Genes controlling mimetic colour pattern variation in butterflies
Nicola J Nadeau

Butterfly wing patterns are made up of arrays of coloured scales. There are two genera in which within-species variation in wing patterning is common and has been investigated at the molecular level, *Heliconius* and *Papilio*. Both of these species have mimetic relationships with other butterfly species that increase their protection from predators. *Heliconius* have a ‘tool-kit’ of five genetic loci that control colour pattern, three of which have been identified at the gene level, and which have been repeatedly used to modify colour pattern by different species in the genus. By contrast, the three *Papilio* species that have been investigated each have different genetic mechanisms controlling their polymorphic wing patterns.

Address
Department of Animal and Plant Sciences, University of Sheffield, Western Bank, Sheffield S10 2TN, UK

Corresponding author: Nadeau, Nicola J (N.Nadeau@sheffield.ac.uk)

**Current Opinion in Insect Science** 2016, 17:24–31
This review comes from a themed issue on Molecular physiology
Edited by Takema Fukatsu and Ryo Futahashi
For a complete overview see the Issue and the Editorial
Available online 27th May 2016
http://dx.doi.org/10.1016/j.cois.2016.05.013
2214-5745/© 2016 The Author. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

**Introduction**
Butterfly wing patterns are examples of evolutionary innovation that have fascinated scientists since the very inception of evolutionary theory [1]. The adaptive significance of these patterns has been established in many cases, and the main function is usually for defence against predators, for example as startle patterns [2], camouflage, or warning colours in chemically defended species [3]. Warning colours are also often shared between species, either through Müllerian mimicry, where multiple chemically defended species have the same pattern, increasing predator learning of these patterns [4], or through Batesian mimicry, where non-defended species copy the patterns of chemically defended species [5]. Wing colours and patterns can also function in mate choice and mate attraction [6], sometimes alongside an anti-predator function [3]. This dual function can lead to interesting evolutionary dynamics, for example the ability to function as ‘magic traits’ in speciation — causing both ecological divergence between populations with different patterns and reproductive barriers due to assortative mating [7].

Wing patterns in the butterflies and moths (Lepidoptera) are made up of arrays of coloured scales (Figure 1). These colours can be conferred either by pigments or by sub-micron-scale structures that produce interference colours (structural colour) or by a combination of these mechanisms [8]. Although the genetic pathways responsible for pigment production are fairly well characterised in most cases, virtually nothing is known about genes controlling structural colour. The most common pigment is melanin and the pathways producing this pigment from the amino acid precursor tyrosine are well known in insects [9,10]. Other butterfly wing pigments include ommochromes, pterins and flavonoids. The first two are synthesised from precursors tryptophan and guanosine triphosphate respectively, but the latter must be obtained from food plants [8].

Although the genetic control of pigment production is reasonably well understood, these genes appear to be fairly conserved in evolutionary terms and contribute relatively little to the variation in wing pigmentation pattern observed in butterflies [11,12] or moths [13], at least over short evolutionary timescales. This contrasts with what is known in vertebrates [14–17] and to some extent also other insects [18,19], and suggests that on the lepidopteran wing there is a greater disconnect between the genes responsible for producing pigments and those responsible for the evolution of colour patterning.

There are two major butterfly groups in which genetic variation underlying pattern variation has been investigated, *Heliconius* and *Papilio* (Figure 2). Both of these show widespread within-species variation in wing pigmentation pattern related to mimicry. This variability has made them excellent systems for identifying genes controlling pattern variation.

**The *Heliconius* ‘Tool Kit’**
As well as within-species variation in pigmentation pattern, the *Heliconius* butterflies have also been studied because of the often near-perfect mimicry between species. This mimicry has also made them an excellent system for studying the extent to which the same genes are used when evolving convergent traits [20]. Extensive genetic work on species within this genus (largely *H. erato, H. melpomene, H. cydno* and *H. numata*) has revealed a ‘tool kit’ of around five unlinked genetic loci (Figure 2)
that control almost all of the colour pattern variation in this group and that have been repeatedly used by different species to produce both convergent and divergent wing colour patterns [21–23,24*]. Over the last few years several of these have been pinned down to individual genes.

**Optix**

Fine-scale mapping and gene expression analyses have identified the transcription factor *optix* as being responsible for turning on and off most red, orange and brown colour pattern elements in *H. erato*, *H. melpomene* and *H. cydno* (Figure 2) [25]. In *Drosophila* the main function of *optix* is in controlling eye development [26]. However, the gene apparently took on a role in wing scale specification within the lepidoptera, initially controlling the development of specialised scales coupling together the forewings and hind-wings, and just within *Heliconius* has it taken on a role in colour patterning [27].

