncture eggs with high internal pressure. The key challenge for these eggs is to use needles that are strong enough to penetrate tough eggshells but are not so wide as to weaken pressure balance or destroy embryos. For instance, needles that are too long and narrow can break easily when used on tough eggs and will clog at a high frequency. Conversely, needles that have a very wide diameter will tend to have problems with pressure loss and backflow. We recommend the needle shape shown in Fig. 8.3a which is characterized by a short rapid taper to a fine point. We have found that this shape provides enough strength to puncture fairly tough eggs, yet is relatively resistant to clogging and pressurization problems. Our initial attempts at pulling needles like this with a traditional gravity needle puller failed. We now pull our needles using a velocity-sensitive Sutter P-97 programmable needle puller, which works very well for crafting nuanced needle shapes. We currently prefer to

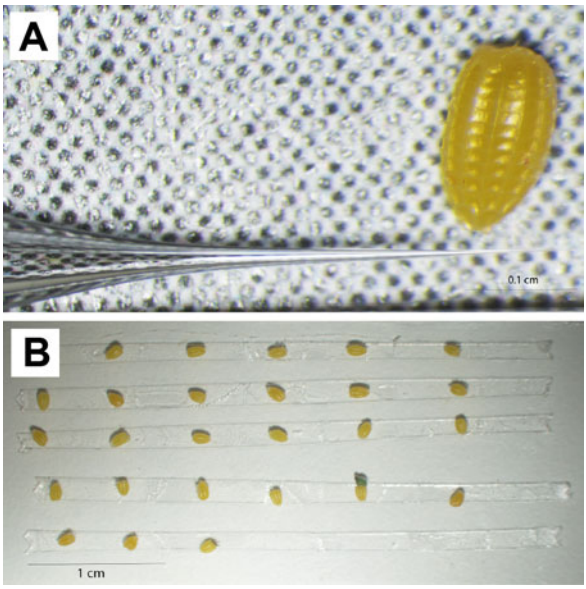

Fig. 8.3 Needle shape and egg arrangement for butterfly embryo injections. (**a**) The injection needle shape we prefer has a steep taper and a relatively large orifice. Here a preferred needle is shown next to a *Heliconius* egg. (**b**) An example of arranging *Heliconius* eggs on double-sided tape on a microscope slide just before injection

use Sutter Instrument 0.5 mm fire-polished glass capillary needles (Sutter BF-100-50-10) and 3 mm square box heating filaments (Sutter FB330B). Although settings will vary by instrument and filament, we use a single-cycle program on our puller with parameters HEAT 537, PULL strength 77, VELOCITY (trip point) 16, and TIME mode (cooling) 60. Among these parameters, the HEAT value has to be adjusted relative to the RAMP value, which is specific to certain instruments – different pullers can produce slightly different needle shapes even with the same parameter setting. We provide our settings as a starting point for other users to work toward optimizing production of needles with a steep taper and a large orifice as shown in Fig. 8.3a.

Egg Treatment Egg treatment is different for eggs from different taxa. For species with soft eggs like *Heliconius*, *Agraulis*, and *Danaus*, freshly collected eggs can be immediately arranged on double-sided adhesive tape on a microscope slide (Fig. 8.3b) and injected. For those eggs with relatively soft chorion but high pressure, like *V. cardui*, collected eggs should be arranged on a slide and then kept in a desiccation chamber for 15 min before injection. We use a simple sealed petri dish filled with desiccant for this purpose. For species with thick-shelled eggs like *J. coenia*, we recommend that eggs be dipped in 5% benzalkonium chloride (Sigma-Aldrich, St Louis, MO, USA) for 90s to soften the chorion and then washed in water for 2 min before mounting on microscope slide. We also tried treatment with 50% bleach solution to soften eggs; however, this significantly reduced the hatch rate. Softened eggs can then be dried in a desiccation chamber for 15 min and injected.

Injection Timing In all published cases we are aware of, injections of sgRNA and Cas9 (either mRNA or recombinant protein) were completed between 20 m and 4 h

after oviposition, when embryos are presumed to be in an acellular syncytial state. Most of our injecting experience has been in eggs 1–3 h old. Although we have not rigorously quantified this effect, after extensive work with pigmentation genes in *V. cardui*, we found that injecting earlier (e.g., at 1 h) typically produces more and larger mutant clones compared to injection performed later (e.g., at 4 h). This is consistent with previous studies that have found a higher deletion frequency when embryos are injected at earlier versus later stages (Li et al. 2015a). Thus, for most of our deletion experiments, we aim to inject ~1–2 h after oviposition. If we expect that deletion of the locus will have a strongly deleterious or embryonic lethal effect, we will begin by injecting at 2–4 h to decrease the magnitude of somatic deletions.

Egg Injection The key concern during injection is to minimize damage as much as possible. An optimum angle for needle insertion is about 30°–40° in our experience. We prefer to use a Narishige MM-3 micromanipulator for full three-dimensional control of the needle during injection. In the butterfly species we have worked with, the location of injection does not seem to have a major impact on editing efficiency, although in *V. cardui* we get a slightly higher survival rate by injecting into the side near the base of the egg. Proper positive balance pressure is critical for successful injection. Users should adjust balance pressure to a point where the needle is just able to retain the solution. Prior to any egg injection, adjust the injection pressure and time to ensure the flowing droplet is visible when pressing the injector's footswitch. We have worked extensively with two different injectors: a Harvard apparatus PLI-100 Pico-Injector and a Narishige IM 300 Microinjector. In our experience, PLI-100 Pico-Injector has better sensitivity in terms of balance pressure, which is very important for species with high-pressure eggs like *V. cardui* and *J. coenia*. The IM 300 does not perform as well with these eggs. The other two injectors we know of that also work well for butterfly eggs are Eppendorf FemtoJet microinjector and Drummond Nanoject III. We use 10 psi injection pressure and 0.5 psi balance pressure for soft-shelled eggs with the Narishige IM 300 injector and 20 psi injection pressure and 0.8 psi balance pressure for *V. cardui* and *J. coenia* eggs with the PLI-100 Pico-Injector. After injection we maintain slides with the injected eggs in a petri dish and move larvae to their rearing containers immediately upon emergence.

8.5 Interpreting Somatic Mosaics

While several studies have been published that describe the germ line transmission of edited alleles in *B. mori*, thus far most studies in Lepidoptera have focused on interpreting deletion phenotypes in G_0 somatic mosaics. Maintaining edited genetic lines is necessary for looking at the homozygous effects of specific edited alleles and will also be essential for a future generation of more sophisticated knockin studies. Maintaining edited lines presents a few challenges in Lepidoptera, however. First, the deletion phenotypes of many interesting genes would likely be embryonic lethal. For example, our lab has thus far been unsuccessful in efforts to produce living larvae with *wingless* or *Notch* coding region deletions, which is

unsurprising because these genes are known to be essential for early embryonic development in insects. For loci like these, we can confirm deletions by PCR and sequencing, but all embryos with deletions die before or shortly after hatching. Second, based on our experience with inbreeding attempts in *Heliconius* spp., *V. cardui*, and *J. coenia*, and through discussions with colleagues working in other systems, it is clear that many lepidopterans are sensitive to inbreeding, and lines will die out quickly unless fairly large stocks are kept. Large stocks then make it more difficult to identify individuals with specific genotypes. So while maintaining lines is possible in many lab-adapted species, it is not always a trivial endeavor.

Because of the challenges posed by the embryonic lethality of many target genes, along with the difficulty of maintaining and genotyping edited lines, most of our attention has focused on analysis of mosaic G_0 phenotypes. One obvious advantage of focusing on somatic mosaics is that data can be collected in a single generation. Another advantage is that the phenotypic effects of lesions are limited to the subset of cell lineages (clones) hosting deletion alleles, thus reducing the deleterious effects of many deletions. Because of their clear phenotypes, knockout work on melanin pigmentation genes has allowed a very useful visual demonstration of the nature of somatic mosaicism in injected animals. Our work on eight pigmentation genes across several butterfly species (Zhang et al. 2017) has allowed us some general insights into work with mosaics. First, as described above, we found a loose association between the number and size of clones and the timing of injection, where earlier injections with higher concentrations of Cas9/sgRNA complexes tend to produce larger clones. We have not attempted to quantify this effect, but across replicated experiments, our tentative conclusion is that this is a real and consistent phenomenon. This is important because it gives rough control over the strength of a phenotype and can thus be important for trying to get small non-deleterious clones for an otherwise lethal gene. Conversely, by injecting at very early stages to knock out minimally pleiotropic genes, we can often produce animals with very large clones, such as entire wings.

One challenge of working with somatic mosaics lies in detecting and interpreting more subtle phenotypes. Most of the phenotypes published to date, such as loss of wing pattern features like eyespots or production of discolored patches, are fairly obvious and/or far outside the range of natural variation. Without having a dramatic phenotype or independent clone boundary marker, however, minor or highly localized effects can be difficult to differentiate from natural variation. It is possible that quantitative image analysis approaches could address this issue, although we are unaware of published examples of this. In our own work, we have relied on two main criteria to validate putative deletion phenotypes: (1) replicates, which, of course, are useful for increasing confidence (we typically aim for a minimum of three, although the number of replicates required to make a particular inference is somewhat arbitrary and there is no standard), and (2) asymmetry, which is perhaps the most powerful criterion for inferring deletion phenotypes. Because natural variation is ordinarily symmetrical, strongly asymmetric phenotypes are best explained by left/right variation in clonal mosaics.

8.6 Genotyping

To validate that genome editing is occurring as expected at the appropriate locus, it is necessary to perform genotyping on experimental animals. We have found that the simplest and most robust genotyping approach is to design PCR primers flanking the deletion sites and then to compare PCR product sizes between wild-type and experimental animals. We recommend that genotyping amplicons cover less than 1.5 kb and be at least 100 bp outside of the closest sgRNA site to allow proper band size resolution and detection of large deletions. This approach works best for long deletions produced by double sgRNAs – indels induced by repair of a single Cas9 cut site will usually be too small to detect by PCR alone. For this reason, our lab always uses double sgRNAs to produce deletions. These PCR products can also be cloned and sequenced for further validation, as well as to better characterize the diversity and nature of deletion alleles. If a single sgRNA is used, it is likely that deletion alleles will need to be sequenced to confirm lesions. To genotype insertions, PCR primers flanking the insertion site may be used similarly, or one may also use a primer inside the transgene (e.g., Fig. 8.2c).

A current challenge in genotyping edited animals is the lack of tools to rigorously confirm specific deletion alleles in specific cell populations. First, there is the physical problem of isolating a population of cells representing a single pure clone. To our knowledge, this has not been done in insects outside of using transgenic cell sorting methods (Böttcher et al. 2014). Even carefully dissected presumptive clones cannot be assumed to be pure clonal cell populations. Indeed, to our knowledge there is not yet a practical method developed to firmly associate specific alleles with specific phenotypes. This challenge also makes it difficult to decisively confirm whether a clone is monoallelic (i.e., has a single edited allele) or biallelic (i.e., has two edited alleles), thus making it difficult to infer dosage effects without additional information. Therefore, even though some previous studies present DNA sequences of edited alleles isolated from tissues including cells with deletion phenotypes (e.g., whole embryos), none of these studies rigorously associate individual alleles with specific clones because it cannot be ruled out that the tissue samples maybe have contained multiple monoallelic or bialleleic clones. A second challenge for genotyping specific clones is that some tissues of special interest, such as adult cuticle structures, including wing scales, do not have genomic DNA of sufficient quality to permit straightforward PCR genotyping, especially for longer amplicons. Thus, even if methods become available for isolating specific clone populations, there will still be limitations when dealing with some tissue types. Given the challenges outlined above, readers should understand that most genotyping to date should be seen as a validation of the experimental approach (editing accuracy and efficiency) and not necessarily as decisive confirmation that specific alleles underlie a certain clone phenotype.

