A micro-RNA drives a 100-million-year adaptive evolution of melanic patterns in butterflies and moths

Shen Tian*, Tirtha Das Banerjee, Jocelyn Liang Qi Wee, Yehan Wang, Suriya Narayanan Murugesan, Antónia Monteiro*

Department of Biological Sciences National University of Singapore, 14 Science Drive 4, Singapore 117543, Singapore

*Co-corresponding authors

Abstract

A long-standing question in biology is how phenotypic variation is generated from genetic variation. In Lepidoptera (butterflies and moths), an enigmatic genomic region containing the protein-coding gene *cortex* is repeatedly used to generate intraspecific wing color pattern polymorphisms in multiple species. While emerging evidence suggest that *cortex* itself is not the effector gene of this locus, the identity of the effector remains unknown. Here, we investigated two deeply conserved miRNAs embedded in the *cortex* locus, and discovered that one of them, *mir-193*, promotes melanic wing color. This miRNA is likely derived from an intron of a gigantic long non-coding RNA spanning the entire locus, and it elicits its function by directly repressing multiple pigmentation genes including a well-known melanin pathway gene, *ebony*. The function of *mir-193* is conserved in a nymphalid and a pierid butterfly, belonging two families that diverged 90 Mya. We propose that *mir-193* is the effector gene of this hotspot locus underlying a 100-million-year evolution of intraspecific wing color pattern polymorphisms in Lepidoptera. Our results suggest that small non-coding RNAs can drive adaptive evolution in animals.

Introduction

Striking examples of intraspecific polymorphisms in wing color patterns can be found in Lepidoptera (butterflies and moths). Multiple studies have mapped the genomic loci underlying these polymorphisms. Surprisingly, only a handful of evolutionarily conserved loci, termed 'hotspot' wing color pattern loci, have been found to control most of these polymorphisms, despite the diversity of associated color pattern variants regulated by these loci (Wang et al., 2022, Joron et al., 2011, Ito et al., 2016, Nadeau et al., 2016, VanKuren et al., 2019, Hof et al., 2016, Reed et al., 2011, Gallant et al., 2014, Kunte et al., 2014, Iijima et al., 2018, Martin et al., 2012, Tunström et al., 2024).

One of the most intensively studied hotspot loci is a genomic region containing the protein coding gene *cortex*, associated with melanic (black and white; dark and bright) wing color pattern variations. The *cortex* locus has been repeatedly used to generate intraspecific polymorphisms across a 100-million-year evolutionary history of Lepidoptera. This includes the textbook example of industrial melanism of the peppered moth *Biston betularia*, the iconic mimetic radiation of *Heliconius* butterflies, and the leaf wing polymorphism of oakleaf butterflies *Kallima inachus*, among others (Wang et al., 2022, Joron et al., 2011, Nadeau et al., 2016, Hof et al., 2016, Edelman et al., 2019). However, the main effector gene within this locus is still largely unknown.

In previous studies centred on the *cortex* locus, the *cortex* gene itself was considered the sole effector gene of the entire locus. RNA-seq data suggested that *cortex* was the only, or one of a few genes within the locus that was differentially expressed, or differentially spliced across different morphs or color pattern variants (Nadeau et al., 2016, Hof et al., 2016, Livraghi et al., 2021, Saenko et al., 2019). However, the spatial expression of *cortex* on the developing wings did not map to the actual wing color patterns (Livraghi et al., 2021, Saenko et al., 2019, Wang et al., 2022). Functional evidence from CRISPR-Cas9 mediated knock out experiments suggested that *cortex* might regulate wing melanisation, but these studies suffered from low success rates since *cortex* is essential for life (Livraghi et al., 2021,

Hanly et al., 2022, Wang et al., 2022), indicating that hidden genomic features disrupted next to *cortex*, but not *cortex* itself, could be causative.

