Colour pattern specification in the Mocker swallowtail *Papilio dardanus*: the transcription factor *invected* is a candidate for the mimicry locus *H*

Rebecca Clark^{1,2,†}, Sarah M. Brown^{3,‡}, Steve C. Collins⁴, Chris D. Jiggins⁵, David G. Heckel^{3,6} and Alfried P. Vogler^{1,2,*}

¹Department of Entomology, Natural History Museum, Cromwell Road, London SW7 5BD, UK

²Division of Biology, Imperial College London, Silwood Park Campus, Ascot SL5 7PY, UK

³CESAR, Department of Genetics, University of Melbourne, Parkville, Vic. 3010, Australia

⁴African Butterfly Research Institute, PO Box 14308, 0800 Nairobi, Kenya

⁵Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK

⁶Max Planck Institute for Chemical Ecology, Hans-Knöll Strasse 8, 07745 Jena, Germany

The swallowtail butterfly, *Papilio dardanus*, is an iconic example of a polymorphic Batesian mimic. The expression of various female-limited colour forms is thought to be controlled by a single autosomal locus, termed *H*, whose function in determining the wing pattern remains elusive. As a step towards the physical mapping of *H*, we established a set of 272 polymorphic amplified fragment length polymorphism (AFLP) markers (*EcoRI-MseI*). Segregation patterns in a ‘female-informative’ brood (exploiting the absence of crossing over in female Lepidoptera) mapped these AFLPs to 30 linkage groups (putative chromosomes). The difference between the *hippocoon* and *cenea* female forms segregating in this family resides on a single one of these linkage groups, defined by 14 AFLPs. In a ‘male-informative’ cross (markers segregating within a linkage group), a pair of AFLPs co-segregated closely with the two female forms, except in four recombinants out of 19 female offspring. Linkage with these AFLP markers using four further female-informative families demonstrated that the genetic factor determining other morphs (*poultoni*, *lamborni* and *trimeni*) also maps to this same linkage group. The candidate gene *invected*, obtained in a screen for co-segregation of developmental genes with the colour forms, resides in a 13.9 cM interval flanked by the two AFLP markers. In the male-informative family *invected* co-segregated perfectly with the *hippocoon/cenea* factor, despite the four crossovers with the AFLPs. These findings make *invected*, and possibly its closely linked paralogue *engrailed*, strong candidates for *H*. This is supported by their known role in eyespot specification in nymphalid butterfly wings.

Keywords: evolutionary genetics; mimicry; phenotype–genotype association; *engrailed*; candidate genes; amplified fragment length polymorphism

1. INTRODUCTION

The African Mocker swallowtail, *Papilio dardanus*, is among the most striking cases of Batesian mimicry and a textbook example of adaptive evolution (Ridley 2003). At least a dozen colour forms have been recognized in this species throughout its range in sub-Saharan Africa, including Bioko island in the Atlantic Ocean and Madagascar and the Comoro Islands in the Indian Ocean, and as many as six forms may be present in a single population (Ford 1936). Mimetic forms are limited to the female sex, while males are monomorphic and non-mimetic. Only in the Indian Ocean populations (and

some Ethiopian highland populations), females are ‘male like’ and closely match the general appearance of the males. Most female forms have been named and refer to specific colour patterns that are mimics of different chemically defended butterflies and day-flying moths. The forms vary in the extent of their black patterning and in the colour and layout of the background pigmentation, which can range from white to a deep red (e.g. *hippocoon* versus *poultoni*; figure 1a).

The seminal breeding work of Clarke & Sheppard (1959, 1960a,b, 1962) showed that the colour variation is largely determined by an autosomal locus, termed *H*, that exhibits at least 10 distinct alleles. Pedigree crosses showed that each female colour form corresponds to a distinct allele of *H* (including *H^c*—the dominant *cenea*-defining allele, and *H^h* or *h*—the universally recessive *hippocoon*-defining allele; figure 1b), which segregates as a single, multi-allelic locus with strict dominance hierarchies among the alleles. The existence of secondary ‘modifier loci’ involved in refining the mimetic wing

* Author and address for correspondence: Department of Entomology, Natural History Museum, Cromwell Road, London SW7 5BD, UK (a.vogler@nhm.ac.uk).

† Present address: Department of Craniofacial Development, King’s College London, Guy’s Tower, London SE1 9RT, UK.

‡ Present address: Department of Life and Environmental Sciences, Deakin University, 22 Burwood Hwy, Burwood, Vic. 3125, Australia.

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rspb.2007.1762> or via <http://journals.royalsociety.org>.

Figure 1. Female colour forms and broods for mapping used in this study. (a) Wing patterns and their nomenclature for female forms used in crosses. (b) The female-informative Brood 99 involving the segregation of forms *cenea* (H^c) and *hippocoon* (h), with the genetic interpretation of parents and offspring and the number of offspring. (c) The male-informative Brood 59.

patterns was mainly inferred from crosses between geographically distant *P. dardanus* populations ('races'). They have been interpreted to support the hypothesis that mimicry evolves in a two-step process by which a first mutation generates substantial resemblance to the model conferring initial protection (genetically controlled at H), followed by the selection for higher accuracy of the patterns mainly at other loci (Clarke & Sheppard 1960c; Turner 1984). However, owing to their minor role in colour pattern determination, these loci have not been genetically characterized.