8.7 Future Prospects

CRISPR/Cas9 genome editing is rapidly revolutionizing genetic work in Lepidoptera, as it is across all of biology. It is now fairly straightforward to quickly and cheaply induce long, targeted deletions in virtually any species that can be reared in captivity. Published reports to date have focused on producing deletions in gene coding regions; however, we anticipate there will be significant interest in also applying long deletion approaches to test the function of noncoding regulatory regions, especially now that *cis*-regulatory elements can be functionally annotated with high resolution, thanks to methods like ChIP-seq (Lewis et al. 2016). Pilot work shown here and elsewhere also demonstrates that targeted insertions are possible as well, thus promising even further developments on the near horizon such as protein tagging, reporter constructs, and tissue-specific expression constructs. Right now the main challenge with knockin strategies is the relatively low efficiency rate, although newer technologies such as NHEJ mediated knockin (Auer et al. 2014) promise to dramatically improve this. Perhaps the most exciting thing about CRISPR-associated genome editing approaches, though, is the straightforward portability of the technology between species. This is truly an exciting time to be a comparative biologist.

Acknowledgments We thank Arnaud Martin for extensive discussions regarding genome editing methods in butterflies, Joseph Fetcho for early assistance with needle pulling and injection procedures, and Katie Rondem for helpful comments on the manuscript. This work was funded by the US National Science Foundation grant DEB-1354318.

Appendix: A Detailed Example of CRISPR/Cas9 Genome Editing in the Painted Lady Butterfly *V. cardui*

The following procedure provides guidelines to generate genomic deletions in the butterfly *V. cardui* using the CRISPR/Cas9 nuclease system. This protocol includes a specific example of the Reed Lab's work deleting the melanin pigmentation pathway gene *Ddc* as previously reported (Zhang and Reed 2016).

Target Design

No genome reference was available for *V. cardui* when we first began our experiment, so we used a transcriptome assembly (Zhang et al. 2017) to identify sequences of the *Ddc* coding region. Primers GCCAGATGATAAGAGGAGGTT AAG and GCAGTAGCCTTTACTTCCTCCCAG were designed to amplify and sequence the target region of the genome, and exon-intron boundaries were inferred

by comparing genomic and cDNA sequences. We recommend designing target sites at exons because they are more conserved than introns and therefore provide more predictably consistent matches between sgRNAs and genomic targets. We design sgRNAs by scanning for $GGN_{18}NGG$ or $N_{20}NGG$ pattern on the sense or antisense strand of the DNA. Target sequences GGAGTACCGTTACCTGATGA**AGG** and **CCT**CTCTACTTGAAACACFACCA (PAM sequences underlined) were designed to excise a region of 131 bp spanning the functional domains of the DDC enzyme. sgRNA oligos containing T7 promoter, target sequences, and sgRNA backbone were synthesized by a commercial supplier (Integrated DNA Technologies, Inc.). Of note, the PAM sequence is not included in the CRISPR forward primer.

CRISPR forward oligos:

Ddc sgRNA1: GAAATTAATACGACTCACTATA**GGGATCAGCTTTCGTCT GCC**GTTTTAGAGCTAGAAATAGC

Ddc sgRNA2: GAAATTAATACGACTCACTATA**GGAGTACCGTTACCTGA TGA**GTTTTA GAGCTAGAAATAGC

CRISPR universal oligo: AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTG ATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC

sgRNA Production

sgRNA Template Generation

- With the oligos generated in the preceding step, use High-Fidelity DNA Polymerase PCR Mix (NEB, Cat No. M0530) to generate the template for each sgRNA with CRISPR forward and reverse oligos. We recommend using DEPC-free nuclease-free water (Ambion, Cat No. AM9938).

PCR Reaction		PCR program
High-Fidelity DNA Polymerases PCR	50 μl	
Mix	5 μl	98 °C for 30 s
CRISPR forward oligo (10 μM)	5 μl	35 cycles (98 °C for 10 s; 60 °C for 30 s; 72 °C for 15 s)
CRISPR universal oligo (10 μM)	40 μl	72 °C for 10 min
Nuclease-free water		4 °C hold

- Purify the PCR reaction with MinElute PCR Purification Kit (Qiagen, Cat No. 28004) following the kit instructions and eluting in 15 μl nuclease-free water.
- Dilute 1 μl of this reaction with 9 μl nuclease-free water, and then run on a gel and a fluorometer (e.g., Qubit) to confirm purity, integrity, fragment length, and yield. It is also possible to use gel extraction at this stage if nonspecific products are present.

- The expected size should be around 100 bp, and the expected yield should be around 200 ng/ul.

In Vitro Transcription (IVT)

- Generate sgRNAs by in vitro transcription of the sgRNA PCR template using the T7 MEGAscript Kit (Ambion, Cat. No. AM1334). When producing and handling RNA, it is important to wear gloves and clean equipment and benches with detergent prior to use to avoid RNAse contamination. Pipette tips with filters can also be beneficial to prevent contamination from pipettes.

IVT reaction mix		Incubation and purification
ATP	2 μl	
CTP	2 μl	37 °C overnight incubation
GTP	2 μl	Add 1 μl Turbo DNAse and incubate for 15 min at 37 °C
UTP	2 μl	Add 115 μl ddH_2O and 15 μl ammonium acetate stop solution
10 × reaction buffer	2 μl	
Template	2 μl	
T7 Enzyme Mix	2 μl	
Nuclease-free water	Up to 20 μl	

- Extract sgRNA by adding 150 μl phenol:chloroform:isoamyl alcohol (25:24:1) at pH 6.7 (Sigma-Aldrich, Cat No. P2069), and vortex thoroughly for 30 s.
- Separate phases by centrifugation at 10,000 × g for 3 min at room temperature, and remove the upper phase to a fresh tube.
- Precipitate the RNA by addition of an equal volume (150 μl) of cold isopropanol (Sigma-Aldrich, Cat No. I9516).
- Mix thoroughly, and incubate at −20 °C for greater than 2 h (can be left overnight).
- Collect RNA by centrifugation at 17,000 × g for 30 min at 4 °C.
- Wash pellet twice in 0.5 ml room temperature fresh made 70% ethanol, centrifuging at 17,000 × g for 3 min at 4 °C between each wash.
- Remove the remaining liquid and dry RNA pellet for 3 min at room temperature.
- Resuspend in 30 ul nuclease-free water.
- Measure concentration on a Qubit. The expected concentration should be around 2 μg/ul; sgRNAs can be stored at −80 °C.
- MEGAclear™ Transcription Clean-Up Kit (Ambion, Cat No. AM1908) also works very well for sgRNA purification.

Cas9 Production

Cas9 is typically provided by injection of a plasmid, mRNA, or recombinant protein. We have tried both Cas9 mRNA and protein injections, and both yield similarly efficient mutation rates in butterflies. However, we recommend using commercially available Cas9 protein (PNA Bio, Cat No. #CP01) because it is more stable than Cas9 mRNA and is easier and faster to use.

- Cas9 mRNA is generated by in vitro transcription of the linearized MLM3613 (Addgene plasmid 42,251) plasmid template. The mMessage mMachine T7 Kit (Ambion, Cat No. AM1344) is used to perform in vitro transcription with T7 RNA polymerase, followed by in vitro polyadenylation with the PolyA Tailing Kit (Ambion, Cat No. AM1350). An Agilent Bioanalyzer, or similar instrument, should be used to check the size and integrity of Cas9 mRNA. Note that Cas9 mRNA can show some degree of degradation yet still produce fairly efficient results.

Egg Injection and Survivor Ratio Calculation

- Collect eggs for 2–4 h by placing a host plant leaf into the butterfly cage.
- For thick chorion eggs (e.g., *J. coenia*), dip eggs in 5% benzalkonium chloride for 90 s.
- Cut double-sided tape into several thin strips and fix them to a glass slide.
- Use a paintbrush to line the eggs onto the double-sided tape.
- For high-pressure eggs (e.g., *V. cardui* or *J. coenia*), place the slide in a desiccation chamber for 15 min before injection.
- Mix Cas9 and CRISPR sgRNAs prior to microinjection.

Injection mix		Incubation
Cas9 mRNA or protein (1 μg/μl)	1 μl	Incubate on ice for 20 min
CRISPR sgRNA1 (375 ng/μl)	1 μl	
CRISPR sgRNA2 (375 ng/μl)	1 μl	
Nuclease-free water	2 μl	

- Break the closed tip of the needle with an optimum angle about 30°–40°.
- Load the needle with 0.5 μl injection mix by capillary action or by using by Eppendorf™ Femtotips Microloader Tips (Eppendorf, Cat No. E5242956003).
- One by one inject the eggs with the injector.
- Generally, higher amounts of sgRNA and Cas9 protein will increase mutation rate and decrease egg survival (hatch rate).

Genotyping for Modification

- In order to investigate the efficiency of CRISPR-/Cas9-mediated *Ddc* knockout, we randomly surveyed 81 first instar caterpillars. DNA was extracted according to Bassett et al. (2013) to confirm CRISPR/Cas9 lesions. Generally, place one caterpillar in a PCR tube and mash the caterpillar for 30 s with a pipette tip in 50 µl of squishing butter (10 mM Tris-HCl, pH 8.2, 1 mM EDTA, 25 mM NaCl, 200 µg/ml proteinase K). Incubate at 37 °C for 30 min, inactivate the proteinase K by heating to 95 °C for 2 min, and store in −20 °C for PCR genotyping. Genotyping can also be done with adult butterfly leg DNA by using proteinase K in digestion buffer. We typically use QIAamp DNA Mini Kit (Qiagen, Cat No. 51304) for DNA extraction when genotyping from muscle tissue.
- Design genotyping primers outside of the target region. For *Ddc*, genotyping forward (GCTGGATCAGCTATCGTCT) and reverse primers (GCAGTAGCCTTTACTTCCTCCCAG) were designed to produce a 584 bp PCR fragment in wild-type individuals.
- Mix PCR reagents. PCR fragments containing two sgRNA target sites are expected to produce smaller mutant bands than wild type.

PCR reaction for genotyping		PCR program
Taq DNA Polymerases PCR Mix (NEB)	12.5 µl	98 °C for 1 min
Genotyping F primer (10 µM)	1 µl	35 cycles (98 °C for 10 s; 55 °C for 30 s; 72 °C for 40 s)
Genotyping R primer (10 µM)	1 µl	72 °C for 10 min
DNA template	1 µl	4 °C hold
Nuclease-free water	9.5 µl	

- Recover mutant bands by gel extraction using MinElute Gel Extraction Kit (Qiagen, Cat No. 28604).
- Ligate recovered DNA fragment to T4 vector for TA cloning using a TA cloning kit (Invitrogen, Cat No. K202020).
- Extract plasmid with mutant DNA fragment using QIAprep Miniprep Kit (Qiagen, Cat No. 27104).
- Sequence plasmids and align mutant sequences to wild-type sequences to confirm deletions (Fig. 8.1a in Zhang and Reed, 2016).

