Supplementary Information

DNA sequence variation at *Yb*

We performed De novo assembly of short read data in Hm and related taxa. This revealed that in the region upstream of *cortex* near the SNP that was perfectly associated with the yellow forewing band (SNP603344) there was a stretch of DNA sequence that was present in all individuals lacking the yellow forewing band (n=5) but absent in all individuals lacking it (n=10, Extended Data Figure 2). However, given the smaller sample sizes this is less significant than SNP603344 itself. Similarly there were two 1bp deletions just downstream of SNP603344 that were present in all individuals with the yellow hind-wing bar but absent in all other individuals. We also found no strong associations between phenotype and the presence of particular unresolved repeat regions.

From the Long-range PCR targeted sequencing of cortex in Hm aglaope and Hm amaryllis we found that *cortex* had the highest mean squared allele frequency difference of the six amplified regions within *Yb* (0.0090 compared to 0.0040, 0.0041, 0.0032, 0.0018 and 0.0008 for the other regions).

The sequenced fosmid clones did not reveal any major structural rearrangements in the region. There were some apparent small translocations, which were mostly associated with TEs (Extended Data Figure 2). TE content was highly variable both between individuals and between alleles from the same individual. The *Heliconius*-specific Metulj elements[53] were particularly common across the region, especially in the large first exon of *cortex*. However these seemed to be largely consistent between individuals and alleles, with comparatively little variation in their presence/absence (Extended Data Figure 3).

In contrast, the second most common type of element, Helitron-like elements, were highly variable between individuals and alleles. The majority of TE variation did not obviously appear to associate with phenotype, for example we did not find any TEs that were present in both *Hm amaryliis* and *Hm rosina* (which have a common colour pattern, sharing the yellow hind-wing bar) but absent in the other morphs (Extended Data Figure 3). However one Helitron-like element was present between exons 3 and 4 in *Hm rosina* that was not present in *Hm melpomene* (which lacks the hind-wing bar). This element is particularly interesting because SNPs in this intron showed the strongest associations with the yellow hind-wing bar phenotype in the genotype-by-phenotype association analysis conducted across multiple races. Unfortunately the region was not covered by the *Hm aglaope* and *amaryllis* fosmid sequences.

Gene Expression Analyses

Tiling microarray analysis of Hm plesseni/malleti

The overall level of expression of *cortex* was similar in wings and eyes at day 3: Average log2 normalised expression of the *cortex* exons from the tiling array is 7.57 for wing tissue and 7.69 for eyes.

In addition to strong differences in expression observed at *cortex* (see main text) we observed weak but significant differences in expression at two previously identified[30] microRNAs within the *Yb* interval (Extended Data Table 1). hme-miR-193 appeared to be expressed more highly (1.01x) in *H. m. malleti* than *H. m. plesseni* in 1 day old pupal hindwings (FDR p=0.025). hme-miR-2788 was also upregulated (1.44x) in *H. m. malleti* as compared to *H. m.*

plesseni in the distal region of 1 day old pupal forewings (FDR p=0.040). These differences are small in both magnitude and significance level and were also only observed in these comparisons and not, for example, in the comparison of different wing regions, therefore we are unable to say whether or not they play a role in patterning, but it leaves open the possibility that they may interact with *cortex* in some way.

5' UTRs and alternative splice forms of cortex

A previous study of transcriptomic data suggested the existence of different splice variants of *cortex* (HM00025) in *Hm* involving both coding exons and alternative 5' UTR exons[12]. We further investigated this using RT-PCR and 5' RACE on RNA from *Hm* individuals. This revealed an extensive set of alternative 5' UTRs with the furthest being over 100kb upstream of the *cortex* coding exons (Extended Data Fig 2A). Using the mRNA sequence of these we were able to detect possible homologous regions upstream of the *He cortex* gene in the *HeCr* BAC sequence tilepath (Fig 2C), although no corresponding transcripts were found in available RNA-sequencing (RNA-seq) data for *He*.

The furthest upstream exon was present in both *Hm* individuals (*Hm aglaope* and *amaryllis)* used for 5' RACE and its presence was confirmed by RT-PCR in 17 additional individuals comprising *Hm aglaope* and *Hm amaryllis* of various developmental stages. Moreover exon 1, which contains the start codon, was found to be alternatively spliced with the first UTR exon, in that isoforms contained either exon 1 or exon U1 (Extended Data Fig 2A). The isoform lacking exon 1 is presumed to utilise the next start codon, which is in exon 3, resulting in a protein that is 365aa rather than 447aa.

We also detected multiple isoforms involving alternative splicing of other coding exons (Extended Data Fig 2 and 4). Isoforms lacking either exon 3 or exon 5 were found to be fairly common and present in multiple individuals. Splicing of exon 3 could lead to a new start codon in exon 2 that would preserve the frame of the rest of the protein and result in a protein of 335aa. Splicing of exon 5 results in a frame shift and premature stop codon in exon 6, and so a truncated protein of 203aa (assuming the exon 1 start codon is used).