In addition to extensive genetic studies, other work using phylogenetics (Vane-Wright *et al.* 1999), morphometrics (Nijhout 2003) and mathematical modelling (Sekimura *et al.* 2000) has been conducted to study the function, genetic architecture and evolution of H . Early literature argued for the control of the phenotype by a 'supergene' composed of multiple genes that each determines a particular portion of the wing pattern, which have become closely linked to allow the maintenance of advantageous combinations (distinct mimicry

patterns) as a unit (Clarke & Sheppard 1960c). However, various elements of the colour patterns are apparently controlled independently (Nijhout 1991, 2003) and current understanding of the hierarchical nature of the developmental patterning mechanisms suggests that H might constitute a regulatory gene controlling the phenotypic switch through differential expression of (monomorphic) downstream genes (Nijhout 1994).

Studies on gene expression during eyespot determination in nymphalid butterflies (Brakefield & French 1999; Brunetti *et al.* 2001) illustrate the co-option of genes with known roles in development to butterfly wing patterning and provide a wealth of candidate regulatory genes for the control of colour forms in *P. dardanus*. A candidate gene approach using this information has been successful in other butterfly species, providing support for the gene *Distal-less* as the eyespot organizer gene in *Bicyclus anynana* (Brakefield *et al.* 1996) and illustrating the linkage of both wing colour preference and mate preference with the *wingless* gene in *Heliconius* species (Kronforst *et al.* 2006). An alternative strategy for the identification of functional loci is based on whole genome mapping with amplified fragment length polymorphisms (AFLPs; Vos *et al.* 1995) as random molecular tags linked to the loci of interest. This strategy has been highly successful in Lepidoptera to map insecticide resistance genes in moths (Gahan *et al.* 2001) and wing pattern genes in *Heliconius* (Jiggins *et al.* 2005; Tobler *et al.* 2005; Joron *et al.* 2006; Kapan *et al.* 2006). Here we applied a combination of these approaches to identify markers co-segregating with the female forms in *P. dardanus*, as a first step towards a molecular identification of the enigmatic H locus.

2. MATERIAL AND METHODS

(a) Breeding and specimen preparation

Crosses were conducted with individuals collected as pupae from Mount Kenya (ssp. *Papilio dardanus polytrophus*) and the Shimba Hills region of Kenya (ssp. *Papilio dardanus tibullus*), and subsequently reared and bred under temperature-controlled conditions. A total of 176 matings were carried out, over three breeding seasons, by hand pairing (Clarke & Sheppard 1956). Mated females were isolated in individual cages under sunlight-adjusted light for 8 hours daily and were fed a weak solution of honey in distilled water. Freshly cut branches of food plants were supplied in the cages for egg deposition. Larvae were reared on Mexican Orange Flower, *Choisya ternata* (Clarke 1959). Crosses involved females of all available colour forms: 53 *hippocoon*; 56 *cenea*; 19 *lamborni*; 28 *poultoni*; 16 *planemoides*; and 4 *trimeni*.

To interpret the crosses we used the female colour forms to infer the allelic state at the hypothetical H locus, under the assumption that a single, multi-allelic locus H segregates for a number of alleles; that colour pattern is entirely specified by H ; and by using the recessiveness/dominance relationships (*hippocoon* < *cenea* < *lamborni* < *poultoni*) compiled by Clarke & Sheppard (1959, 1960a,b, 1962). Only those crosses for which the parental genotypes could be identified with certainty from the segregation patterns of the phenotype among the offspring were used for subsequent steps. DNA was extracted from each individual in the selected families using small sections of the abdomen and the DNeasy tissue kit from Qiagen (insect protocol).

(b) Building an AFLP map

The generation of AFLPs used a two-step process (Heckel *et al.* 1999) and initially took advantage of the lack of intra-chromosomal recombination during meiosis in female Lepidoptera in order to identify the linkage groups and to assign the *H* locus to one of them. For this purpose, an initial 'female-informative' family was selected in which the dominant phenotype has been inherited from the female parent, i.e. the locus controlling the colour forms does not recombine relative to other loci on that chromosome. A 'male-informative' family (dominant phenotype inherited from the male parent, i.e. affected by the crossover *within* a linkage group) was subsequently used to identify the AFLPs most tightly linked to *H*.

The AFLP Analysis System 1 (Life Technologies, Gaithersburg, MD, USA) was used according to the manufacturer's instructions, with the substitution of *Eco*RI primers having two instead of three selective bases. The *Eco*RI primer was labelled with ³³P, and second-stage amplification products were run on a 6% denaturing acrylamide sequencing gel and visualized using Kodak X-Omat film. Twenty-one primer combinations were screened initially, and the 14 producing the largest number of polymorphisms were scored in the female-informative family. Customized software (DBMLnk3.p; Heckel *et al.* 1999) was used to identify groups of co-segregating AFLPs. This program detects groups of co-segregating, non-recombinant loci while allowing for a small number of genotyping errors, to avoid artificially inflating the number of independently segregating groups. Owing to the absence of AFLP information on the grandparents of the broods, the gametic phase of each marker was not known. Therefore both phase possibilities were considered for each AFLP marker in the grouping. After grouping, the more frequent of the two possibilities was chosen to arbitrarily represent the gametic phase, such that presence of the AFLP band so chosen was coded by a one and its absence by a zero. Finally, the segregation pattern across progeny of each AFLP linkage group corresponding to a different chromosome was summarized as an array of ones and zeros.