References

Auer TO, Duroure K, De Cian A, Concordet J-P, Del Bene F (2014) Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. Genome Res 24:142–153

Bassett AR, Tibbit C, Ponting CP, Liu J-L (2013) Highly efficient targeted mutagenesis of *Drosophila* with the CRISPR/Cas9 system. Cell Rep 4:220–228. doi:10.1016/j.celrep.2013.06.020

Beldade P, Peralta CM (2017) Developmental and evolutionary mechanisms shaping butterfly eyespots. Curr Opin Insect Sci 19:22–29

Böttcher R et al (2014) Efficient chromosomal gene modification with CRISPR/cas9 and PCR-based homologous recombination donors in cultured *Drosophila* cells. Nucleic Acids Res 42:e89–e89

He X et al (2016) Knock-in of large reporter genes in human cells via CRISPR/Cas9-induced homology-dependent and independent DNA repair. Nucleic Acids Res gkw064

Khan SA, Reichelt M, Heckel DG (2017) Functional analysis of the ABCs of eye color in *Helicoverpa armigera* with CRISPR/Cas9-induced mutations. Sci Rep 7

Kolliopoulou A, Swevers L (2014) Recent progress in RNAi research in Lepidoptera: intracellular machinery, antiviral immune response and prospects for insect pest control. Curr Opin Insect Sci 6:28–34

Lewis JJ, van der Burg KR, Mazo-Vargas A, Reed RD (2016) ChIP-Seq-annotated *Heliconius erato* genome highlights patterns of cis-regulatory evolution in Lepidoptera. Cell Rep 16:2855–2863

Li X et al (2015a) Outbred genome sequencing and CRISPR/Cas9 gene editing in butterflies. Nat Commun 6

Li Z et al (2015b) Ectopic expression of ecdysone oxidase impairs tissue degeneration in Bombyx mori. Proc R Soc B 282:20150513

Ling L et al (2015) MiR-2 family targets awd and fng to regulate wing morphogenesis in *Bombyx mori*. RNA Biol 12:742–748

Liu Y et al (2014) Highly efficient multiplex targeted mutagenesis and genomic structure variation in *Bombyx mori* cells using CRISPR/Cas9. Insect Biochem Mol Biol 49:35–42

Ma S et al (2014) CRISPR/Cas9 mediated multiplex genome editing and heritable mutagenesis of *BmKu70* in *Bombyx mori*. Sci Rep 4:4489

Markert MJ et al (2016) Genomic access to monarch migration using TALEN and CRISPR/Cas9-mediated targeted mutagenesis. G3: Genes| Genomes| Genetics 6:905–915

Merlin C, Beaver LE, Taylor OR, Wolfe SA, Reppert SM (2013) Efficient targeted mutagenesis in the monarch butterfly using zinc-finger nucleases. Genome Res 23:159–168

Perry M et al (2016) Molecular logic behind the three-way stochastic choices that expand butterfly colour vision. Nature 535:280–284. doi:10.1038/nature18616

Takasu Y et al (2010) Targeted mutagenesis in the silkworm *Bombyx mori* using zinc finger nuclease mRNA injection. Insect Biochem Mol Biol 40:759–765

Takasu Y et al (2013) Efficient TALEN construction for *Bombyx mori* gene targeting. PLoS One 8:e73458

Tamura T et al (2000) Germline transformation of the silkworm *Bombyx mori* L. using a piggyBac transposon-derived vector. Nat Biotechnol 18:81–84

Terenius O et al (2011) RNA interference in Lepidoptera: an overview of successful and unsuccessful studies and implications for experimental design. J Insect Physiol 57:231–245

Wang Y et al (2013) The CRISPR/Cas system mediates efficient genome engineering in *Bombyx mori*. Cell Res 23:1414–1416

Wei W et al (2014) Heritable genome editing with CRISPR/Cas9 in the silkworm, *Bombyx mori*. PLoS One 9:e101210

Xin Hh et al (2015) Transcription factor Bmsage plays a crucial role in silk gland generation in silkworm, Bombyx Mori. Arch Insect Biochem Physiol 90:59–69

Zeng B et al (2016) Expansion of CRISPR targeting sites in *Bombyx mori*. Insect Biochem Mol Biol 72:31–40

Zhang L, Reed RD (2016) Genome editing in butterflies reveals that *spalt* promotes and *Distal-less* represses eyespot colour patterns. Nat Commun 7. doi:10.1038/ncomms11769

Zhang Z et al (2015) Functional analysis of Bombyx Wnt1 during embryogenesis using the CRISPR/Cas9 system. J Insect Physiol 79:73–79

Zhang L et al (2017) Genetic basis of melanin pigmentation in butterfly wings. Genetics 102632. doi:10.1534/genetics.116.196451

Zhu L, Mon H, Xu J, Lee JM, Kusakabe T (2015) CRISPR/Cas9-mediated knockout of factors in non-homologous end joining pathway enhances gene targeting in silkworm cells. Sci Rep 5:18103

Chapter 9
What Can We Learn About Adaptation from the Wing Pattern Genetics of Heliconius Butterflies?

Chris D. Jiggins

Abstract *Heliconius* wing patterns are an adaptive trait under strong selection in the wild. They are also amenable to genetic studies and have been the focus of evolutionary genetic analysis for many years. Early genetic studies characterised a large number of Mendelian loci with large effects on wing pattern elements in crossing experiments. The recent application of molecular genetic markers has consolidated these studies and led to recognition that a huge range of allelic variation at just a few major loci controls patterns across most of the *Heliconius* radiation. Some of these loci consist of tightly linked components that control different aspects of the phenotype and can be separated by occasional recombination. More recent quantitative analyses have also identified minor-effect loci that influence the expression of these major loci.

Studies of a single locus polymorphism in *Heliconius numata* provide an example of a 'supergene', in which a single major locus controls segregation of a variable phenotype. This supports 'Turner's Sieve' hypothesis for the evolution of supergenes, whereby sequential linked mutations arise at the same locus. In addition, inversion polymorphisms are associated with wing pattern variation in wild populations, which reduce recombination across the supergene locus. This provides direct evidence that the architecture and organisation of genomes can be shaped by natural selection. There is also evidence that patterns of dominance of the alleles at this locus have also been shaped by natural selection. Mimicry therefore provides a case study of how natural selection shapes the genetic control of adaptive variation.

Keywords Mimicry • Heliconius • Convergent evolution • Input-output gene • Developmental pathway • Adaptive radiation

A major research effort in evolutionary biology is devoted to determining the molecular changes in DNA sequences that control adaptive phenotypic changes. By identifying the number and identity of genes controlling traits, and the relative

C.D. Jiggins (✉)
Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK
e-mail: c.jiggins@zoo.cam.ac.uk

T. Sekimura, H.F. Nijhout (eds.), *Diversity and Evolution of Butterfly Wing Patterns*, DOI 10.1007/978-981-10-4956-9_9

contribution of individual mutations to changes in the appearance of an organism, we can address a wealth of questions in evolutionary biology including some that were debated by early geneticists, such as the importance of large versus small mutations in evolution. Mimicry patterns in *Heliconius* butterflies have contributed significantly to our understanding of the genetic basis for adaptation over the past 40 years. Here I review what is known of the genetic basis for these bright colour patterns and some of the implications for our understanding of evolution.

9.1 Phenotypic Effects of Major Loci: The Red Locus Optix

The most striking aspect of *Heliconius* wing pattern genetics is that a few major loci control large phenotypic changes (Fig. 9.1). This major locus control of adaptive traits is an emerging pattern in other organisms, but studies of butterflies provided some of the first clear examples (Nadeau and Jiggins 2010) and were already evident in early work (Sheppard et al. 1985). The locus that is best understood at a molecular level and has perhaps the largest phenotypic effect controls red patterns (Table 9.1). Alternate alleles represent regulatory switches controlling expression of the transcription factor *optix*. The most studied red patterns controlled by this locus can be divided into three main elements: the red forewing band, the red ray pattern on the hindwing and the basal patch on the forewing. The latter is known as the 'Dennis' patch, after an individual butterfly that William Beebe named '*Dennis the Menace'*. Once linked genetic markers were identified, it became clear that there is a remarkable degree of homology between species in the control of these elements (Baxter et al. 2008).

This shows that convergent patterns in mimetic species are controlled by the same genetic mechanism. But what about other types of patterns? It turns out that a huge diversity of patterns are controlled by the same genetic loci. For example, this locus also controls orange patches in silvaniform butterflies, *H. hecale* and *H. ismenius* (Huber et al. 2015), and the brown forceps-shaped pattern on the ventral hindwing of *H. cydno* (Naisbit et al. 2003; Chamberlain et al. 2011). In fact, in every species so far investigated genetically, this locus has major phenotypic effects on red and orange pattern elements.

The *optix* locus actually consists of distinct, tightly linked elements. Direct estimation of recombination rates between these has proven difficult, but there are rare natural recombinants. For example in *H. erato*, a single individual with *ray* but not *dennis* was collected in a Peruvian hybrid zone, and similar individuals are known in *H. melpomene* (Mallet 1989). There are also established races that have recombinant genotypes, such as *H. e. amalfreda* and *H. m. meriana* that have *dennis* but not *ray*, while *H. timareta timareta f. contigua* is a form with *ray* but not *dennis*.

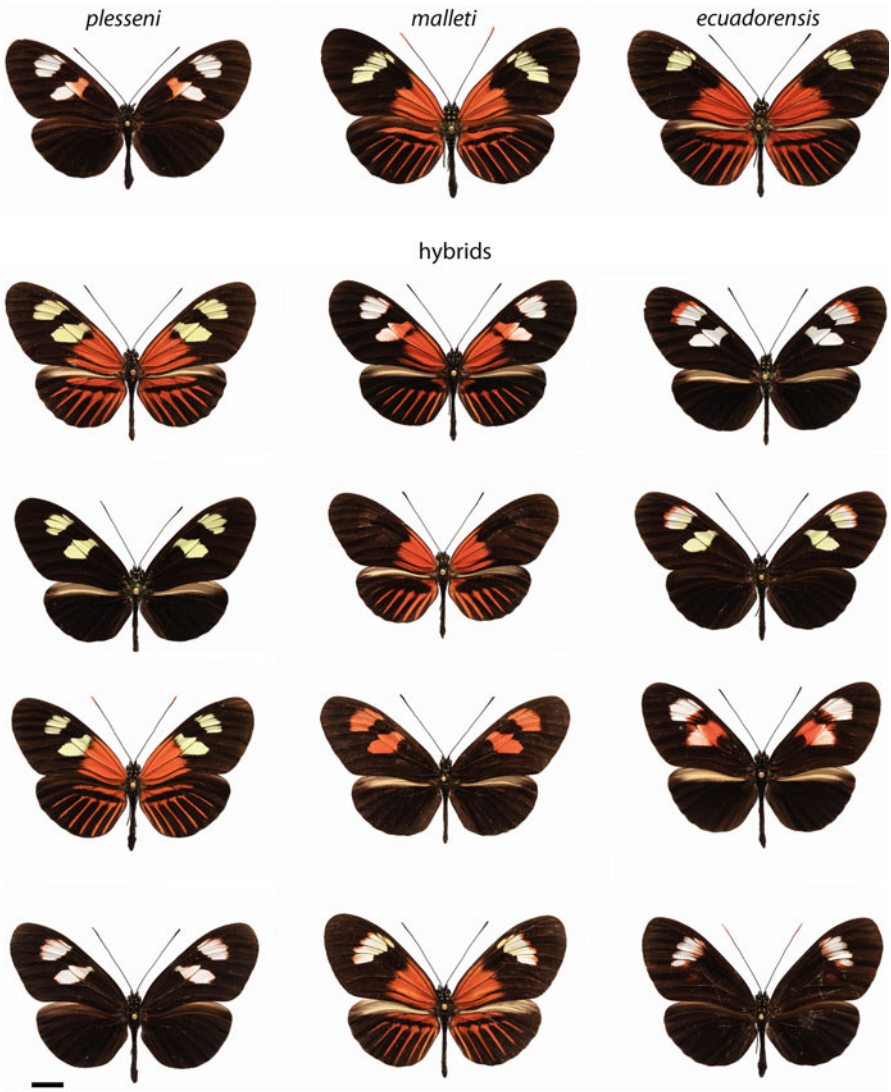

Fig. 9.1 Phenotypes from a hybrid zone in Eastern Ecuador
There are three parental races that contribute variation to the hybrid zone, pictured here along the top row *H. m. plesseni*, *H. m. malleti* and *H. m. ecuadorensis*. Three major loci control the wing patterns, *D* controls red/orange pattern elements, *Ac* controls the shape of the forewing band (two spots or one) and *Yb* produces the yellow forewing band. These butterfly hybrids are all from the Neukirchen collection. Scale bar is 1 cm