RNAseq and qRT-PCR analysis of Hm amaryllis/aglaope

These races have a hybrid zone in Peru and differ at the *HmYb* and *HmN* loci controlling the presence of the yellow hind-wing bar and yellow forewing band respectively. We detected twelve, 95 and 208 genes as being differentially expressed between races at final instar larvae, day 2 and day 3 respectively using multiple analysis methods (Table S3). Only two genes were detected as being differentially expressed within the *HmYb* mapped region and both were only differentially expressed in the day 2 wings. HM00052 was upregulated in the yellow barred hind-wings of *Hm amaryllis* (p=0.018) while *cortex* was upregulated in the rayed hind-wings of *Hm aglaope* (p=0.035). This difference in expression of *cortex* is consistent with the upregulation that we detected in the phenotypically similar *Hm malleti*, and could be linked to the role of the *HmYb/N* locus in controlling the length of the hind-wing anterior red bar [7].

The *cortex* expression difference was confirmed by qRT-PCR using day 2 hind-wings from 10 *Hm aglaope* and 11 *Hm amaryllis*. On average expression was 1.6 times higher in *Hm aglaope* (SD=0.7, Wilcoxon rank sum test p=0.035) using primers in the coding exons 5 and 6 (Extended Data Fig 2B). However, using the same samples, we found 8.5x higher expression in *Hm aglaope* when assaying exons 1 and 2 (SD=0.54, Wilcoxon rank sum test p=1.08e-05, Extended Data Fig 2B). This suggests that *Hm aglaope* and *Hm amaryllis* have

differential expression of the isoforms that contain alternative exons 1 and U1, which contain different start codons.

In addition we found that the isoform lacking exon 3 was differentially expressed between these races. It was detected in all rayed *Hm aglaope* individuals (developing hind-wings from final instar larvae, day 1 and day 2 pupae, 24 individuals in total) but appeared to be completely absent from all yellow barred *Hm amaryllis* (same stages and sample sizes used, Extended Data Fig 2C, 4B).

Table S3. RNAseq results from 4 analysis methods

	Mean No. of genes detected as expressed		No. of genes detected as differentially expressed between races					
	RSEM	Stampy -Htseq	RSEM-DEseq	RSEM-BaySeq	Stampy -DEseq	Stampy-BaySeq	Consistent across all 4	within *Yb*
5th instar larvae	7776	7664	408	424	116	125	12	0
day 2 pupae	7903	7789	263	330	302	327	95	2
day 3 pupae	8445	8558	485	738	498	700	208	0

qRT-PCR analysis of Hm rosina/melpomene

These races have a hybrid zone in Panama and differ only in the presence of the yellow hind-wing bar, with *Hm rosina* having a bar and *Hm melpomene* lacking it. Comparisons of these races were conducted by RT-PCR and qRT-PCR of *cortex* transcripts only. Unlike the previous comparison no difference in expression was detected when using assays spanning either exons 5 and 6 or exons 1 and 2 (Day 2 pupal wings, n=25, Wilcoxon rank sum test p=0.8517 and p=0.205 respectively). Neither was there any clear race association with the isoform lacking exon 3, with a limited number of both *Hm melpomene* and *rosina* expressing this isoform (Extended Data Fig 4C). This could suggest that these differences that were detected in the previous comparisons are associated with the control of the shape of the anterior red bar on the hind-wing that is present in both *Hm aglaope* and *malleti* but not in either *Hm rosina* nor *melpomene*.

However, in this comparison we did detect one isoform that was differentially expressed between races. An isoform lacking exon 5 was detected in all *Hm rosina* individuals, which have a yellow hind-wing bar (developing hind-wings from final instar larvae, day 1 and day 2 pupae, 17 individuals in total) but was not present in any *Hm melpomene* individuals, which lack the bar (same stages and sample size). This isoform showed allele specific expression in an F2 cross between *Hm rosina* and *Hm melpomene*, demonstrating *cis*-regulatory control of the alternative splicing patterns. Using markers within the *HmYb* region we were able to identify individuals as heterozygous or homozygous for *HmYb* from the parental populations. Individuals both hetero- and homozygous for the *Hm rosina* allele expressed the isoform lacking exon 5, while those homozygous for the *Hm melpomene* allele did not (Extended Data Fig 4H). Using a diagnostic SNP within exon 4, we found that in heterozygous

individuals only the *Hm rosina* allele produced this isoform, while other isoforms contained alleles from both parents (Extended Data Fig 4I).

We also found the isoform lacking exon 5 to be expressed in *Hm cythera* (pool of 17, and 2 further individuals), which again possess the yellow hind-wing bar, and to be absent from a pool of 6 *Hm malleti* individuals, which lack the bar (Extended Data Fig 4G). However, we did not find a consistent difference in expression of this isoform between *Hm aglaope* and *amaryllis* (Extended Data Fig 4F), although the lower expression detected at exons 5 and 6 in *Hm amaryllis* (Extended Data Fig 2B) could indicate relatively higher prevalence of isoforms lacking exon 5 in this race. Therefore, isoforms lacking exon 5 may be important in formation of the yellow hind-wing bar.

Expression of *H. melpomene cortex* in *D. melanogaster* wings

D. melanogaster Cortex is known to generate an irregular microchaete phenotype when ectopically expressed in the posterior compartment of the adult fly wing[24]. We performed the same assay using *H. melpomene cortex* in order to test if this functionality was conserved. *Dm_Cortex* reproduced the irregular microchaete expression observed by Swan and Schüpbach (2007) (Extended Data Fig 6C), but *Hm_Cortex* did not generate any phenotype (Extended Data Fig 6B). Successful expression of *Hm_Cortex* was confirmed by IHC against an HA tag inserted at the N terminal of either protein (Extended Data Fig 6E).