To screen for AFLP markers tightly linked to the *H*-locus in a male-informative family, offspring was analysed by bulked segregant analysis (Michelmore *et al.* 1991). All 32 combinations of the *Eco*RI primers *Eco*-ACA, *Eco*-ACC, *Eco*-ACG and *Eco*-AGG with the *Mse*I primers *Mse*-CAA, *Mse*-CAC, *Mse*-CAG, *Mse*-CAT, *Mse*-CTA, *Mse*-CTC, *Mse*-CTG and *Mse*-CTT were used. Three pools of offspring DNA were created, containing all females carrying the dominant *H* phenotype, all recessive females and all males, respectively. AFLP bands present in the first and third pools and the father, but not the mother or the second pool, were identified. These bands were then screened on all offspring separately to verify the association with the *H*-locus.

To obtain the sequence information for AFLP bands of interest, they were excised from the dried gel after rehydration using the superimposed autoradiogram as a guide to identify relevant parts of the gel, followed by PCR amplification with the original primers. The products were cloned into pGEM-T (Promega, Madison, WI, USA) for DNA sequencing. Specific PCR primers were then designed from the sequence obtained (table 1 in the electronic supplementary material) for genotyping. Sequencing was with BIGDYE terminator technology and an ABI sequencer.

(c) Amplification of candidate genes

Genes with known functions in butterfly wing patterning were selected for genetic mapping. Primers for *apterous*, *distal-less*, *engrailed*, *invected*, *patched*, *scalloped* and *wingless* were from Kronforst (2005) and primers for *hedgehog*, *decapentaplegic*, *notch*, *delta*, *cubitus interruptus* and *ultrabithorax* were newly designed here from aligned sequences for insects available at GenBank, attempting to minimize degeneracy. A total of 50 primer pairs were tested for amplification under various annealing temperatures and in many cases temperature gradients to optimize annealing. Successful primer pairs that are not available in the literature, and their annealing temperatures, are given in the electronic supplementary material, table 1. PCR fragments were cloned using a pMOSblue blunt-ended cloning kit (Amersham) and then amplified and sequenced using the vector-derived primers T7 and M13 (table 1 in the electronic supplementary material). The identity of the amplified gene, and location of the amplified region, was then confirmed through blastx searches against the GenBank database. Introns were identified using the GeneSplicer software (www.tigr.org/tdb/genesplicer/gene_spl.html).

(d) cDNA synthesis and rapid amplification of cDNA ends

The rapid amplification of cDNA ends (RACE) protocol for the amplification of full-length cDNA for the candidate gene *invected* was conducted on total RNA extracted from final instar larval wing discs. RNA extractions were conducted immediately following dissection using the RNeasy Kit (Qiagen). The BD Smart RACE cDNA Amplification Kit (Clontech) was used for cDNA preparation and RACE. First-strand cDNA synthesis was from 1 µg of total RNA following the protocol supplied with the kit. RACE was conducted in the 5' direction using primer 5'-CTGGAACCAGATCTTGATCTGCG specific to the conserved homeobox region near the 3' terminus. Primer design was from initial sequences obtained for the 3' region from genomic DNA as described previously. Amplification was again conducted according to the Clontech protocol and resulted in a single fragment of 1700 bp. The amplified fragment was cloned into a pMOSblue vector, again amplified from there and then sequenced.

(e) Screening for marker co-segregation

Identification of parental genotypes for the AFLP-derived markers was by the sequencing of PCR products using primers and annealing temperatures supplied in the electronic supplementary material, table 1 and cloned into pMOSblue vectors. Cloning was required due to insertion/deletion (indel) polymorphisms in these markers; between 4 and 10 clones were sequenced from each specimen to ensure both the alleles were identified. Sequence editing and alignment was performed in SEQUENCHER (v. 4.1).

Heterozygotes in parental alleles of the candidate genes were identified as double peaks in an otherwise clean sequence by direct sequencing and confirmed by sequencing in both directions. Polymorphic sites that were heterozygous in the informative parent and homozygous in the other parent were used for the unambiguous identification of the allele that had been inherited from the informative parent in offspring sequences. Screening for marker co-segregation was performed by amplifying each marker in all offspring specimens, in some cases taking advantage of size differences between

Table 1. Segregation of markers within broods. (Brood sizes are given as the number of female offspring of each segregating phenotype (both phenotype names and the number of female offspring are given in the order of decreasing dominance) followed by the number of males. Alleles in the AFLP markers (*ACT*, *Pd*) were scored with a variety of assays identifying the female and male genotypes in the offspring to assess co-segregation with *inv* (scored through direct sequencing of either the 5' or 3' region). Male-inherited alleles from each brood (with the exception of 14 whose male was homozygous for *ACT*) were used to assess the number of recombinants between the markers, and these numbers are given in parentheses. Note that the number of scored individuals do not always add up to the total number of individuals in the broods due to the failure of PCR in some cases. F, female; M, male; Fi, female-inherited alleles; Mi, male-inherited alleles; rec, recombinants.)