Table 9.1 Summary of published wing patterning loci

Species	Locus	Phenotypic effect	Reference (corresponding to the caption)
D – Optix – LG18			
H. melpomene	D	Dennis patch	1
	B	Red FW band	1
	R	HW rays	1
	M	Yellow FW band	2
H. erato	Y	Yellow/red FW band	1
	D	Dennis patch	1
	R	HW rays	1
	Wh	White in FW	1
H. cydno	Br	Brown cydno 'C'	3
H. pachinus/heurippa	G	Red HW spots	3, 4
H. hecale	HhBr	HW orange/black	6
H. ismenius	HiBr	HW orange/black	6
Yb – cortex – LG15			
H. melpomene/cydno	Yb	Yellow HW bar	1,3
	N	Yellow FW band	1,3
	Sb	HW white margin	3,5
	Vf	Pale ventral FW band	3
H. erato	Cr	Cream rectangles	1
H. hecale	HhN	FW submarginal spots	6
H. ismenius	HiN	FW submarginal spots	6
H. ismenius	FSpot	FW subapical spots	6
H. ismenius	HSpot	HW marginal spots	6
H. numata	P	All pattern variants	7
Ac – WntA – LG10			
H. melpomene/cydno	Ac	FW band shape	1,3
	C	Broken FW band	1
	S	Shortens FW band	1,8
H. erato	Sd	FW band shape	1,9
	Sd	HW bar	1,9,10
	St	Split FW band	1,9
	Ly	Broken FW band	1,9
	Yl	Yellow FW line	1,11
H. hecale	HhAc	Yellow FW band	6
H. ismenius	HiAc	Yellow FW band	6

(continued)

Table 9.1 (continued)

Species	Locus	Phenotypic effect	Reference (corresponding to the caption)
LG1			
H. melpomene/cydno	K	FW band colour (yellow/white)	3,12
	Khw	HW margin colour (yellow/white)	13
LG13			
H. melpomene	Unnamed	FW band shape	14
H. erato	Ro	Rounded FW band	15
Unknown			
H. melpomene	Or	Orange/red switch	1
H. cydno	L/Wo	Forewing white spots	16
H. cydno/pachinus	Ps	Pachinus 'shutter'	17
H. cydno	Fs	Forewing 'shutter'	17
H. cydno	Cs	Cydno 'shutter'	17

A summary of previously described wing patterning loci and their homology to major effect genes. HW and FW refer to hindwing and forewing respectively. Notes: [1] Sheppard et al. (1985). [2] The *M* locus interacts with *N* to influence the forewing yellow band in *H. melpomene* (Mallet 1989). Unpublished work (Baxter and Mallet pers. Comm.) indicates that *M* is an effect of the *optix* locus. [3] Naisbit et al. (2003). [4] Mavarez et al. (2006). [5] Linares (1996). [6] Huber et al. (2015). [7] The *P* supergene locus in *H. numata* controls all aspects of phenotype. The locus is homologous to *Yb* although it seems likely that the supergene includes several functional loci (Joron et al. 2006).[8] Nijhout (1990). [9] Papa et al. (2013). [10] Mallet (1989). [11] Sheppard et al. (1985) infer that *Yl* and *Sd* are linked, but that *Yl* and *Ly* segregate independently. *Sd* and *Ly* are now known to be the same locus, so it is unclear whether *Yl* is unlinked. Further crosses of Brazilian forms would be needed to test this. [12] Kronforst et al. (2006). [13] Joron et al. (2006). [14] Baxter et al. (2009). [15] The *Ro* locus was mapped to linkage group 13 by means of a hybrid zone association study (Nadeau et al. 2014). [16] L and Wo are linked loci that control forewing white elements in *H. cydno* and may be homologous to *Ac* (Linares 1996). [17] *Ps*, *Fs* and *Cs* from Nijhout (1990) are included for completeness but patterns of segregation and linkage are not known. These may be effects of the *WntA* locus

Recent molecular analysis has confirmed that these phenotypes are indeed recombinants between tightly linked elements located in non-coding DNA near to *optix* (Wallbank et al. 2016). Thus, there are at least three very tightly linked elements that independently control different patches of red on the wing.

9.2 Phenotypic Effects of Major Loci: The Yellow Locus Cortex

This second major locus is similar in many ways to the red locus – it consists of tightly linked elements that similarly control different patches of yellow and white pattern. The *cortex* locus represents a cluster of tightly linked loci located on linkage group 15. These include effects known as *Yb*, *Sb* and *N* in *H. melpomene* and *Cr* in *H. erato* (Sheppard et al. 1985; Mallet 1986). Alleles that produce a yellow band are recessive to the absence of the band, although heterozygotes typically show an alteration in scale morphology in the band region that can be seen in altered reflectance in the otherwise black hindwing. Another allele at the same locus produces a band only on the underside of the hindwing and is present in the west Colombian race *H. m. venustus*. The same genomic region also controls a white hindwing margin found in the west Ecuador races *H. e. cyrbia* and *H. m. cythera* (Jiggins and McMillan 1997; Ferguson et al. 2010).

Many of the coloured patches on *Heliconius* wings are controlled in this very simple one-allele makes one-phenotype manner. However, there are also more complex interaction effects between loci. For example, in East Andean populations of *H. erato*, the yellow hindwing bar results from the joint effects of two loci, *cortex* and *WntA*. Thus, in Peruvian *H. e. favorinus*, recessive alleles at both loci are required for full expression of the hindwing bar (Mallet 1989) (although in Central American *H. erato*, a very similar bar results from a recessive allele at one locus). There is also evidence for rare recombination events between tightly linked loci at this locus. Thus, for example, *Yb* and *Sb* were mapped to within ~1 cM of one another, with two recombinant phenotypes identified in 175 individuals (Ferguson et al. 2010). Similar results are seen in crosses between *H. melpomene rosina* and *H. c. chioneus* (Naisbit et al. 2003).

In summary, these two loci both consist of a set of tightly linked genetic elements that control major phenotypic changes. Each locus controls pattern elements with broadly similar phenotypic effects: yellow and white patches in the case of *cortex* and red and orange patches in the case of *optix*. Patterns of dominance are also predictable, with alleles for red elements dominant, and those for yellow or white elements recessive, giving a dominance series of red > black > white > yellow. In both cases, loci most likely represent tightly linked *cis*-regulatory elements of the same protein-coding gene, with linkage a result of genetic architecture rather than being favoured by selection.

9.3 Phenotypic Effects of Major Loci: The Shape Locus *WntA*

The third major locus is located on linkage group 10 and primarily controls the shape of the forewing elements. For example, in crosses between *H. melpomene rosina* and *H. cydno chioneus,* a recessive allele *ac* places a triangle that forms a white hourglass shape in the main forewing cell of *H. cydno* (Naisbit et al. 2003). In the Ecuadorean *H. m. plesseni*, this locus produces the 'split' forewing band – the largely recessive *H. m. plesseni* allele expresses the more proximal of the two white patches of this form and also influences the shape of the more distal patch (Salazar 2012). This locus likely results from variation in expression of the gene *WntA* (Martin et al. 2012).

A wide variety of loci have previously been described (*St*, *Sd* and *Ly*) which all map to the same genomic location (Papa et al. 2013), corresponding to *WntA*. These loci influence the shape of forewing band elements. In some cases the phenotypic effects of this locus are extremely similar to those seen in *H. melpomene*; thus, for example, in *H. e. notabilis*, which is mimetic with *H. m. plesseni, Sd* also acts to generate the split forewing band phenotype (Salazar 2012). In Amazonian forms, the allele at this locus also generates the broken yellow forewing band (Sheppard et al. 1985; Papa et al. 2013).

9.4 Phenotypic Effects of Other Loci

A further locus, termed *K*, controls the colour change between yellow and white pigments in *H. melpomene*, *H. cydno* and *H. pachinus*. Most strikingly, this locus controls a polymorphism of yellow and white forms in *H. cydno alithea* in western Ecuador. The *K* locus is located on linkage group 1 and is linked with the gene *wingless* (Kronforst et al. 2006). This differs from other loci in that it influences solely colour, with no effect on pattern. There are also a number of minor-effect loci described in the older literature, but in most cases, these have been found to represent allelic effects of the major loci described above. Nonetheless, some of these loci are likely to be distinct. For example, a locus named *Or* described in both *H. melpomene* and *H. erato* controls the switch between red and orange colours (Sheppard et al. 1985). 'Postman' races typically have a bright red forewing band, while Amazonian forms have orange *dennis* and *ray* patterns. Another locus that has been better characterised is *Ro*, which generates a rounded forewing band phenotype such as that seen in *H. e. notabilis* (Salazar 2012; Papa et al. 2013; Nadeau et al. 2014). Some of the most beautiful but poorly characterised are the iridescent blue and green colours that result from structural variation in the wing scales. These traits vary continuously and are difficult to quantify (Jiggins and McMillan 1997). However, while most analysis of *Heliconius* genetics has relied on

the scoring of presence/absence of major pattern elements, a better characterisation of these minor-effect loci is gained by a quantitative analysis of pattern segregation.

9.5 Quantitative Analysis

A comprehensive QTL analysis was carried out by Papa et al. using crosses between *H. e. notabilis* and *H. himera* (Papa et al. 2013). This confirmed the subjective finding from generations of earlier researchers that major loci control the segregation of most of the wing variation in crosses. For example, an additive model showed that the *optix* locus controlled 87% of variance in the amount of white versus yellow in the forewing, while the amount of red was best described by an epistatic model in which *optix* explained ~56% of the variation. The sizes of the two forewing spots showed a less skewed distribution of effect sizes and were controlled by several QTL of moderate effect (>5%), some as large in effect as the major locus *WntA*. For example, four QTL together explained 63% of the variance in the 'big spot', one of which was the *WntA* locus. This spot shape analysis therefore suggests a less skewed, more quantitative genetic architecture. Nonetheless, the overall variance explained across the complete set of *H. erato* crosses described in this paper is strongly dominated by large-effect loci.

These QTL analyses of specific wing pattern traits still fail to capture and quantify both segregation of the presence and absence of major pattern elements in the same analysis as quantitative variation in the expression of those traits. More recently, analytical methods have been developed that capture all of the variation in colour and pattern into a single PCA analysis (Huber et al. 2015; Le Poul et al. 2014), which was used to analyse broods of *H. hecale* and *H. ismenius*. All of the significant QTL identified corresponded to the existing major wing patterning loci. More minor QTL did not pass the significance threshold, although some of these additional loci would likely become significant with larger sample sizes. These quantitative analyses therefore support the conclusion that most variations are controlled by a handful of major-effect loci, although their expression is modified by minor-effect loci. In the future, there is a clear need for studies that combine large mapping families with objective methods for pattern analysis to better characterise the distribution of wing patterning variants.

9.6 Non-genetic Effects and Plasticity

There has been considerable interest recently in the role of phenotypic plasticity in evolution, and it has been proposed that plasticity can promote evolutionary novelty, for example, by allowing populations to explore new phenotypes without genetic change (Pfennig et al. 2010; Moczek et al. 2011). However, there is little evidence for phenotypic plasticity in the expression of *Heliconius* wing patterns.

First, most of the variation in wing pattern among hybrid butterflies can be explained by genetic variation at just a handful of major loci. Second, in the wild there is very little phenotypic variation in wing pattern among individuals occurring across a wide range of altitudes and habitats – apart from genetically divergent wing pattern races. Some pigment colours do fade with age, or in stressed individuals, but this is not adaptive plasticity. In summary, while plasticity may play a role in many aspects of *Heliconius* biology, such as learning of behaviour, there is no evidence that it plays a role in wing pattern evolution.

9.7 A Distribution of Effect Sizes?

Early workers used major genes in butterfly mimicry as an argument for major mutations driving evolution, but Fisher countered that mutations with a large effect on the organism will virtually always be deleterious (Fisher 1930). More recently Orr has shown that during an adaptive walk, we expect an exponential distribution of mutational effect sizes (Orr 1998, 2005). Early in the process, there is a high likelihood of mutations that move the population a large distance relative to the optimum. Later on, smaller effect mutations are more probable, that act to 'fine-tune' the adaptation. To some extent this modern view therefore reconciles the two camps.

The theory developed by Orr and others hypothesised a population evolving towards a single adaptive peak. However, the frequency-dependent nature of mimicry and warning colour means that these traits have a different dynamic. If a population of butterflies has a bright warning colour pattern (hereafter the 'mimic'), predators will learn this pattern, and the population will generally be well protected from predation. There may be other butterfly species locally that are perhaps more abundant or more toxic (the 'model') and therefore have a better-protected wing patterns, so the mimic species would gain in fitness by evolving mimicry of the model pattern. However, an individual 'mimic' that deviates from the rest of the population would be selected against, even if it becomes slightly more similar to the model. The two patterns would have to be very similar for predators to generalise between them, in order for gradual evolution towards the model to be possible (Turner 1981). Most current *Heliconius* patterns in different mimicry rings are sufficiently different from one another that gradual convergence seems unlikely. There is a valley of low fitness between the model and mimic which would seem to prevent gradual evolution of mimicry. This difficulty can be overcome if a single mutation causes a large change, sufficient to induce enough similarity to the model in one step that overall fitness is increased. This initial mutation is unlikely to produce a perfect mimic, so subsequent mutations will then be needed to perfect the phenotype. This argument was first outlined by Nicholson (1927) and termed the 'Nicholson two-step model' by John Turner (1977, 1984, 1987). Mimicry may therefore have a different genetic architecture to traits evolving under a single-peak-climbing model (Baxter et al. 2009).