While previous investigations exclusively focused on protein-coding genes, non-coding RNAs were ignored, and the *cortex* locus contains two such annotated elements, two micro-RNAs (miRNAs). MiRNAs are well-known post-transcriptional gene regulators, but largely understudied in the genetics of morphological diversification. They are small (20-22nt) non-coding RNAs, mostly transcribed from the intronic regions of messenger RNAs (mRNAs) or from long noncoding RNAs (lncRNAs). Upon transcription they form a local hairpin-like structure, termed a primary miRNAs (pri-miRNAs). These pri-miRNAs are processed by Drosha to generate precursor miRNAs (pre-miRNAs), and then processed by Dicer, to generate double stranded mature miRNAs. The guide strand of a mature miRNA, once loaded onto the RNA-induced silencing complex (RISC), recognizes its mRNA targets via sequence complementarity, and mediates its decay or translational repression (Fig. 1A) (Ha and Kim, 2014).

In this study, we tested if the two previously ignored miRNAs present in the *cortex* locus are the effector genes of the locus. We used CRISPR-Cas9 to investigate their roles in color pattern development in two model butterfly species, *Bicyclus anynana* and *Pieris canidia*. We then further investigated the potential primary transcript that gives rise to both miRNAs, as well as the direct targets of one of the miRNAs, in a model system *B. anynana*.

Results

One of the two miRNAs embedded in the *cortex* locus, *mir-193*, but not *mir-2788*, promotes melanic wing color

Taking advantage of a complete set of miRNAs annotated in *B. anynana* butterflies (Tian and Monteiro, 2022), we screened the whole *cortex* locus and only found two deeply conserved miRNAs embedded in it, *mir-193* and *mir-2788* (Fig. 1B).