brood	segregating phenotypes	M/F informative	brood size		allele scoring		<i>inv</i> region	linkage	
			F	M	<i>Pd</i> Fi (Mi)	<i>ACT</i> Fi (Mi)		<i>Pd-inv</i> (rec)	<i>inv-ACT</i> (rec)
99	<i>cenea/hippocoon</i>	F	6/5	11	AFLP	AFLP	—	—	—
59	<i>cenea/hippocoon</i>	M	13/11	18	n.a. (size)	n.a. (<i>Nla</i> III)	5'	35 (4)	35 (1)
41	<i>lamborni/cenea/hipp</i>	F	7/6/5	21	<i>Rsa</i> I	sequence (<i>Spe</i> I)	3'	36 (2)	37 (2)
09	<i>trimeni/hippocoon</i>	F	13/18	22	<i>Mse</i> I (size)	<i>Rsa</i> I	3'	49 (4)	49 (2)
06	<i>poultoni/lamborni</i>	F	5/2	0	sequence	sequence	5'	7 (1)	7 (0)
14	<i>poultoni/hippocoon</i>	F	3/3	0	sequence	sequence	—	—	—

the alleles or restriction enzyme sites, which distinguished alleles when digested PCR products were electrophoresed on 3% Electran high-resolution agarose. Co-segregation of polymorphic sites was analysed in MAPMAKER/EXP (v. 3.0; Lander *et al.* 1987).

3. RESULTS

(a) Production of families for genetic mapping and AFLP analysis

Among 176 crosses performed, 57 (32.4%) produced adult offspring. The parents of only 14 broods could be genotyped for *H* with certainty (§2) and only 6 of these contained over 5 female offspring and over 10 offspring in total. A female-informative family (Brood 99; figure 1b, table 1) was chosen for two-generation segregation analysis of AFLPs. This cross between a heterozygous *cenea/hippocoon* female and a homozygous *hippocoon* male produced six *hippocoon* and five *cenea* daughters (plus 11 male offspring that contributed AFLP information). A total of 272 variable AFLPs were obtained and could be assigned to 30 linkage groups with a minimum of 4 (linkage group 30) and a maximum of 16 (linkage group 2) loci to each group (table 2 in the electronic supplementary material). The linkage group in which all markers co-segregated with offspring gender corresponds to the segregation of the Z and W chromosomes from the mother and was defined as linkage group 1. The remaining autosomal linkage groups were named in the decreasing order of marker numbers. Colour pattern segregation in this family was confidently assigned to linkage group 4, defined by 14 co-segregating AFLPs. In linkage group 4, one allele of each AFLP is exclusively found in *hippocoon* and the other in *cenea* among the daughters of Brood 99; no other linkage group has this property. Most of the segregating AFLPs were dominant, except for one of the AFLPs mapping to linkage group 4, which was found to be co-dominant. EcoAT-MseCAG selective primers revealed two allelic fragments of 608 bp (present in *cenea* daughters) and 617 bp (present in *hippocoon*). This marker was cloned and sequenced for screening in further families and is hereafter arbitrarily named *Pd*.

A male-informative cross between a male heterozygous for *cenea* and *hippocoon* and a homozygous recessive

(*hippocoon*) female (Brood 59; figure 1c, table 1) permitted estimates of the distance of a given molecular marker from *H* if located on the same chromosome. This cross produced 11 *hippocoon* and 13 *cenea* daughters (plus 18 male offspring) reflecting segregation of the two alleles in the male parent. Brood 59 was the only male-informative family produced in this study that contained sufficient adult offspring for genetic analysis. Bulked segregant analysis on Brood 59 using separate pools of *cenea*, *hippocoon* and male offspring identified a 400 bp AFLP that co-segregated with the female phenotype, despite the presence of recombination within its linkage group. This marker was also developed for further screening and was named *ACT*.

Both markers were highly variable except for the female parent of Brood 59, which was homozygous for both *Pd* and *ACT*. Alleles segregating in each family were distinguished through either size differences or restriction fragment length polymorphism assays (table 1). The positions of *Pd* and *ACT* relative to *H* were established from a recombinational map in the male-informative Brood 59 (figure 2a). The 12 *cenea* offspring that could be scored (out of 13 total) in this brood inherited alleles *ACT*-59-3, indicating complete linkage with *H*, while only 9 of 12 inherited *Pd*-59-4 due to three recombinants between *H* and *Pd*. All 9 *hippocoon* offspring successfully scored (out of 11 total) showed co-segregation of *Pd*-59-3; however, one individual showed *ACT*-59-3 instead of the *ACT*-59-4 expected under complete linkage, indicating a further crossover (figure 2). No additional recombinants between *Pd* and *ACT* were found in the males of this cross. The fact that recombination appears to have occurred only between one marker and *H*, but not between both markers and *H* (figure 2), suggests that *Pd* and *ACT* are flanking *H* on either side. MAPMAKER/EXP analysis of these data gave this arrangement as the most likely (log likelihood = 17.46) with a map distance between *Pd* and *H* of 7.9 cM and between *H* and *ACT* of 4.2 cM giving a total map distance of 12.1 cM (figure 2a). Two alternative arrangements with *H* to either side of the AFLP markers gave map distances of 17.3 and 23.3 cM and log-likelihood values of -18.38 and -19.41, respectively.