The major locus control of *Heliconius* patterns seems to fit with the predictions of the 'Nicholson two-step model' (Huber et al. 2015; Papa et al. 2013; Turner 1981; Baxter et al. 2009), with a few major loci and additional modifiers of small effect. However, there are a number of reasons to be sceptical of this simple interpretation. First, many races within both *H. erato* and *H. melpomene* differ at several unlinked major-effect loci. For example, hybrid zones in both Peru and Ecuador between races of both *H. melpomene* and *H. erato* differ in at least two major loci (Mallet 1989; Salazar 2012; Nadeau et al. 2014). It is not clear whether a substitution at just one of these loci would be sufficient to gain enough mimetic similarity to provide protection, while the population 'waited' for a subsequent mutation at the second locus. Turner has acknowledged this difficulty but suggested either multiple rounds of 'two-step' evolution or that changes at just one of the loci would be sufficient to confer a fitness advantage (Turner 1977).

Another mismatch between the theory and empirical data is that the data from crossing experiments refers to the phenotypic effects of genetic loci, not separate mutations (Baxter et al. 2009). As pointed out by Fisher (1930), and more recently in dissection of major effect QTL in other organisms (Stam and Laurie 1996; Linnen et al. 2013), major-effect loci can result from accumulation of many mutations at a single locus. It seems likely that single large-effect genetic loci harbour many mutations corresponding to adaptive steps towards the peak. Testing the 'two-step model' therefore becomes a much more challenging problem of separating the order and effect size of individual mutations at a single locus. Nonetheless, mimicry can arise through hybridisation, in which an already well-adapted large-effect allele is acquired from a related species. This represents a clear case of single-step 'major-effect' evolution, so there certainly are at least *some* cases in which large changes are involved (The Heliconius Genome Consortium, 2012). Overall therefore, the 'rugged' adaptive landscape of mimicry likely favours adaptation via large steps as described under a two-step theory, and this might provide some part of the explanation for the major-effect loci involved in *Heliconius* mimicry.

9.8 Supergenes and Polymorphism

The broad picture of wing pattern genetics outlined above applies to most *Heliconius* that have been studied, but there is one species in the genus that has a very different pattern: *H. numata*. Mimicry patterns in *Heliconius numata* are polymorphic, with different morphs mimetic with different species mostly in the genus *Melinaea*. These dramatic differences are controlled by a single genetic locus, with several alternate alleles. Such loci are known as '*supergenes',* which we have defined as '*A genetic architecture involving multiple linked functional genetic elements that allows switching between discrete, complex phenotypes maintained in a stable local polymorphism*' (Thompson and Jiggins 2014). There are two major characteristics of the *Heliconius numata* supergene that maintain an

integrated phenotype. First, a lack of recombination – all aspects of the phenotype are inherited as a single non-recombining locus – and second, dominance: alternate alleles show complete dominance relationships such that heterozygote genotypes develop the wing pattern of one or other parent.

The *P* supergene is genetically homologous to the region of the *cortex* locus in *H. melpomene* (Joron et al. 2006). The genetic architecture of 3–4 major loci is ancestral because it is shared by all other species in the genus that have been studied (Huber et al. 2015), so in *H. numata* this locus has 'taken over' control of all aspects of pattern variation (Jones et al. 2012). There are several hypotheses to explain the gradual evolution of tightly linked elements in a supergene. A long-standing hypothesis is that alleles located in different regions of the genome might be translocated into tight linkage (Turner 1967). However, there is no evidence for long-range movement of genes; the gene content of the region is similar in all *Heliconius*. The *P* locus has therefore evolved control of pattern variation normally influenced by genes on different chromosomes, rather than by moving those genes into linkage. The second hypothesis is that sequential mutations might arise in tight linkage with the polymorphic locus and be favoured by selection (Turner 1977; Charlesworth and Charlesworth 1976). Mutations that improve one mimetic form are likely to make things worse for other forms. However, if a new mutation is tightly linked at the *P* locus, then it will always be inherited with the alleles with which it is coadapted. This process has become known as 'Turner's sieve', because it involves sieving of the genetic variation that arises in order to select only linked variants (Turner 1977; Charlesworth and Charlesworth 1976; Turner 1978). The fact that *P* consists of linked elements suggests that these must have arisen through multiple sequential mutations.

Once linked elements have arisen, theory predicts that selection can act to further reduce recombination between them (Turner 1967; Charlesworth and Charlesworth 1976; Kirkpatrick and Barton 2006). Mathieu Joron and his group have identified large genomic inversions (400 kb) that segregate in polymorphic populations around the *P* locus (Fig. 9.2). Alternate gene arrangements are fully associated with wing pattern phenotypes in natural populations and show strong linkage disequilibrium in natural populations. Effectively, there is a block of about 400 kb of DNA sequence that is inherited in complete association with different wing pattern forms (Joron et al. 2011). Similar inversions have been seen in complex polymorphisms in other species – notably a behavioural and plumage polymorphism in the white-throated sparrow, a social polymorphism in fire ants and a behavioural polymorphism in the ruff, a wading bird (Thompson and Jiggins 2014; Huynh et al. 2011; Wang et al. 2013; Küpper et al. 2015; Lamichhaney et al. 2015). In all cases, inversions lock together inheritance of a large part of one chromosome. Perhaps more similar to the *Heliconius numata* case is *Papilio polytes,* in which a very localised inversion around the *Dsx* gene controls a wing pattern mimicry polymorphism (Kunte et al. 2014; Nishikawa et al. 2015). These examples all suggest that the evolution of inversions to reduce recombination between coadapted alleles may be a common phenomenon.

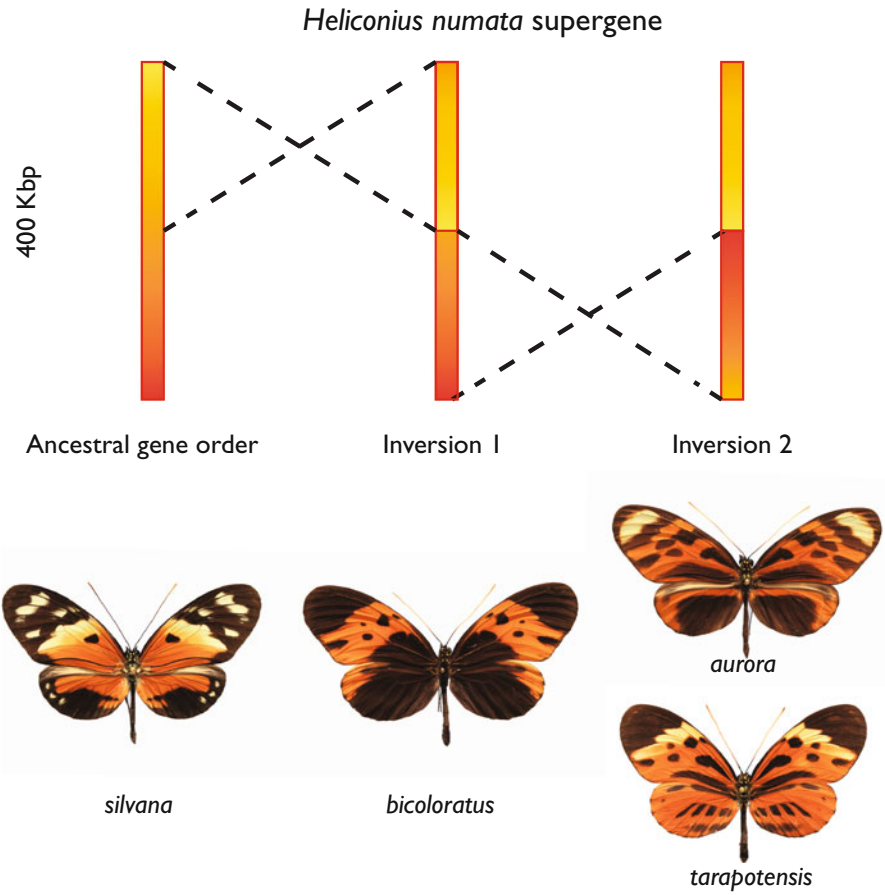

Fig. 9.2 Structural variation associated with the *Heliconius numata* supergene
At least two genetic inversions are associated with the *H. numata* supergene. The ancestral gene order, which matches that in *H. melpomene* and *H. erato* is shown on the left and is associated with ancestral phenotypes such as *H. n. silvana*. Two sequentially derived inversions are associated with dominant alleles and are shown in the middle and right. Redrawn from (Joron et al. 2011)

The second aspect of a supergene that ensures mimicry is a strong pattern of dominance (Llaurens et al. 2015). Alternate alleles show complete dominance, with an allelic series between morphs (Le Poul et al. 2014; Joron et al. 2011; Brown 1976). Remarkably, one heterozygote genotype is distinct but appears to have been stabilised because of its effective mimicry of a different species (Le Poul et al. 2014). In most *Heliconius*, there are predictable rules for dominance. Red/orange pattern elements are generally dominant over black, while yellow/white pattern elements are recessive. The complete dominance of alleles across the entire wing surface in *H. numata* therefore represents a derived state that apparently overturns typical 'rules of inheritance'. Dominance has been optimised by natural selection.

These patterns of dominance could be controlled by mutations within the supergene itself or unlinked loci acting to control dominance at *P*. Although there is evidence for both of these processes, recent analysis provides strong evidence for evolution of dominance at the *P* locus itself. Patterns of dominance *between* derived and ancestral alleles show unusual patterns of dominance in which the typical dominance patterns are overruled. In contrast, *among* derived alleles, patterns of dominance follow the typical colour hierarchy seen in other *Heliconius* species (Le Poul et al. 2014). These patterns suggest that dominance is a property of the alleles themselves, rather than the genetic background. This will be a fascinating system in which to explore mechanisms underlying the evolution of dominance.

9.9 Conclusions

The extraordinary diversity of wing patterns among the *Heliconius* butterflies has provided insights into the diversification of animal form and its genetic control. An important discovery has been the repeated role of just a handful of loci in diversification of not just convergent mimetic patterns but also diverse and novel phenotypes. Nonetheless, there is still a need for better quantitative analysis of patterns that will reveal the distribution of loci controlling adaptation. These patterns parallel discoveries in other systems, for example, sticklebacks, where similarly there are a few loci with major effects on phenotype (Colosimo et al. 2005; Chan et al. 2010), but many traits are also influenced by more polygenic control (Peichel and Marques 2017).

I have reviewed our understanding of wing patterning based on genetic crossing experiments, but have not considered in detail the developmental basis for pattern diversity, which has recently been reviewed elsewhere (Jiggins et al. 2017).