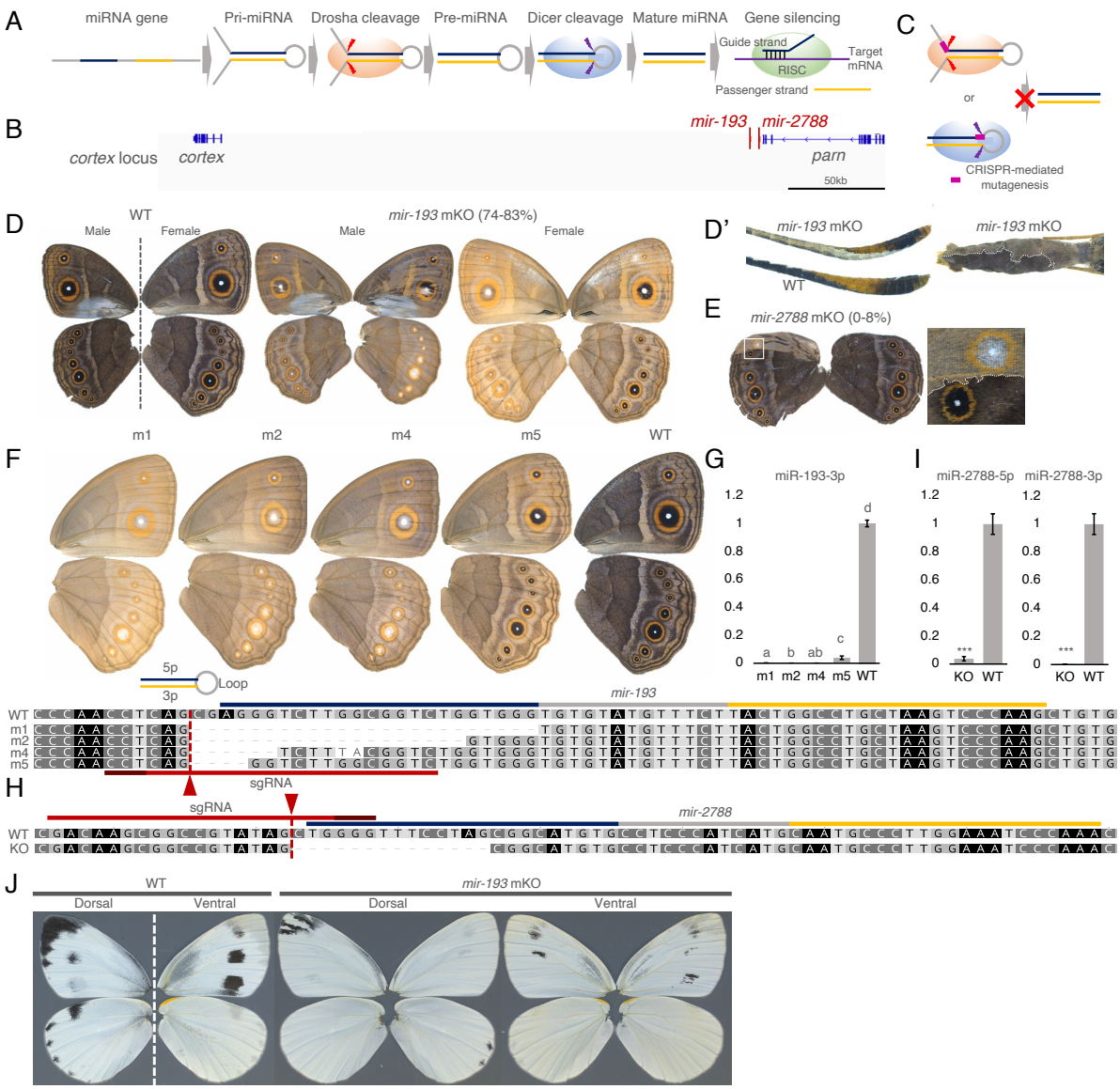

Figure 1. *Mir-193*, but not *mir-2788*, promotes melanic wing color. (A) Biogenesis and silencing machinery of miRNAs. **(B)** Two deeply conserved miRNAs embedded in the *cortex* locus. **(C)** CRISPR-Cas9 mediated mutagenesis strategy to block the biogenesis of mature miRNAs. Mosaic knock-out (mKO) phenotypes of **(D-D')** *mir-193* and **(E)** mir-2788 in *B. anynana*. **(F)** homozygous *mir-193* mutant lines of *B. anynana* and their corresponding genotypes **(G)** Expression levels of the guide strand of *mir-193*, miR-193-3p, across the mutant lines and wild-type (WT), determined by qPCR (N=3). Expression levels with the same letter are not significantly different from each other, as determined by Tukey's test. **(H)** Genotype of the homozygous *mir-2788* mutant line and **(I)** expression levels of the two mature strands of *mir-2788*, miR-2788-5p and miR-2788-3p, across the mutant line and WT, determined by qPCR (N=4). Student's t-test ns, not significant; *, p<0.05, **, p<0.01, ***, p<0.001. **(J)** mKO phenotypes of *mir-193* in *P. canidia*.

Both miRNAs are in the intergenic region between *cortex* and the adjacent protein-coding gene *poly(A)-specific ribonuclease* (*parn*) (Fig. 1B). Phylogenetic studies and the current version of miRBase suggest that *mir-193* is a deeply conserved miRNA across the animal kingdom, while *mir-2788* is conserved in

Arthropoda, (Ma et al., 2021, Quah et al., 2015, Griffiths-Jones et al., 2006). From sequence alignments, we confirmed that both miRNAs are well conserved in all the major model species of Lepidoptera (Fig. S1).

To elucidate the function of both miRNAs we performed embryonic CRISPR-Cas9 mediated knock-out experiments in *B. anynana*. Single guide RNAs (sgRNAs) were designed to disrupt the essential Drosha or Dicer processing sites in the miRNA biogenesis pathway to effectively block the biogenesis of mature miRNAs (Fig. 1C)(Chang et al., 2016). Two sgRNAs were designed for each miRNA and each sgRNA was injected separately, all sgRNAs showed efficient cutting of the genomic DNA (Fig. S2, Tables S1-2).