Table 2. Blastx similarity for developmental genes amplified in this study. (Species and GenBank accession numbers of aligned protein sequences identified from blastx searches of GenBank, and GenBank accession numbers of the *P. dardanus*-specific sequences of each of the successfully amplified candidate genes.)

locus	aligned protein sequence	amplified region ^a	score (bits)	expect (identities %)	GenBank accession
<i>invected</i> (<i>inv</i>)	<i>Bombyx mori</i> P27610	3–314	278	4×10^{-77} (60)	EU124665
<i>engrailed</i> (<i>en</i>)	<i>Bombyx mori</i> P27609	279–333	70.9	1×10^{-12} (94)	EU157963
<i>wingless</i> (<i>wg</i>)	<i>Bombyx mori</i> P49340	241–360	226	2×10^{-59}	EU157964
<i>decapentaplegic</i> (<i>dpp</i>)	<i>D. melanogaster</i> P07713	513–568	98.6	4×10^{-21} (71)	EU157962
<i>cubitus interruptus</i> (<i>ci</i>)	<i>D. melanogaster</i> P19538	521–552	72.0	4×10^{-13} (96)	EU157961

^a Amino acid position in full-length protein sequence.

(b) The transcription factor gene *invected* co-segregates with *H*

Among the developmental genes tested for co-segregation, only a small proportion could be successfully amplified. *Papilio dardanus*-specific sequences for *wingless*, *cubitus interruptus*, *decapentaplegic* and *engrailed* were readily obtained by PCR amplification from cDNA, but of these only *wingless* could be amplified consistently from genomic DNA. The identities of the sequences were confirmed by blastx searches against the GenBank Swissprot database (table 2). *Wingless* showed polymorphic sites only in the female parent of Brood 41 (below) and screening of *wingless* co-segregation among the offspring of this family showed that *wingless* is not located on the same linkage group as *H* (data not shown). By contrast, the conserved 3' homeobox region of *invected* (*inv*) was amplified consistently from genomic DNA using published primers (Kronforst 2005), as confirmed against the Swissprot database. The 3' fragment of the putative *inv* locus includes part of the homeobox sequence and a 222 bp intron, and shows 83% amino acid identity with the *Bombyx mori* *Invected* amino acid sequence (residues 358–432) but only 62% identity with the equivalent region of the *Engrailed* protein. *Engrailed* (*en*) is a paralogue of *inv* that shares several conserved DNA binding domains, including the 3' homeobox domain (Peel et al. 2006) but the levels of sequence similarity support the conclusion that *inv* rather than *en* has been amplified.

When testing for co-segregation, single nucleotide polymorphisms (SNPs) in the 3' region of *inv* showed 100% linkage with the colour form inherited from the female parent in the female-informative Brood 99. This confirmed that *inv* also maps to the same linkage group as *Pd*, *ACT* and *H*. The male-informative Brood 59 proved to be homozygous for this region of the *inv* gene. Hence, the highly conserved homeobox region was used to obtain an additional fragment of the gene through 5' RACE and the development of new primers. The resulting fragment near the 5' end of the gene was much less conserved than the 3' region and showed only 58% amino acid identities with *Invected* (*B. mori* residues 35–301); however, it contained a 5' *Invected* signature region not present in *Engrailed* (Peel et al. 2006). Sequence comparisons with *B. mori* therefore fully discriminate between *Invected* and *Engrailed* and identified the cloned fragments as the *inv* gene.

Sequencing of the 5' *inv* region in Brood 59 revealed polymorphisms separating parental alleles and showed perfect co-segregation of *inv* alleles and female colour form in the 19 female offspring that could be scored for

Figure 2. Segregation pattern and genetic distances in the *H* region. (a) Co-segregation of various alleles of the AFLP markers and *inv* in two types of female offspring (*cenea* above, *hippocoon* below) of the male-informative Brood 59. The number of female offspring (*n*) maintaining the parental phase or showing recombination is also shown, with inferred crossover events marked in the figure. (b) A map of the *H* genome region based on co-segregation data from 133 offspring of four female-informative families.

inv from this cross (of 24 females total in this brood, see above). This includes those individuals that were affected by recombination between *H* and the *Pd* or *ACT* markers. Hence, *inv* was more closely linked to *H* than to either AFLP marker, and its most likely location is therefore between these two markers as shown above for *H* (figure 2a).