References

Baxter SW, Johnston SE, Jiggins CD (2009) Butterfly speciation and the distribution of gene effect sizes fixed during adaptation. Heredity 102:57–65. doi:10.1038/hdy.2008.109

Baxter SW, Papa R, Chamberlain N, Humphray SJ, Joron M, Morrison C, ffrench-Constant RH, McMillan WO, Jiggins CD (2008) Convergent evolution in the genetic basis of Müllerian mimicry in *Heliconius* butterflies. Genetics 180:1567–1577. doi:10.1534/genetics.107.082982

Brown KS (1976) An illustrated key to the silvaniform Heliconius (Lepidoptera: Nymphalidae) with descriptions of new subspecies. Trans Am Entomol Soc 102:373–484

Chamberlain NL, Hill RI, Baxter SW, Jiggins CD, Kronforst MR (2011) Comparative population genetics of a mimicry locus among hybridizing Heliconius butterfly species. Heredity 107:200–204. doi:10.1038/hdy.2011.3

Chan YF et al (2010) Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. Science 327:302–305. doi:10.1126/science.1182213

Charlesworth D, Charlesworth B (1976) Theoretical genetics of batesian mimicry II. Evolution of supergenes. J Theor Biol 55:305–324. doi:10.1016/S0022-5193(75)80082-8

Colosimo PF et al (2005) Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. Science 307:1928–1933. doi:10.1126/science.1107239

Ferguson L et al (2010) Characterization of a hotspot for mimicry: assembly of a butterfly wing transcriptome to genomic sequence at the HmYb/Sb locus. Mol Ecol 19:240–254. doi:10.1111/j.1365-294X.2009.04475.x

Fisher RA (1930) The Genetical theory of natural selection, 1st edn. Oxford University Press, Oxford

Huber B et al (2015) Conservatism and novelty in the genetic architecture of adaptation in Heliconius butterflies. Heredity 114:515–524. doi:10.1038/hdy.2015.22

Huynh LY, Maney DL, Thomas JW (2011) Chromosome-wide linkage disequilibrium caused by an inversion polymorphism in the white-throated sparrow (*Zonotrichia albicollis*). Heredity 106:537–546. doi:10.1038/hdy.2010.85

Jiggins CD, McMillan WO (1997) The genetic basis of an adaptive radiation: warning colour in two Heliconius species. Proc R Soc Biol Sci 264:1167–1175. doi:10.1098/rspb.1997.0161

Jiggins CD, Wallbank RWR, Hanly JJ (2017) Waiting in the wings: what can we learn about gene co-option from the diversification of butterfly wing patterns? Philos Trans R Soc B 372:20150485. doi:10.1098/rstb.2015.0485

Jones RT, Salazar PA, ffrench-Constant RH, Jiggins CD, Joron M (2012) Evolution of a mimicry supergene from a multilocus architecture. Proc R Soc Lond B Biol Sci 279:316–325. doi:10.1098/rspb.2011.0882

Joron M et al (2006) A conserved supergene locus controls colour pattern diversity in Heliconius butterflies. PLoS Biol 4:1831–1840. doi:10.1371/journal.pbio.0040303

Joron M et al (2011) Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. Nature 477:203–206. doi:10.1038/nature10341

Kirkpatrick M, Barton N (2006) Chromosome inversions, local adaptation and speciation. Genetics 173:419–434. doi:10.1534/genetics.105.047985

Kronforst MR, Young LG, Kapan DD, McNeely C, O'Neill RJ, Gilbert LE (2006) Linkage of butterfly mate preference and wing color preference cue at the genomic location of wingless. Proc Natl Acad Sci U S A 103:6575–6580. doi:10.1073/Pnas.0509685103

Kunte K, Zhang W, Tenger-Trolander A, Palmer DH, Martin A, Reed RD, Mullen SP, Kronforst MR (2014) doublesex is a mimicry supergene. Nature 507:229–232. doi:10.1038/nature13112

Küpper C et al (2015) A supergene determines highly divergent male reproductive morphs in the ruff. Nat Genet 48:79–83. doi:10.1038/ng.3443

Lamichhaney S et al (2015) Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). Nat Genet 48:84–88. doi:10.1038/ng.3430

Le Poul Y, Whibley A, Chouteau M, Prunier F, Llaurens V, Joron M (2014) Evolution of dominance mechanisms at a butterfly mimicry supergene. Nat Commun 5:5644. doi:10.1038/ncomms6644

Linares M (1996) The genetics of the mimetic coloration in the butterfly *Heliconius cydno weymeri*. J Heredity 87(2):142–149

Linnen CR, Poh Y-P, Peterson BK, Barrett RDH, Larson JG, Jensen JD, Hoekstra HE (2013) Adaptive evolution of multiple traits through multiple mutations at a single gene. Science 339:1312–1316. doi:10.1126/science.1233213

Llaurens V, Joron M, Billiard S (2015) Molecular mechanisms of dominance evolution in Müllerian mimicry. Evolution 69:3097–3108. doi:10.1111/evo.12810

Mallet J (1986) Hybrid zones of Heliconius butterflies in Panama and the stability and movement of warning color clines. Heredity 56:191–202

Mallet J (1989) The genetics of warning color in Peruvian hybrid zones of *Heliconius erato* and Heliconius melpomene. Proc R Soc Lond Ser B-Biol Sci 236:163–185

Martin A et al (2012) Diversification of complex butterfly wing patterns by repeated regulatory evolution of a Wnt ligand. Proc Natl Acad Sci 109:12632–12637. doi:10.1073/pnas.1204800109

Mavarez J, Salazar CA, Bermingham E, Salcedo C, Jiggins CD, Linares M (2006) Speciation by hybridization in Heliconius butterflies. Nature 441(7095):868–871

Moczek AP, Sultan S, Foster S, Ledón-Rettig C, Dworkin I, Nijhout HF, Abouheif E, Pfennig DW (2011) The role of developmental plasticity in evolutionary innovation. Proc R Soc Lond B Biol Sci 278:2705–2713. doi:10.1098/rspb.2011.0971

Nadeau NJ, Jiggins CD (2010) A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. Trends Genet 26:484–492. doi:10.1016/j.tig.2010.08.004

Nadeau NJ et al (2014) Population genomics of parallel hybrid zones in the mimetic butterflies, H. melpomene and *H. erato*. Genome Res 24:1316–1333. doi:10.1101/gr.169292.113

Naisbit RE, Jiggins CD, Mallet J (2003) Mimicry: developmental genes that contribute to speciation. Evol Dev 5:269–280

Nijhout HF (1990) A comprehensive model for colour pattern formation in butterflies. Proc Royal Soc B Biol Sci 239(1294):81–113

Nicholson AJ (1927) A new theory of mimicry in insects. Aust Zool 5:10–104

Nishikawa H et al (2015) A genetic mechanism for female-limited Batesian mimicry in Papilio butterfly. Nat Genet 47:405–409. doi:10.1038/ng.3241

Orr HA (1998) The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. Evolution 52:935–949. doi:10.2307/2411226

Orr HA (2005) The genetic theory of adaptation: a brief history. Nat Rev Genet 6:119–127. doi:10.1038/nrg1523

Papa R, Kapan DD, Counterman BA, Maldonado K, Lindstrom DP, Reed RD, Nijhout HF, Hrbek T, McMillan WO (2013) Multi-allelic major effect genes interact with minor effect QTLs to control adaptive color pattern variation in *Heliconius erato*. PLoS One 8:e57033. doi:10.1371/journal.pone.0057033

Peichel CL, Marques DA (2017) The genetic and molecular architecture of phenotypic diversity in sticklebacks. Philos Trans R Soc B 372:20150486. doi:10.1098/rstb.2015.0486

Pfennig DW, Wund MA, Snell-Rood EC, Cruickshank T, Schlichting CD, Moczek AP (2010) Phenotypic plasticity's impacts on diversification and speciation. Trends Ecol Evol 25:459–467. doi:10.1016/j.tree.2010.05.006

Salazar PCA (2012) Hybridization and the genetics of wing colour-pattern diversity in Heliconius butterflies. Cambridge University, Cambridge

Sheppard PM, Turner JRG, Brown KS, Benson WW, Singer MC (1985) Genetics and the evolution of Muellerian mimicry in Heliconius butterflies. Philos Trans R Soc Lond Ser B Biol Sci 308:433–610. doi:10.2307/2398716

Stam LF, Laurie CC (1996) Molecular dissection of a major gene effect on a quantitative trait: the level of alcohol dehydrogenase expression in *Drosophila melanogaster*. Genetics 144:1559–1564

The Heliconius Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487:94–98. doi:10.1038/nature11041

Thompson MJ, Jiggins CD (2014) Supergenes and their role in evolution. Heredity 113:1–8. doi:10.1038/hdy.2014.20

Turner JRG (1967) On supergenes. I. The evolution of supergenes. Am Nat 101:195–221

Turner JRG (1977) Butterfly mimicry – genetical evolution of an adaptation. In: Hecht MK, Steere WC, Wallace B (eds) Evolutionary biology. Plenum Press, New York, pp 163–206

Turner JRG (1978) Why male butterflies are non-mimetic: natural selection, sexual selection, group selection, modification and sieving. Biol J Linn Soc 10:385–432. doi:10.1111/j.1095-8312.1978.tb00023.x

Turner JRG (1981) Adaptation and evolution in Heliconius – a defense of Neodarwinism. Annu Rev Ecol Syst 12:99–121

Turner JRG (1984) Mimicry: the palatability spectrum and its consequences. In: Vane-Wright RI, Ackery PR (eds) The biology of butterflies. Academic Press, London, pp 141–161

Turner JRG (1987) The evolutionary dynamics of batesian and muellerian mimicry: similarities and differences. Ecol Entomol 12:81–95. doi:10.1111/j.1365-2311.1987.tb00987.x

Wallbank RWR et al (2016) Evolutionary novelty in a butterfly wing pattern through enhancer shuffling. PLoS Biol 14:e1002353. doi:10.1371/journal.pbio.1002353

Wang J, Wurm Y, Nipitwattanaphon M, Riba-Grognuz O, Huang Y-C, Shoemaker D, Keller L (2013) A Y-like social chromosome causes alternative colony organization in fire ants. Nature 493:664–668. doi:10.1038/nature11832

Chapter 10
Molecular Mechanism and Evolutionary Process Underlying Female-Limited Batesian Mimicry in *Papilio polytes*

Haruhiko Fujiwara

Abstract Mimicry is an important evolutionary trait involved in prey-predator interactions. In a swallowtail butterfly *Papilio polytes*, only mimetic-form females mimic the unpalatable butterfly, *Pachliopta aristolochiae*, but it remains unclear how this female-limited polymorphic Batesian mimicry is generated and maintained. To explore the molecular mechanisms, we determined two whole genome sequences of *P. polytes* and its related species *P. xuthus* for comparison. The genome projects revealed a single long-autosomal inversion outside *doublesex* (*dsx*) between mimetic (*H*) and non-mimetic (*h*) chromosomes (Chr25) *in P. polytes*. The inversion site was just same as the mimicry locus *H* identified by linkage mapping. The gene synteny around *dsx* among Lepidoptera suggests that *H*-chromosome originates from *h*-chromosome. The 130 kb inverted region includes three genes, *doublesex* (*dsx*), *UXT*, and *U3X*, all of which were expressed from *H*-chromosome, but rarely from *h*-chromosome, indicating that these genes in *H*-chromosome are involved in the mimetic trait as supergene. Amino acid sequences of Dsx were substituted at over 13 sites between *H*- and *h*-chromosomes. To certify the functional difference of Dsx, we performed electroporation-mediated knock-down and found that only female *dsx* from *H*-chromosome (*dsx_H)* induced mimetic patterns but simultaneously repressed non-mimetic patterns on female wings. We propose that *dsx_H* switches the coloration of predetermined patterns in female wings and that female-limited polymorphism is tightly kept by chromosomal inversion. In this chapter, I will introduce the above results and discuss about the molecular mechanism and evolutionary process underlying the female-limited Batesian mimicry in *P. polytes*.

Keywords Batesian mimicry • Female-limited polymorphic mimicry • *Papilio polytes* • Whole genome sequence • *Doublesex* • Supergene • Chromosomal inversion • Electroporation-mediated functional analysis • Wing coloration pattern

H. Fujiwara (✉)
Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba 277-8562, Japan
e-mail: haruh@edu.k.u-tokyo.ac.jp

T. Sekimura, H.F. Nijhout (eds.), *Diversity and Evolution of Butterfly Wing Patterns*, DOI 10.1007/978-981-10-4956-9_10

10.1 Research Background

One of the most essential problems in evolutionary biology is to elucidate the molecular basis of various and adaptive morphological phenotypes in living organisms. The morphological diversity plays an important role in adaptation to the surrounding environment in many cases (Darwin 1872). Insects at the bottom of the food chain have been continuously attacked by the predators and thus developed various defense strategies to avoid predation during evolution (Ruxton et al. 2005).