Similar CRISPR phenotypes were observed for the two miRNA loci, but at different frequencies. For *mir-193*, 74-83% of the mosaic knock-out individuals showed reduced melanin levels across all the body parts covered with scale cells, such as wings, antenna, legs, and abdomen (Fig. 1D-D', Table S2). Reduced melanin levels were not observed from body parts with bare black cuticles without scale cells such as antennal tips (Fig. 1D'). Variations in wing color were observed between individuals and sexes (females are generally lighter in color than males). Many of the mosaic mutants could not fly and had wing hinge weakness. For *mir-2788*, similar color phenotypes were observed, but only 0-8% of the adults showed wing color changes, with no flight defects observed (Fig. 1E, Table S2). These phenotypes were also observed when the injected embryos were reared at lower temperatures (Supplementary notes, Table S2).

To document the mutant alleles that each individual was carrying, we sequenced some of the individuals with variant phenotypes. For *mir-193*, we genotyped five male mutant hindwings with homogeneous color changes, but with variable melanin levels in the eyespot black disk region (Fig. S3A). We used Illumina short-read amplicon sequencing to sequence a 362bp region around the sgRNA cut sites (Table S3). We discovered that the mutation rate approached 100% in all the samples, and each had 1-3 major mutant alleles with indels of 3 to 123 bp in length (Fig. S3B-C). For *mir-2788*, we suspected rare long

indels around the sgRNA cutting site that disrupted adjacent genomic causative regions. To capture these putative long indels, we performed Pacbio long-read sequencing of a 3820 bp region around the cut site from both hindwings of a mutant male, where one wing showed reduced melanin levels and the other wing had no visible phenotypic changes (Fig. S3D-E, Table S3). Long deletions were detected in both wings only by one of the two analyses tools, but were not observed when the sequence alignments were inspected manually. However, short indels were observed at a near 100% rate around the cutting site in both samples (Fig. S3F). As a result, all the DNA extracted from the crispant wings from both *mir-193* and *mir-2788* knock-out experiments was near completely mutated. For *mir-193*, the various mutant phenotypes are likely the result of individual and combinatorial effects of different mutant alleles around the miRNA locus. For *mir-2788*, complete mutations with short indels around the miRNA locus were observed regardless of the presence of wing phenotypes, suggesting that the rare color phenotypes are not caused by the disruption of the small miRNA locus itself, but of a flanking genomic locus, potentially *mir-193*.

To further confirm the genotype-phenotype association, we generated homozygous mutant lines with short deletions for both miRNAs. (Fig. S4, see Materials and Methods for more details). For *mir-193*, we generated four mutant lines with a 4bp (m5), 6bp (m4), 19bp (m2), and 24bp (m1) deletion around the *mir-193* 5' Drosha processing site (Fig. 1F). A phenotype series was observed where m1, m2, and m4, with 6-24bp deletions, showed almost equivalent light brown wing color, with "black disk" eyespot wing color patterns turning white (Fig. 1F). Homozygotes of these lines could not fly due to severe wing hinge weakness. m5, with a shorter 4bp deletion, showed milder phenotypes, with slightly darker wing colors and black "black disks". m5 homozygotes were, for the most part, able to fly. Heterozygotes of all the four mutant lines were visibly WT-like, in both color and behavior, indicating the recessive nature of these alleles. Quantification of the level of the guide strand of *mir-193*, miR-193-3p, using qPCR across mutants and WT pupal wings indicated that increasing levels of miR-193-3p correlated with darker wing colors across a phenotypic gradient, suggesting a dose-dependent effect of *mir-193* in wing coloration (Fig. 1G). For *mir-2788*, one mutant line was generated with a 14bp deletion around the 5' Drosha processing site (Fig. 1H). No visible phenotypic changes were observed in both mutant

heterozygotes and homozygotes, although depletion of both mature strands in the mutant line was confirmed using qPCR (Fig. 1I, Table S4). To sum up, we confirmed that *mir-193*, but not *mir-2788*, promotes melanic wing color in *B. anynana*,

To examine whether the function of *mir-193* is conserved across butterflies, we performed CRISPR knock out experiments on *mir-193* in *P. canidia* butterflies, a species that diverged from *B. anynana* around 90 Mya (Kawahara et al., 2023). Similar phenotypes were observed in this species, where the black/grey wing color patterns became white (Fig. 1J), confirming that the function of *mir-193* is conserved across Pieridae and Nymphalidae.