(c) Improving the map of the H-linked region

Four further broods segregating for three additional female forms, *lamborni*, *poultoni* and *trimeni* were of a size sufficient to be genotyped. All were female-informative with regard to the colour form (figure 2b, table 1). Again, the levels of polymorphisms were high and only the male parent of Brood 14 was homozygous for *ACT*. In total, the variation in the 12 parental individuals of families analysed throughout this study exhibited a remarkable 20 SNP sites and 7 indels of 1–21 bp in length in *ACT* and 27 SNP sites and 2 indels in *Pd*. We screened the male-inherited alleles of these markers in all offspring of Broods 41, 09 and 06 to obtain greater mapping resolution of the relative positions of the AFLP-derived markers and *inv*. Again, the three markers showed not more than two alleles each in all parental individuals and

these alleles co-segregated among offspring as expected from Mendelian principles. When combined with data from Brood 59, the number of offspring scored for segregation of the three markers sums to 133 individuals (5 of these individuals could only be scored for the AFLP markers and 2 for *inv* and 1 other marker) of which 16 were recombinants (table 1). MAPMAKER favoured the arrangement *Pd-inv-ACT* as the most likely (log likelihood = -67.06) with the closer of the two alternative arrangements showing a decrease in log likelihood of 3.78, i.e. alternative arrangements of these markers are over a thousand times less likely. MAPMAKER analysis gave a map distance between *Pd* and *inv* of 9.1 cM and a distance between *inv* and *ACT* of 4.8 cM, for a total map distance for this region of 13.9 cM (figure 2b).

4. DISCUSSION

We provide the first coarse-grain genetic map for *P. dardanus*, with the primary aim of tagging the hypothetical *H* locus, as a prerequisite for its molecular identification and studies of its function. Our approach was to first narrow down the position to a single linkage group presumably corresponding to one of 30 chromosomes in AFLP analysis, which makes use of the fact that in Lepidoptera there is no crossing over in the female sex. We mapped five further wing-pattern-determining factors (presumed alleles of *H*) to the same linkage group in this manner. This was followed by a detailed analysis mapping the *hippocoon* (*h*) and *cenea* (*H^c*)-determining factors relative to two AFLP markers using segregation in the males. We show that the interval delimited by the two AFLP markers corresponds to a 13.9 cM region. In another lepidopteran, *B. mori* (Yasukochi *et al.* 2006), this genetic distance would correspond to approximately 3 Mb, although wide margins need to be applied to this number given the great differences in recombination rates among species and within a single genome.

Critically, we found that this interval includes the gene for the transcription factor *inv*, which we consider a plausible candidate for *H*. Based on five crosses and well over 100 offspring genotyped, *inv* was placed between the two AFLP markers and hence *within* the interval, rather than on either side of the markers. If nothing else, the *inv* locus provides an additional marker in the critical *H*-linked region. Owing to its perfect co-segregation with the female forms even in individuals undergoing recombination with the AFLP markers, we can conclude that *H^c/h* is more closely linked to *inv* than the flanking AFLPs, and consequently *H^c/h* and *inv* together occupy a region that has to be smaller than the 13.9 cM interval. Applying this approach to further male-informative families like Brood 59 will, in future, further refine the map position of the *hippocoon/cenea* morphological difference. This cross requires the double-recessive *hippocoon* females, which one can unambiguously select from the population, with *hippocoon/cenea* heterozygous males. The latter are more difficult to obtain because all males are morphologically identical. Nevertheless, in any cross giving progeny that segregates *hippocoon* or *cenea*, every recombination between the *Pd* and *ACT* markers and the phenotype refines the position of *H^c* and *h*. If these males are selected from highly polymorphic populations such as race *P. d. polytrophus*, additional female forms are expected to segregate in this

type of cross and can be used to determine the relative position of the various *H* alleles. Based on the prior genetic work by Clarke & Sheppard (1959, 1960a,b, 1962) and the apparent lack of recombination between the hypothetical *H* alleles, other phenotype-determining factors are expected to be closely linked to the *hippocoon/cenea*-linked region identified here.

Our genetic mapping of *H* was hampered by the limited availability of specimens for use as parents in crosses and the small proportion of crosses in which the parental alleles of *H* could be determined from phenotypic segregation among the offspring, in addition to general difficulties of inducing oviposition and maintaining healthy breeding populations in captivity. This constrained the number of recombinants and the precision about the physical distance of *inv* and *H*. As a result, despite perfect co-segregation of *inv* and *H^c/h*, the possibility remains that these markers are far apart and they merely define a broader region that contains many more genes. Yet, *inv* is an interesting candidate for *H*. *Inv* is part of the *engrailed* (*en*) family of transcription factors and both *inv* and *en* may have roles in the developmental biology of wing pattern formation in other butterflies. Based on the expression studies using immunohistochemistry, *inv* or *en* (these functional studies using conserved portion of the proteins do not distinguish between both genes) affect the formation of the eyespot in nymphalid butterflies in late larval wing disc (Keys *et al.* 1999; Brunetti *et al.* 2001; Reed *et al.* 2007). *Inv* and *en* are located adjacently in the genomes of all hexapod species in which their genomic organization has been studied, including *Heliconius* and *B. mori* in the Lepidoptera (Peel *et al.* 2006; Pringle *et al.* 2007). The tandem arrangement of these two genes, the potential for complex control mechanisms that this arrangement provides and their redundancy of function during segmentation that could allow the diversification of function at other developmental stages (Peel *et al.* 2006) are all also suggestive of the supergene hypothesis. *En* is therefore also a candidate for *H*, along with any further genes that may be found to be linked to *inv* or the *H*-linked markers.