Population genomics approaches have identified a 65 kb interval ~100 kb downstream of *optix* that likely contains cis-regulatory elements controlling red colour patterns in both *H. erato* and *H. melpomene* [28,29]. Detailed analysis of this region in *H. melpomene* has revealed two discrete regulatory modules, one 10 kb in length containing variants that control red patches at the wing bases (‘dennis’ patches) and one 25 kb in length controlling red ‘rays’ on the hind-wing (Figure 3b) [30**]. It seems likely that each of these modules contain one or more transcription factor binding sites that specify the expression pattern of *optix*. However, discovering exactly what the functional variants within these regions are will likely remain unresolved until transgenic techniques are developed in these species. It is also presumed that there is a third, currently unidentified, regulatory module for *optix*, which controls the presence of a red forewing band [22].

**Cortex**

A second major locus is responsible for switching on and off most white and yellow colour pattern elements in *H. erato*, *H. melpomene* and *H. cydno* (Figure 2) [21,23]. Interestingly this locus also overlaps with two inversions present in certain morphs of *H. numata*, which control quite different colour patterns of black, orange and yellow spots [31]. *H. numata* differs from most other *Heliconius* species in that multiple colour patterns are usually present within a single population and that all colour pattern variation is controlled by multiple alleles at single genetic loci with a strict dominance hierarchy between these alleles [32]. The gene *cortex* appears to be, at least partially, responsible for these colour pattern variants, with population genomics approaches mapping colour pattern variation within *H. erato*, *H. melpomene* and *H. numata* to within or near this gene and *H. melpomene* and *H. numata* showing colour-pattern-associated expression differences of *cortex* [33**].

*Cortex* belongs to a family of cell cycle regulators [34], which includes two genes that are highly conserved in all eukaryotes, *CDC20*/*fzy* and *cdk1/fsr*, and have a fundamental role in cell cycle progression [35]. *Cortex* itself appears to be insect specific and to have a much higher evolutionary rate [33**]. It seems likely that it could control scale cell colour through control of scale developmental rate, as melanin scales are known to develop at a slower rate than scales of other colours across a diversity of lepidoptera [36]. Indeed, the *cortex* gene also appears to regulate melanin pigmentation in the peppered moth, with the insertion of a transposable element in this gene producing the melanin form that proliferated during the
industrial revolution [37*]. Therefore, it seems likely that cortex has a role in scale cell development and pigmentation across all lepidoptera.

Again, the precise functional variants of cortex causing differences in pigmentation patterning are unknown, but appear to be cis-regulatory rather than coding. Cortex has several 5' untranslated exons (5' UTRs) spanning a region of over 100 kb, suggesting a complex of dispersed regulatory elements [33**]. In addition to splicing variation of these 5' UTRs, there are also alternative coding isoforms, some of which show associations with colour pattern. Further work is needed to understand if this splicing variation affects scale pigmentation.
### WntA

A third gene, WntA, controls several aspects of the size and shape of the colour pattern elements switched on and off by the previous two loci in both *H. erato* and *H. melpomene* (Figure 2) [38]. Unlike the previous two genes, some functional information does exist for this gene, with pharmacological treatments that enhance wnt signalling increasing the amount of melanin pigmentation on the wing and mirroring the natural effects of this locus [38]. On the other hand this locus has not been fine-mapped in the same detail as the previous two, so the location of functional sites is less clear. The evidence again seems to point to cis-regulatory variation, although mapping data places these closer to the coding region than is the case for the previous two genes [38,39], and coding variants have not been completely ruled out.

Like *cortex*, WntA’s role in wing patterning seems fairly ubiquitous, at least within the nymphalid butterflies [40]. Further, WntA also controls colour pattern differences between Batesian mimetic and non-mimetic populations of the admiral butterfly *Limenitis arthemis* in the eastern USA. In this species colour pattern variation shows a perfect association, again with the insertion of a transposable element, upstream of the coding exons of WntA [41*].

### Other Heliconius loci

At least two other loci are known to control aspects of pigmentation patterning variation in *Heliconius*. Another locus controlling the shape of the forewing band has been found on *H. melpomene* chromosome 13 in both *H. erato* and *H. melpomene* (Figure 2) [39,42]. Further work is needed to identify the gene responsible, although the current mapping data implicates the *radial spoke head 3* gene [39].

The *K* locus controls a switch between yellow and white in *H. melpomene* and *H. cydno* (Figure 2) and has been mapped to a region of chromosome 1 that contains *wingless* [43], although the exact gene responsible is not known. Despite causing a simple switch in yellow pigment deposition it seems unlikely that the gene is involved in production of the yellow pigment since this is synthesised in the haemolymph, not *in situ* [44], and both yellow and white patterns can be present on the wing of a single individual with a particular *K* allele.