Mir-193 and _mir-2788_ are likely processed from a gigantic long non-coding RNA

Identifying the primary transcripts, or host genes, of miRNAs is essential to elucidate their transcriptional control and tissue/cell-specific expression patterns. Since miRNAs in proximity are usually co-transcribed together as a cluster (Altuvia et al., 2005), we hypothesized that *mir-193* and *mir-2788*, which are only 5.2kb apart, are processed from the same primary transcript. Due to the high turnover rate of Drosha processing, pri-miRNAs are usually expressed at low levels and are not easily captured in RNA-seq data (Chang et al., 2015). In fact, none of the annotated lepidopteran genomes showed any annotated transcripts overlapping the two miRNAs in the *cortex* locus. Blocking Drosha processing, however, is a feasible way to trigger the accumulation of these primary transcripts (Chang et al., 2015).

Taking advantage of the miRNA mutant lines that disrupted the Drosha processing sites, we performed RNA-seq to profile whole transcriptomes from sib-paired WT and mutant homozygotes. For *mir-193*, wing tissues from sib-paired female WT and the m4 homozygotes were sequenced from 60% wanderer stage (late 5th instar larval stage), Day 1 (early) pupal stage, and Day 6 (late) pupal stage. For *mir-2788*, wing tissues from sib-paired female WT and the mutant homozygotes were sequenced from Day1 pupal stage only. We then inspected all aligned transcripts, from all the RNA-seq data, to identify those spanning the miRNA region. A gigantic transcript spanning a 370kb chromosomal region was observed