However, the molecular nature of *H* may also greatly impact our ability to produce a finer scale map of this region, to distinguish between linked candidate genes and to identify specific functional mutations. For example, the supergene hypothesis for *H* requires the prevention of recombination (such as would be achieved by a chromosomal inversion) in order to allow advantageous gene combinations (and therefore advantageous mimetic wing patterns) to be maintained. As genetic mapping approaches rely on information provided by recombination, this would greatly hamper all efforts to map *H* in any detail, even if greater numbers of offspring could be bred. Such a scenario might also predict that *H* occupies a relatively large genome region and the different phenotype-determining factors, while appearing closely linked, may be spread over a large physical distance. Our work provides an important first step towards the molecular characterization of the mimicry locus to address these possibilities. Already, preliminary data from ongoing studies on wild-caught populations (R. Clark & A. P. Vogler 2007, unpublished data) that use the 3' and 5' portions of the *inv* locus have illustrated the complete lack

of the linkage disequilibrium in this region, arguing against an inversion or other mechanisms that might suppress recombination in this region. This work has also suggested that a specific mutation in the *en/inv* region can be associated with a particular colour form. In a species with limited genetic tools, studies of variation in natural populations probably are the best way forward.

In conclusion, this work takes the initial steps towards the molecular characterization of the mimicry locus, *H*, by producing the first *H*-linked markers and candidate genes. The *H* locus has fascinated evolutionary biologists as a model for morphological change, where a relatively small region of the genome produces great diversity. A dual approach of studying the genetic organization of the locus and the action of regulatory genes present in this region will elucidate the mechanism of this regulatory switch. Interpreting such data in the light of similar work in other rather distantly related butterfly species, such as the mimetic patterns of *Heliconius* species or eyespot development in the Nymphalidae, will be intriguing and could establish the deep homologies, or divergences, of pattern elements between various lineages of Lepidoptera. Therefore, the mechanism of *H* action has wider implications for the understanding of colour pattern determination across the Lepidoptera and will contribute to our understanding of how evolutionary changes in the developmental mechanisms produce variation in morphological form.

This work was funded by grants from Human Frontier Science Program (HFSP; RG00323/1999-M), NERC (NE/D/008484/1), and a PhD fellowship of BBSRC (02A1G08223), with additional funding from the Natural History Museum and CESAR, a Special Research Centre of the Australian Research Council. We are grateful to Fred Nijhout (Duke University), Ali Cieslak (MNCM Madrid) and Dick Vane-Wright (NHM London) for their advice and discussions, and five anonymous reviewers and the Editor for their helpful comments on the manuscript.