### The importance of gene-exchange for Heliconius pattern variation

In addition to this tool-kit of loci that can be used flexibly to generate a wide range of patterns, gene exchange between species also appears to have played an important role in pattern evolution in this group [45]. There are now several well-supported cases of species that have gained novel wing patterns as a result of rare hybridisation events with other species, allowing introgression of colour pattern genes [45–47]. This mode of evolution is likely to be particularly effective, as it means that an entire locus,
containing multiple co-evolved mutations that have built up over evolutionary time, can be acquired instantaneously. It is also likely to particularly advantageous in Heliconius, where positive frequency dependent selection drives mimicry between species [4], so a species moving into a new area can rapidly join a mimicry ring by acquiring genes from other species already in that area.

However, gene-flow between species appears to be able to do more than just transfer existing patterns between species. In some cases it also appears to be able to generate novel patterns. Recent work has shown that the two distinct modules producing the ‘dennis’ and ‘rays’ pattern in Amazonian H. melpomene and H. elevatus have distinct evolutionary origins, with dennis arising first in the ancestor in H. elevatus and then being shared with H. melpomene, and rays arising later in H. melpomene and then being transferred in to H. elevatus [30**]. Therefore the current phenotype of both of these species is a chimera of different patterns that evolved separately in each species with hybridisation acting to bring them together (Figure 3a).

Papilio supergenes
Within the swallowtail butterfly genus Papilio, female-limited Batesian mimicry has evolved multiple times, with males being non-mimetic and females mimicking other, chemically defended, species [48]. In several of these species the females are also polymorphic, often with a male-like non-mimetic morph and morphs that mimic either one or several toxic species [5]. The genes controlling the switch between different female morphs have often been described as ‘supergenes’ because of their ability to influence multiple aspects of the phenotype from a single genetic locus [49]. Two such genes, underlying female-limited polymorphism, have been identified (Figure 2). By contrast to the Heliconius system, the genes involved are not the same between different species, although both are transcription factors.

Papilio polytes
In this species there are multiple female morphs, including a non-mimetic male-like morph and three mimic morphs resembling distantly related, toxic, Pachliopta swallowtails. Two teams independently mapped the female-limited polymorphism to the doublesex (dsx) gene [50*,51**]. This autosomal gene controls sexual dimorphism in all insects that have been investigated [52]. Fascinatingly, in at least one of the mimetic morphs, dsx is inverted relative to the ancestral orientation found in the non-mimetic morph [51**]. This has repressed recombination between the mimetic and non-mimetic alleles, allowing multiple sequence differences to accumulate.

As in other insects, there are multiple female-specific splicing isoforms of dsx in P. polytes, but the studies disagree on whether these are differentially expressed between morphs [50*,51**]. However, knockdown of dsx confirmed the functional role of this gene in specifying pattern and implied that coding or structural differences found in the gene could be important [51**]. Knockdown of the mimic dsx allele produced a switch to a non-mimetic pattern, whereas knockdown of the non-mimetic allele in heterozygous individuals, which have the mimetic phenotype (it is dominant), produced no phenotypic effect, suggesting that changes in the expression level of dsx alone are insufficient to produce a change in colour pattern. Nevertheless, there must also be some regulatory component that prevents the mimic dsx allele from affecting male phenotype.

Papilio dardanus
This species also has multiple mimetic female morphs, but in this case they mimic very distantly related nymphalid butterfly species and non-mimetic female morphs are less common [53]. Mapping and population genomics analyses have identified the gene responsible for switching between morphs as the autosomal gene engrailed [54*]. No inversions were present in the region, but one of the morphs had a duplication of engrailed, which could similarly act to reduce recombination and promote divergence between the alleles. In this case too, coding sequence changes are present and may have a functional role, although this remains to be tested. Engrailed expression patterns have previously been shown to correlate with adult wing patterns in the butterfly Bicyclus anynana [55], suggesting that the transcription factor may have a widespread role in regulating butterfly wing colour patterning.

Other Papilio species
Papilio memnon is similar to P. dardanus in having a large number of female morphs that are largely controlled by a single genetic locus [56], but the molecular genetics in this species has not been investigated. Papilio [Pterourus] glaucus has a mimetic and a non-mimetic female morph, largely controlled by a locus on the W chromosome, with a low frequency of a Z-linked modifier alleles coming from hybridisation with P. canadensis [57]. The fact that control is sex-linked in this species demonstrates that the genes involved are again distinct from those controlling polymorphism in P. polytes or P. dardanus.