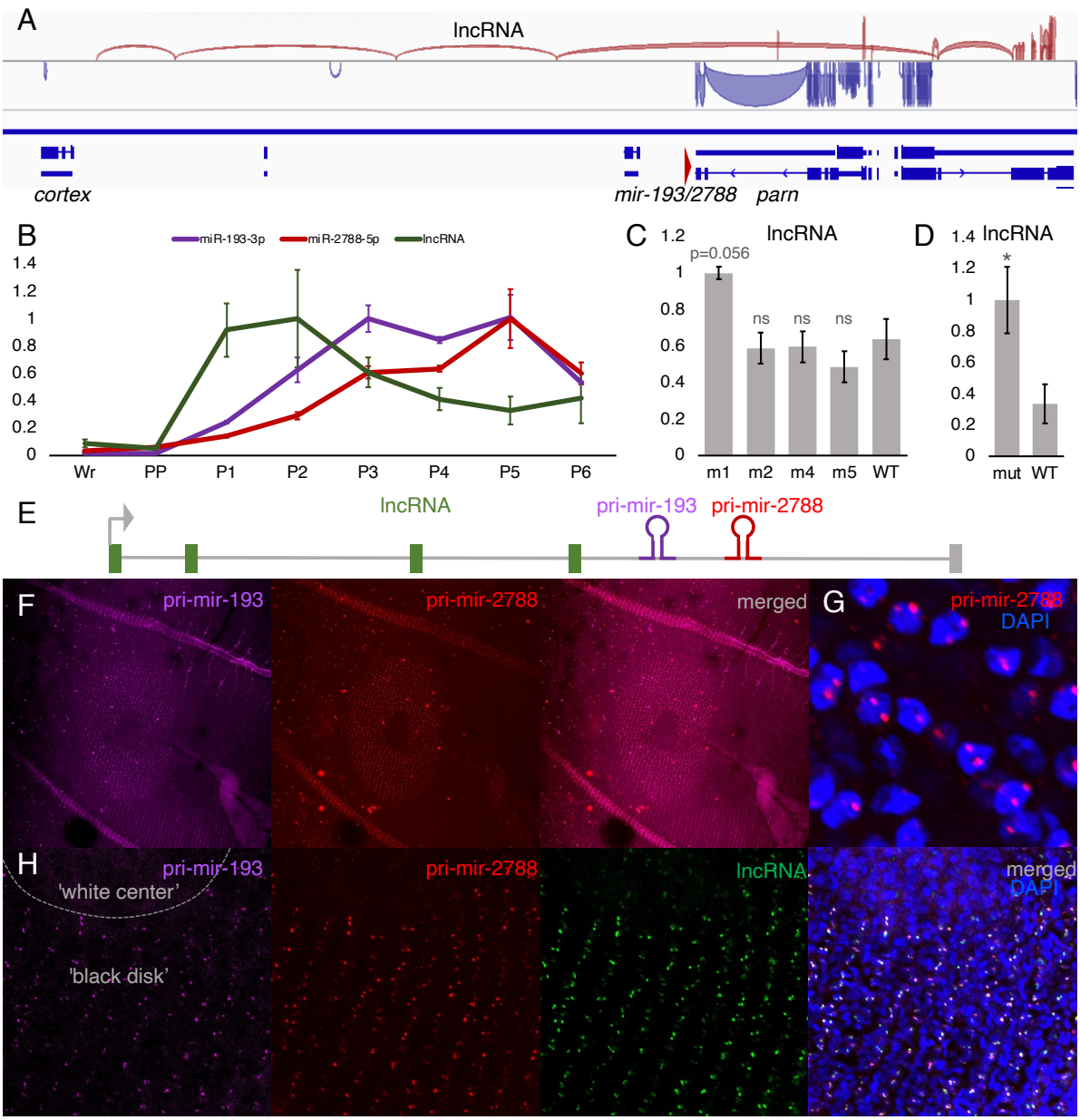

Figure 2. *mir-193* and *mir-2788* are likely derived from a gigantic lncRNA. (A) A gigantic lncRNA spanning a 370kb genomic region was discovered in the RNA-seq data of some *mir-193* mutant pupal wings, of which an intronic region spans the two miRNAs, with the same transcription direction. **(B)** Time series expression profiles of the mature guide strands of the miRNAs, *miR-193-3p* and *miR-2788-5p*, as well as the lncRNA, across development. Wr, wanderer stage (late 5th larval stage); PP, pre-pupal stage; P1-P6, Day1-6 pupal stages. Expression levels of the lncRNA **(C)** across the four *mir-193* mutant lines and WT (N=3), and **(D)** across the *mir-2788* mutant line and WT (N=4). Student's t-test: ns, not significant; *, $p<0.05$, **, $p<0.01$, ***, $p<0.001$. **(E)** Probe design (colored region) for the HCR staining experiments, for the lncRNA, exonic regions were used; for miRNAs, a 1kb region flanking each of the miRNA hairpin region was used to probe pri-miRNA expression. **(F)** Co-expression of the two miRNAs in the 'black disk' region of *B. anynana* eyespots. **(G)** Nuclear signals of *pri-mir-2788*. **(H)** Co-expression of the two miRNAs and the lncRNA in the 'black disk' regions.

in two *mir-193* Day1 and one *mir-193* Day6 mutant pupal wings, but not in *mir-193* mutant larval wings, WT wings, or *mir-2788* mutant wings (Fig. 2A). No open-reading frame was found in this newly discovered transcript, rendering it as a long non-coding RNA (lncRNA). There was no clear sequence homology for the lncRNA across Lepidoptera, except for the core promoter region, just upstream of the first exon of the lncRNA, suggesting its transcription is conserved across Lepidoptera.

One intron of the lncRNA overlaps the two miRNAs and this lncRNA is the sole transcript overlapping the two miRNAs with the same transcription orientation (Fig. 2A). Thus, we hypothesized that this lncRNA is the primary transcript of *mir-193* and *mir-2788*, and that it will be highly expressed in the miRNA mutants during the pupal stage, just as its derived miRNAs.