REFERENCES

- Brakefield, P. M. & French, V. 1999 Butterfly wings: the evolution of development of colour patterns. *Bioessays* **21**, 391–401. (doi:10.1002/(SICI)1521-1878(199905)21:5<391::AID-BIES6>3.0.CO;2-Q)
- Brakefield, P. M., Gates, J., Keys, D. N., Kesbeke, F., Wijngaarden, P. J., Monteiro, A., French, V. & Carroll, S. B. 1996 Development, plasticity and evolution of butterfly eyespot patterns. *Nature* **384**, 236–242. (doi:10.1038/384236a0)
- Brunetti, C. R., Selegue, J. E., Monteiro, A., French, V., Brakefield, P. M. & Carroll, S. B. 2001 The generation and diversification of butterfly eyespot color patterns. *Curr. Biol.* **11**, 1578–1585. (doi:10.1016/S0960-9822(01)00502-4)
- Clarke, C. A. 1959 The breeding of *Papilio dardanus* Brown, in England. Technique and some results. *Entomologist* **92**, 89–95.
- Clarke, C. A. & Sheppard, P. M. 1956 Handpairing of butterflies. *Lepidopt. News* **10**, 47–53.
- Clarke, C. A. & Sheppard, P. M. 1959 The genetics of *Papilio dardanus* Brown. I. Race *cenea* from South Africa. *Genetics* **44**, 1347–1358.
- Clarke, C. A. & Sheppard, P. M. 1960a The genetics of *Papilio dardanus* Brown. II. Races *dardanus*, *polytrophus*, *meseres*, and *tibullus*. *Genetics* **45**, 439–456.
- Clarke, C. A. & Sheppard, P. M. 1960b The genetics of *Papilio dardanus* Brown. III. Race *antiorii* from Abyssinia and race *meriones* from Madagascar. *Genetics* **45**, 683–698.
- Clarke, C. A. & Sheppard, P. M. 1960c Super-genes and mimicry. *Heredity* **14**, 175–185.
- Clarke, C. A. & Sheppard, P. M. 1962 The genetics of *Papilio dardanus* Brown. IV. Data on race *ochracea*, race *flavicornis*, and further information in race *polytrophus* and *dardanus*. *Genetics* **47**, 910–920.
- Ford, E. B. 1936 The genetics of *Papilio dardanus* Brown (Lep.). *Trans. R. Entomol. Soc. Lond.* **85**, 435–466.
- Gahan, L. J., Gould, F. & Heckel, D. G. 2001 Identification of a gene associated with Bt resistance in *Heliothis virescens*. *Science* **293**, 857–860. (doi:10.1126/science.1060949)
- Heckel, D. G., Gahan, L. J., Liu, Y. B. & Tabashnik, B. E. 1999 Genetic mapping of resistance to *Bacillus thuringiensis* toxins in diamondback moth using biphasic linkage analysis. *Proc. Natl Acad. Sci. USA* **96**, 8373–8377. (doi:10.1073/pnas.96.15.8373)
- Jiggins, C. D., Mavarez, J., Beltrán, J., McMillan, W. O., Johnston, J. S. & Bermingham, E. 2005 A genetic linkage map of the mimetic butterfly, *Heliconius melpomene*. *Genetics* **171**, 557–570. (doi:10.1534/genetics.104.034686)
- 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)
- Kapan, D. D. et al. 2006 Localization of Mullerian mimicry genes on a dense linkage map of *Heliconius erato*. *Genetics* **173**, 735–757. (doi:10.1534/genetics.106.057166)
- Keys, D. N., Lweis, D. L., Selegue, J. E., Pearson, B. J., Goodrich, L. V., Johnson, R. L., Gates, J., Scott, M. P. & Carroll, S. B. 1999 Recruitment of a hedgehog regulatory circuit in butterfly eyespot evolution. *Science* **283**, 532–534. (doi:10.1126/science.283.5401.532)
- Kronforst, M. R. 2005 Primers for the amplification of nuclear introns in *Heliconius* butterflies. *Mol. Ecol. Notes* **5**, 158–162. (doi:10.1111/j.1471-8286.2004.00873.x)
- Kronforst, M. R., Young, L. G., Kapan, D. D., McNeely, C., O'Neill, R. J. & Gilbert, L. E. 2006 Linkage of butterfly mate preference and wing color preference cue at the genomic location of *wingless*. *Proc. Natl Acad. Sci. USA* **103**, 6575–6580. (doi:10.1073/pnas.0509685103)
- Lander, E. S., Green, P., Abrahamson, J., Barlow, A., Daly, J. J., Lincoln, S. E. & Newburg, L. 1987 MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**, 174–181. (doi:10.1016/0888-7543(87)90010-3)
- Michelson, R. W., Paran, I. & Kesseli, R. V. 1991 Identification of markers linked to disease resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions using segregating populations. *Proc. Natl Acad. Sci. USA* **88**, 9828–9832. (doi:10.1073/pnas.88.21.9828)
- Nijhout, H. F. 1991 *The development and evolution of butterfly wing patterns*. Washington, DC: Smithsonian Institution Press.
- Nijhout, H. F. 1994 Developmental perspectives on evolution of mimicry in butterflies. *Bioscience* **44**, 148–157. (doi:10.2307/1312251)
- Nijhout, H. F. 2003 Polymorphic mimicry in *Papilio dardanus*: mosaic dominance, big effects, and origins. *Evol. Dev.* **5**, 579–592. (doi:10.1046/j.1525-142X.2003.03063.x)
- Peel, A. D., Telford, M. J. & Akam, M. 2006 The evolution of hexapod *engrailed*-family genes: evidence for conservation and concerted evolution. *Proc. R. Soc. B* **273**, 1733–1742. (doi:10.1098/rspb.2006.3497)
- Pringle, E. G., Baxter, S. W., Webster, C. L., Papanicolaou, A., Lee, S. F. & Jiggins, C. D. 2007 Synteny and

- chromosome evolution in the Lepidoptera: evidence from mapping in *Heliconius melpomene*. *Genetics* **177**, 417–426. (doi:10.1534/genetics.107.073122)
- Reed, R. D., Chen, P. H. & Nijhout, H. F. 2007 Cryptic variation in butterfly eyespot development: the importance of sample size in gene expression studies. *Evol. Dev.* **9**, 2–9.
- Ridley, M. 2003 *Evolution*, 3rd edn. London, UK: Blackwell Science.
- Sekimura, T., Madzvamuse, A., Wathen, A. J. & Mani, P. K. 2000 A model for colour pattern formation in the butterfly wing of *Papilio dardanus*. *Proc. R. Soc. B* **267**, 851–859. (doi:10.1098/rspb.2000.1081)
- Tobler, A., Kapan, D., Flanagan, N. S., Gonzalez, C., Peterson, E., Jiggins, C. D., Johnstone, J. S., Heckel, D. G. & McMillan, W. O. 2005 First-generation linkage map of the warningly colored butterfly *Heliconius erato*. *Heredity* **94**, 408–417. (doi:10.1038/sj.hdy.6800619)
- Turner, J. R. G. 1984 Mimicry: the palatability spectrum and its consequences. In *The biology of butterflies* (eds R. I. Vane-Wright & P. R. Ackery), pp. 141–161. New York, NY: Academic Press.
- Vane-Wright, R. I., Raheem, D. C., Cieslak, A. & Vogler, A. P. 1999 Evolution of the mimetic African swallowtail butterfly *Papilio dardanus*: molecular data confirm relationships with *P. phorcas* and *P. constantius*. *Biol. J. Linn. Soc.* **66**, 215–229.
- Vos, P. *et al.* 1995 AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res.* **23**, 4407–4414. (doi:10.1093/nar/23.21.4407)
- Yasukochi, Y., Ashakumary, L. A., Baba, K., Yoshido, A. & Sahara, K. 2006 A second-generation integrated map of the silkworm reveals synteny and conserved gene order between lepidopteran insects. *Genetics* **173**, 1319–1328. (doi:10.1534/genetics.106.055541)