Conclusions
‘Supergenes’ controlling butterfly colour were initially thought to be made up of multiple tightly linked genes [49]. However, in both investigated Papilio species only a single gene seems to be involved [51**,54*]. The situation in Heliconius is less clear. H. numata, the only species with classical supergene architecture, does have large inversions between different mimetic alleles, which lock together multiple genes [31]. Current evidence points to just one of these genes, cortex, as having a major effect on phenotype [33**], but it is still too early to say whether
other genes in the inversions might act with cortex to produce some of the phenotypes. Molecular investigation of other systems has also shown that in some cases supergenes can involve the action of multiple genes locked together in inversions [49,58,59].

Something that is clear is that these loci have not evolved through genomic rearrangements that have brought together previously unlinked genes from around the genome [60]. Indeed, in Heliconius the steps involved in building a supergene can be observed. The three major loci described above (optix, cortex and WntA) control most colour pattern variation not only in the co-mimetic species H. erato and H. melpomene but also in H. hecale and H. ismenius, which have spotted patterns like H. numata (Figure 2) [24*]. This, together with other studies [39], illustrates that each of these loci can have diverse effects on phenotype and that these effects can sometimes be overlapping and can vary in their magnitude. It is therefore not a great leap to see how accumulation of mutations concentrated at just one of these loci could take on broad phenotypic effects, with polymorphism at other loci being reduced. Indeed traces of this process can still be seen in H. numata, where variation linked to wntA, optix, K, and chromosome 13 was found to have minor effects on phenotype (Figure 2) [61].

This also demonstrates that the genetic variants in these systems are in fact the product of a, probably lengthy, process of refinement, that has likely led to a reduction in the number of loci controlling colour pattern. Selection will act against unfit recombinant phenotypes, and will be strongest in fully polymorphic populations and weaker (but still present) where morphs are parapatic [32]. Therefore, we need to be cautious about making inferences from these systems about the earliest stages of divergence and the distributions and effect sizes of the first mutations that were targeted by selection. It is likely that multiple mutations at each of these loci have led to the current polymorphic alleles, and evolution may also have been facilitated by mutations at unlinked loci, at which polymorphism was later lost due to selection [60].

A key remaining question is why the patterns of gene re-use are so different between Heliconius and Papilio, especially when, superficially, the workings of the different Papilio species seem so similar. One obvious possibility is the different forms of mimicry involved: Heliconius are Müllerian mimics, with different species converging on the same patterns, while the Papilio species mimic different, distantly related species. Maybe this is why the different Heliconius species use the same loci, while the Papilio species do not. However, the use of the same loci even in species that have very different patterns, like H. melpomene and H. hecale, suggests that this is not the whole story. Another plausible explanation could be the ubiquity of both colour pattern polymorphism and gene-flow throughout Heliconius. This could have helped to maintain polymorphism at the tool-kit genes, making them predictable targets for selection whenever a new colour pattern became favourable. By contrast, the polymorphic Papilio species are more sparsely distributed both geographically and on the phylogeny, perhaps making the evolution of mimicry somewhat more ‘independent’ between each species. This argument does not hold for other examples of gene re-use, however. For example, why the peppered moth and admiral butterfly have also used two of the Heliconius tool-kit genes [37*,41*], as these events are clearly evolutionarily distinct. Ultimately, the question of what drives patterns of gene use can only be answered by comparing more systems and understanding the genetic basis of further adaptive and polymorphic traits.

Acknowledgements

NJJ is funded by a NERC Independent Research Fellowship (NE/K008498/1).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest


Using population genomics approaches the authors identify discrete regulatory modules of the *optix* gene controlling colour pattern elements and show that these have distinct evolutionary histories, arising in different species before being brought together by hybridisation.


Using population genomics and gene expression analyses the authors identify the cell cycle gene cortex as controlling major aspects of colour pattern variation in several *Heliconius* species.


Using population genomics and gene expression analyses the authors identify that an insertion of a transposable element into an intron of the cortex gene caused industrial melanism in the peppered moth. This is the same gene that controls major aspects of colour pattern variation in *Heliconius* butterflies.


Association mapping demonstrates that the insertion of a transposable element within an intron of the WntA gene is responsible for colour pattern variation in *Limenitis arthemis*. *In situ* hybridisation and harpin injections confirm that wntA has similar effects on wing pattern in both *Heliconius* and *Limenitis*.


Using mapping, genomics and expression analyses the authors identify the doublesex gene as controlling colour pattern polymorphism in Papilio polytes.


From whole-genome sequencing the authors identify an inversion of the doublesex gene associated with mimetic colour pattern. They perform knock-down experiments to show that this allele is responsible for switching colour pattern.


Using population genomics the authors identify the engrailed gene as being associated with colour pattern polymorphism in Papilio dardanus. They also identify a duplication of this gene in one morph.