Among the various strategies used by butterflies to avoid predators, some butterflies have become unpalatable and inform predators to their toxicity by exhibiting the conspicuous wing patterns. Some unpalatable butterflies share similar wing patterns to provide mutualistic protection called Mullerian mimicry (Müller 1878). In contrast, some palatable butterflies have evolved Batesian mimicry, in which they resemble unpalatable model to protect them from predators (Bates 1862; Brower 1958; Uesugi 1996). Multiple loci are involved in the expression of Mullerian mimicry phenotypes in *Heliconius* butterflies (Jiggins et al. 2005; Kapan et al. 2006), whereas the phenotypes of Batesian mimicry species reported so far are determined by a single locus (Clarke and Sheppard 1959, 1962, 1972).

It is noteworthy that the Common Mormon butterfly, *Papilio polytes*, shows a female-limited Batesian mimicry (Clarke and Sheppard 1972). The females have two forms: non-mimetic female (also called *cyrus*) which wing patterns are almost identical to monomorphic males and mimetic female (also called *polytes*) which resembles wing patterns of the distasteful butterfly, the Common Rose, *Pachliopta aristolochiae* (Fig. 10.1). This female-limited dimorphism is controlled by a single autosomal locus *H*, and the mimetic phenotype (genotype: *HH* or *Hh*) is dominant (Clarke and Sheppard 1972). However, how the female-limited Batesian mimicry is generated or how the female dimorphism is maintained is largely unknown.

There are two models for the *H* gene: a conceptual "supergene" that comprises a series of the neighboring genes tightly linked to each other (Clarke and Sheppard

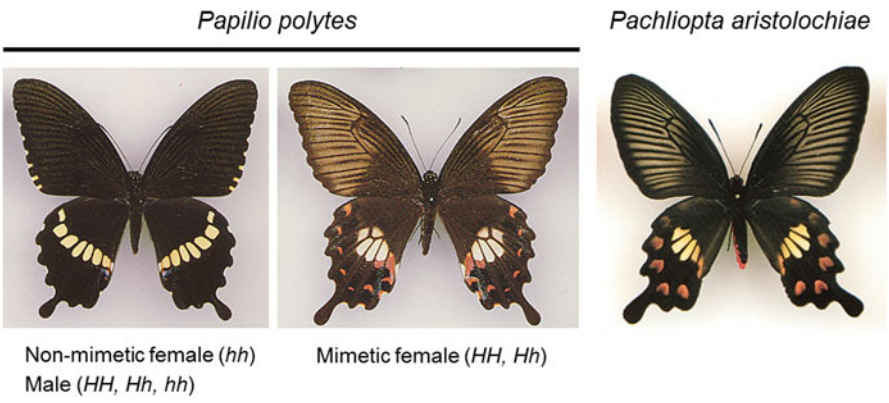

Fig. 10.1 *P. polytes* and model butterfly, *Pachliopta aristolochiae*

1960, 1972) and a single regulatory gene that controls downstream, unlinked genes affecting the color pattern. It is hypothesized that a supergene unit is created by recombination events and fixed by inhibitory effects of a chromosomal inversion on recombination (Nijhout 2003; Joron et al. 2011), although the mechanism underlying this hypothesis has remained obscure.

Recently, we found that drastic changes of gene networks not only in red but also pale-yellow regions can switch wing color patterns between non-mimetic and mimetic female of *P. polytes* (Nishikawa et al. 2013). It is presumed that these pigmentation processes involved in Batesian mimicry of *P. polytes* should be downstream of the *H* gene. To elucidate the evolutionary processes of this mimicry comprehensively, it is important to clarify the *H* locus and its structure and function. More recently, Kunte et al. (2014) and our group have identified the *H* gene locus and revealed its structure, independently (Nishikawa et al. 2015).

10.2 *Papilio* Genome Projects Reveal the *H* Locus and Chromosomal Inversion Near *dsx*

To reveal the *H* locus and its flanking structure, we first determined the whole genome sequences of *P. polytes* and *P. xuthus* for comparison (Nishikawa et al. 2015). We have prepared the *P. polytes* genome DNA (Ishigakijima Island strain in Japan) from one inbred female (genotype, *H*/*h*) after four generations of laboratory inbreeding and the *P. xuthus* genome DNA from a male captured in the field near Tokyo, Japan. We used a whole genome shotgun approach with next-generation sequencing platform. Filtered paired-end reads (135.2 Gb pairs for *P. polytes* and 73.8 Gb pairs for *P. xuthus*) were assembled using Platanus (version 1.2.1) (Kajitani et al. 2014) with some mate-pair libraries sequenced by Illumina Hiseq2000 and Hiseq2500. Consequently, we obtained 3873 and 5572 scaffolds, with an N50 of 3.7 Mb and 6.2 Mb pairs, spanning 227 Mb and 244 Mb pairs of the genome sequences for *P. polytes* and for *P. xuthus*, respectively.

In validating resulting assembled scaffolds, we noticed that there were two independent scaffolds including *dsx* in *P. polytes* while only one scaffold including *dsx* in *P. xuthus*. Because these two scaffolds in *P. polytes* were significantly different in sequences and the genome DNA was prepared from one heterozygous (*Hh*) mimetic female, we assumed that each haplotype (*H* or *h*) was highly diverged around the *dsx* locus in the two independent scaffolds. To survey such heterozygous regions in the whole genome of *P. polytes*, we picked windows in which the coverage depth was $\leq$350, which is approximately half the homozygous peak of 600. After clustering overlapping windows, we found 15 highly diverse (identity of $\leq$90%) and long ($\geq$100 kbp) heterozygous regions; 14 were mapped on heterozygous sex chromosome-1 (ZW) and one on chromosome-25 near *dsx* (Nishikawa et al. 2015). In the heterozygous region near *dsx*, we detected an approximately 130 kbp autosomal inversion (Fig. 10.2b). Strikingly, in the whole genome data of

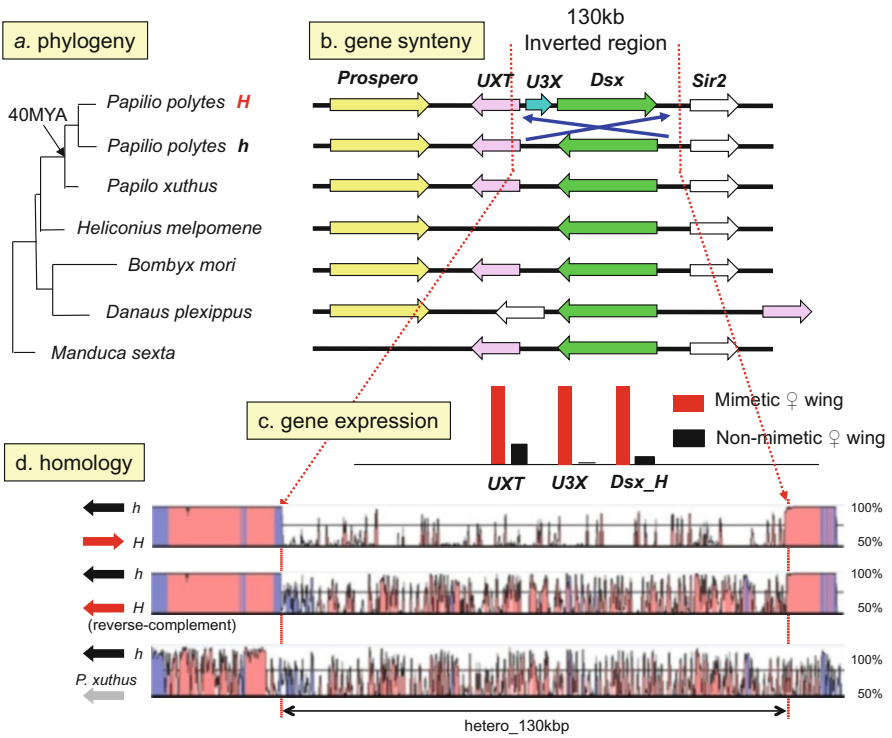

Fig. 10.2 Schematic view of the *H* locus in *P. polytes*. (**a**) Phylogenetic tree of Lepidoptera based on Dsx amino acid sequences. (**b**) Gene structure around *dsx* for the *h* and *H* alleles and gene synteny among Lepidoptera. (**c**) Expression level of genes within inverted region of the *H* locus, at early pupal stage in hind wings of mimetic (*red*) and non-mimetic (*black*) females. (**d**) Sequence homology around the *H* locus between *H* and *h* alleles and between *h* and *P. xuthus*

P. polytes, we could not find a long heterozygous region other than in the sex chromosomes (*Z*/*W*) which include many various repetitive sequences. Thus, the putative *H* locus region located on the chromosome-25 is thought to be only a long and unique heterozygous site among the whole autosomal chromosomes, which structure is maintained by reduced recombination due to the chromosomal inversion.

10.3 Linkage Mapping of the *H* Locus

To identify the mimicry locus *H*, we also performed the linkage mapping using non-mimetic type of *P. polytes* in Minamidaitōjima Island in Japan and mimetic type of *Papilio alphenor* in the Philippines (Nishikawa et al. 2015) (this work was performed mainly by Dr. H. Hori). After analyzing 84 F2 backcrossed females with

mimetic phenotype in heterozygote of *H* (*Hh*) and 69 of non-mimetic females (*hh*) using amplified fragment length polymorphism (AFLP) and restriction fragment length polymorphism (RFLP) markers, we mapped the mimicry locus in *P. polytes* within 800 kbp genomic region containing 41 genes between two markers designed in *kinesin* and *intermediate* on chromosome-25. The association between the region and mimicry phenotype in natural populations was further examined using single nucleotide polymorphisms (SNPs) in 54 wild-caught females (Nishikawa et al. 2015). Consequently, eight SNPs in *dsx* showed significantly higher association (chi-squared test of independence, $P < 10^{-10}$) but none outside the gene. This is consistent with the result of the association study by K. Kunte et al. (2014) using laboratory-reared *P. polytes alphenor*. It is noteworthy that the *H* locus revealed by linkage mapping coincides completely with the long heterozygous region revealed by whole genome sequencing. This means that a genetic locus responsible for some polymorphic trait with a long heterozygous region can be identified only by genome sequencing without linkage mapping.

10.4 Detailed Structure of a Long Heterozygous Region Linked to the *H* Locus

Gene prediction and RNA-seq mapping showed that most of the inverted region of the *H* locus was occupied by *dsx* and the intron/exon structure was reversed in the *h*- and *H*-chromosomes, suggesting that a simple inversion occurred near both ends of *dsx* (Kunte et al. 2014; Nishikawa et al. 2015). Sequence comparison of the inverted region between *H* and *h* showed low-level homology not only directly but also in reverse (Fig. 10.2d). However, it is remarkable that some scattered regions including exons (shown by blue) for *dsx* were highly conserved (Fig. 10.2d).

To estimate the evolutionary process of the chromosomal inversion between *H*- and *h*- chromosomes, we further compared the gene synteny around *dsx* of *P. polytes*, with those of other Lepidoptera (Fig. 10.2a, b). We found that all tested genomes (*Papilio* species, *Heliconius*, *Bombyx* and *Manduca*) except *Danaus* have the same oriented synteny as the *h*-chromosome of *P. polytes*. Only in *H*-chromosome of *P. polytes*, *dsx* resides in the reverse orientation. These observations suggest that the *H*-chromosome may have originated from *h*-chromosome by a single inversion. Based on the gene synteny, we speculate that different types of inversion may have occurred near *dsx* in the *Danaus* genome independently. When comparing the inverted region (named hetero_130kbp) with a corresponding region of *P. xuthus*, the homology between *h* and *P. xuthus* was a little bit lower than that between *h* and *H* (Fig. 10.2d). This fact and phylogenetic tree suggest that the chromosomal inversion may have occurred after the branch of *P. polytes* and *P. xuthus* (Fig. 10.2a, d).

To clarify structural features of the inverted region of the *H* locus, we identified the exact place for the chromosomal inversion (Nishikawa et al. 2015). We have

detected the sharp decline of the sequence conservation at both ends of inverted region between *H*- and *h*-chromosomes and considered these as putative breakpoints (Fig. 10.2d). The left breakpoint which closes on *Prospero* resides on about 700 bp downstream of the sixth exon of *dsx* in *h*-chromosome but on about 14.6 kbp upstream of the first exon of *dsx* in *H*-chromosome. The right breakpoint which closes on *Sir2* resides on about 8.9 kb upstream of the first exon of *dsx* in *h*-chromosome but on about 1.1 kb downstream of *dsx* in *H*-chromosome. Compared with *dsx* in *h*-chromosome (*dsx_h*), *dsx* in *H*-chromosome (*dsx_H*) was longer in the second, fourth, fifth, and sixth introns and sixth exon. Just outside of both breakpoints, in contrast, more than 99% homology was carried on between the *h*- and *H*-sequences (Fig. 10.2d). These structures implied that many mutations and several insertion and deletion events may have accumulated in the inverted region for *H* after the inversion and were maintained by suppression of recombination between two chromosomes.

10.5 Dimorphic Dsx Structure Associated with the *H* and *h* Alleles

The fact that a complete *dsx* was encoded inside of the inversion region indicates a possible involvement of the gene on the mimetic phenotype. RNA-seq assembly from mimetic (*HH*) and non-mimetic female (*hh*) revealed three types of female isoforms of *dsx* (F1, F2, and F3) in wings (Fig. 10.3). Dsx isoforms are generated by alternative splicing between the third and the forth exon both on the *h* and *H* alleles. Translational stop codon appeared in the fourth exon for F1 and for F3 and in the third exon for F2. Amino acid differences among isoforms were restricted merely in the C-terminal region (4–23 amino acids); three isoforms shared the first 244 amino acids including *dsx* DNA-binding motif and oligomerization domain (Fig. 10.3). Although there were 13–15 amino acid changes in three *dsx* isoforms between *H* and *h* alleles (Fig. 10.3), most substitutions occurred around the DNA-binding motif and dimerization domain (An et al. 1996). The comparison of *dsx* sequences among Lepidoptera showed that only five amino acids were specifically changed in *dsx_H* of *P. polytes* (Fig. 10.3, *). Recently, we have revealed dimorphic structure of Dsx sequences in another polymorphic, female-limited Batesian mimic species *P. memnon*, which shows different sites of amino acid changes between mimetic and non-mimetic alleles, in comparison with *P. polytes* (Komata et al. 2016). This finding suggests parallel evolution of the mimicry locus in two *Papilio* species, and further researches are necessary to clarify the structural features of Dsx involved in the mimicry traits.

Respective differences of amino acid sequence for F1, F2, and F3 between *dsx_H* and *dsx_h* are 2, 0, and 1, respectively (Fig. 10.3). This indicates that at least the C-terminal region of F2 may not be involved in the specific function of *dsx_H* on the mimetic phenotype. Furthermore, sequence comparison of these

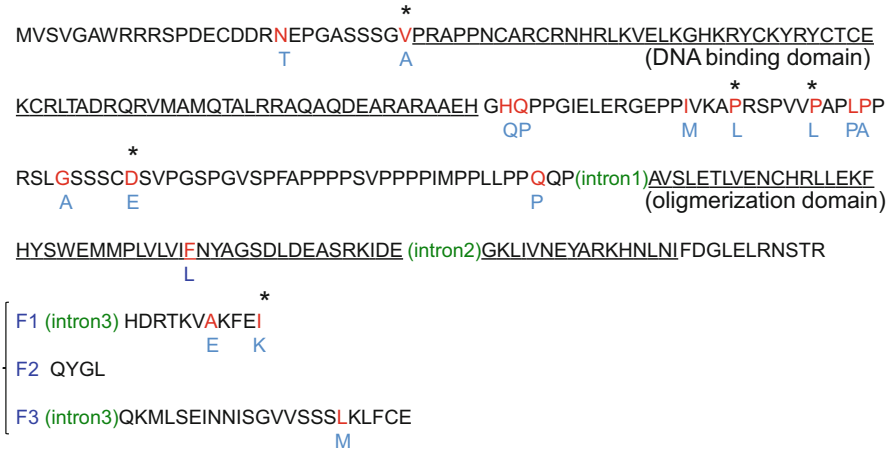

Fig. 10.3 Amino acid sequences of Dsx for the *H* and *h* alleles. The N-terminal 244 amino acid sequence of Dsx which is common to female-specific isoforms (F1, F2, and F3) is shown on *top*. Each sequence of three isoforms for C-terminal region is shown at the *bottom*. The *h* allele sequence (Dsx_h) is shown. The amino acids substituted in Dsx_H are shown in *blue* below the sequence. * indicates an amino acid residue which is substituted only in Dsx_H among Lepidoptera

isoforms with those in other Lepidoptera revealed highly conserved structure except the C-terminal amino acid in F1 isoform. These observations suggest that no special isoform of *dsx_H* seems to be involved in the mimetic wing coloration, although it needs further evidence to show this possibility.

In males which show merely non-mimetic phenotype, we found only one isoform of *dsx* which skips exons 3 and 4 included in all female isoforms, implying the importance of exons 3 and 4 for the mimicry. In these regions of three female isoforms, however, there was only one amino acid (the C-terminal end of F1) changed specifically in *dsx_H*, as described above. The male-specific isoform of *dsx_H* was scarcely expressed in prepupal to pupal wings, suggesting that male *dsx_H* is not involved in the mimetic phenotype (Fig. 10.4).

10.6 Expression Profiles of Genes Around the Inverted Region of *H* Locus

To clarify the transcribed regions around the inverted regions, we mapped reads of RNA-seq to *h* and *H* alleles and found that three independent transcripts near left breakpoints, *ubiquitously expressed transcript* (*UXT*, transcriptional regulator) (Schroer et al. 1999), *unknown-3-exons (U3X*, long noncoding RNA emerged in *H*), and unknown transcript downstream of *Prospero*. These genes were highly expressed in wings of mimetic females (*HH* or *Hh*) compared with that in wings in

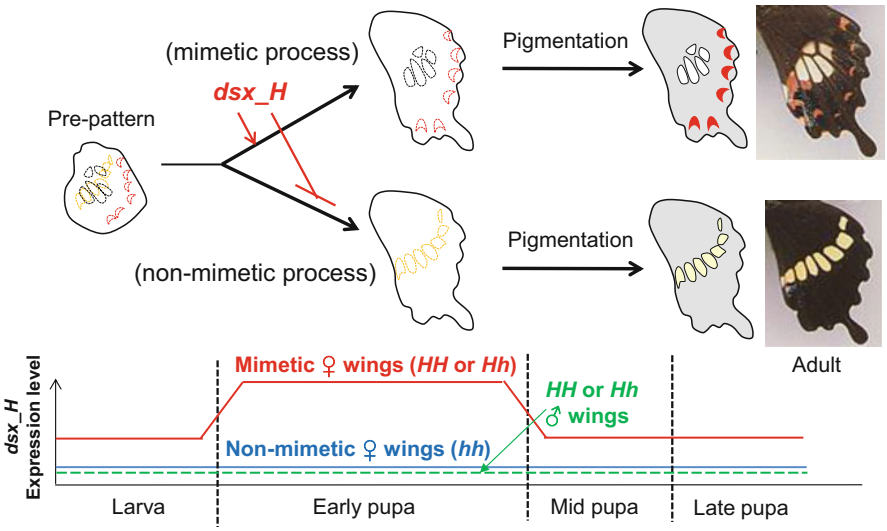

Fig. 10.4 Hypothetical model of Dsx_H function on mimetic and non-mimetic wing coloration patterns. The expression patterns of *dsx_H* in wings of mimetic (*red*) and non-mimetic (*blue*) females and males (*dotted green*) during larval to pupal stages are shown below

non-mimetic females (*hh*) (Nishikawa et al. 2015) (Fig. 10.2c), suggesting a possible involvement of these genes in the mimicry. The 5′ untranslated region (UTR) structure and transcriptional start site for *UXT* were altered by the inversion event between *H* and *h*, while the open reading frame was the same. A newly emerged gene *U3X* was found merely in the heterozygous region of the *H*-chromosome in the whole genome of *P. polytes*. The downstream sequence of *Prospero* was differently expressed between *h*- and *H*-chromosomes while located outside of inverted regions. These facts demonstrate that the chromosomal inversion affects not only the genome structure but also the expression of neighboring genes drastically probably through changes of gene regulatory elements.

We found that there seemed no significant differences in expression level of each isoform (F1, F2, and F3) of *dsx* between mimetic and non-mimetic wings in P1–2, P4–5, and P10.5 stages (Nishikawa et al. 2015). However, the expression level of *dsx_H* in mimetic female wings was quite higher than in non-mimetic wings in early pupal stages, but *dsx_h* did not show such expression profiles. RNA-seq analyses support the results that *dsx_H* was dominantly expressed in *Hh* female wings (Nishikawa et al. 2015). Differential expression level between *dsx_h* and *dsx_H* becomes significant on female wings at P2 stage when the patterning of wing pigmentation may be determined (Nishikawa et al. 2013) (Fig. 10.4). In contrast to *Hh* male, *dsx_H* was scarcely expressed, while *dsx_h* was dominantly expressed both in wandering and early pupal stages. In the report of Kunte et al. (2014), however, the expression pattern of *dsx_H* was upregulated at late pupal stage, which was different with our result.

The comparison of promoter regions of *dsx_H* and *dsx_h* showed highly nucleotide conservation near the transcriptional start site, but the conservation gradually reduced in more upstream regions (Nishikawa et al. 2015). Some of the nucleotide differences in the regulatory regions or intron regions between *dsx_H* and *dsx_h* may be responsible for the specific regulation of *dsx_H* in the female wings. The above results suggest that not only amino acid substitution but also regulatory changes for female *dsx_H* are possibly involved in the mimetic phenotype.

10.7 Functional Analysis of *dsx*

To verify the *dsx* function on the mimetic wing pattern formation, we performed the functional analysis with electroporation-mediated siRNA incorporation optimized for pupal wings, which enables mosaic analysis by knocking down target genes (Ando and Fujiwara 2013; Yamaguchi et al. 2011; Fujiwara and Nishikawa 2016). First, to confirm the validity of this newly established method, we knocked down *tyrosine hydroxylase* (*TH*) that is involved in melanin synthesis and found that the black pigmentation in adult wings was clearly repressed in the siRNA incorporated region. When injecting Universal Negative Control siRNA which is used generally as a negative control, no phenotypic change was observed.

Using this method, we injected siRNA designed to knock down *dsx_H* but not *dsx_h* into the whole hind wings of mimetic female and applied electroporation. This treatment caused the mimetic wing pattern switching to non-mimetic wing patterns. Furthermore, electroporation of *dsx_H* siRNA into part of early pupal hind wings of mimetic females also resulted in severe repression of mimetic wing patterns in the siRNA incorporated region; the peripheral red spots became the small pale orange ones; the central white marking mostly disappeared in the mosaic area. By this treatment, the ectopic white patterns for non-mimetic females emerged in the predicted position (Nishikawa et al. 2015). These results indicated that *dsx_H* not only induces the mimetic wing patterns but also simultaneously represses the emergence of the non-mimetic wing patterns. In contrast, *dsx_h* siRNA in mimetic females did not influence wing phenotype. When knocking down both *dsx_H* and *dsx_h* expression by siRNA which was designed in the common region between *H* and *h* (*dsx_H/h*), the same phenotype was observed as *dsx_H* siRNA alone (Nishikawa et al. 2015). These results implied that *dsx_h* is not involved both in mimetic and non-mimetic wing pattern formation. This strongly suggests that only *dsx_H* is involved in the mimicry phenotypes. It is noteworthy that *H/H* homozygous individuals are viable. This means that *dsx_H* should have basic functions for sexual differentiation in addition to the wing coloration.

It is unexpected that *dsx_H* not only induces the formation of mimetic color patterns but also represses non-mimetic patterns (Fig. 10.4). We previously showed that white (pale-yellow in mimetic) regions on mimetic and non-mimetic female wings of *P. polytes* are composed of different pigments (Nishikawa et al. 2013). Additionally, kynurenine/N-beta-alanyldopamine (NBAD) synthesis and *Toll*