To test this hypothesis, we first generated a time series expression profile for the guide strands of the two miRNAs, as well as for the lncRNA. As expected, the three non-coding RNAs exhibited very similar expression profiles – negligible expression during larval or pre-pupal stages, and high expression levels throughout the pupal stage (Fig. 2B). Interestingly, the lncRNA exhibited an expression peak earlier in development than the two mature miRNAs, which aligns with the general pattern of miRNA processing in *Drosophila*: a fast turnover rate of the primary transcript produces long-lasting mature miRNA products (Zhou et al., 2018).

As miRNA primary transcripts are supposed to accumulate upon the disruption of the Drosha processing sites, we quantified the level of the lncRNA across all *mir-193* and *mir-2788* mutant lines. Across the four mir-193 mutant lines, the lncRNA is marginally ($p=0.056$) overexpressed in m1 mutants, compared with WT, but not in the other mutants with shorter deletions (Fig. 2C). In the *mir-2788* mutants, the lncRNA is significantly overexpressed ($p=0.034$) compared with WT (Fig. 2D). This provided moderate evidence that the two miRNAs were derived from the lncRNA.

Finally, we used hybridization chain reaction (HCR) to test whether the expression of the three non-coding RNAs is co-localized to the same cells and maps to the same dark patterns on pupal wings. To

localize pri-miRNAs, a 1kb genomic region flanking each of the miRNA hairpin regions was used to design HCR probes (Aboobaker et al., 2005) (Fig. 2E). We captured high expression signals of both *pri-mir-193* and *pri-mir-2788* in the black disk region of the eyespot pattern of *B. anynana* (Fig. 2F). The expression signal for each pri-miRNA appeared as two nuclear dots, corresponding to the two chromosomal transcription sites, as previously described (Aboobaker et al., 2005). Similarly, the lncRNA was also highly expressed in the black disk region of the eyespots, overlapping the nuclear signals of pri-mir-193 and pri-mir-2788. This provided strong evidence that both miRNAs were co-transcribed, or derived, from the lncRNA.

Transcription of intronic miRNAs can involve both host gene-dependent, and independent transcriptional regulations (Monteys et al., 2010). To test if *mir-193* has its own transcription start and end sites just up- and downstream of the miRNA locus, we did 5'- and 3' RACE directly extending 3' and 5' ends from the miRNA precursor. This recovered an alternative transcription start 200 bp 5' of the *mir-193* precursor and a polyadenylation signal 250 bp 3' of precursor. However, the RACE signals could only be captured in some, but not all individuals included in the RACE experiment. This suggest that *mir-193* might also have its own transcriptional control, independent of the regulation of the parent lncRNA locus.

Mir-193* directly targets multiple pigmentation genes including a melanin repressor *ebony

MiRNAs elicit their regulatory effects by repressing target mRNAs. To discover both direct and indirect targets, we examined the transcriptomes generated from the m4 *mir-193* mutants and sib-paired WT. A total number of 4 (larva), 11 (Day1 pupa), and 218 (Day6 pupa) genes were found to be differentially expressed (DEGs, padj<0.01) across mutant and WT wings, suggesting that *mir-193* triggers large transcriptomic changes mostly during (late) pupal development, when it is highly expressed. Except for an uncharacterized protein (NCBI gene ID: LOC112054702) which is ~680kb away from *cortex*, none of the genes within the *cortex* locus (2Mb genomic region flanking *cortex*) appeared as DEGs, suggesting the *trans*-, rather than *cis*-acting nature of the miRNA. Multiple genes previously associated with ommochrome and melanin pigmentation pathways in butterfly wings appeared as DEGs in Day6

pupal wings (Zhang et al., 2017). Since direct targets of *mi-193* are expected to be overexpressed in the miRNA mutants, all genes that were up-regulated in the mutants were pooled and searched for complementary binding sites to *miR-193-3p*, the guide strand of *mir-193*. In total, 49 out of 118 genes highly expressed in the mutant were predicted to be direct targets of *miR-193-3p* (Fig 3A). One of those targets is *ebony* (*e*), a well-known insect melanin pathway gene that catalyzes the conversion of dopamine to NBAD – a light yellow pigment. Two *miR-193-3p* binding sites were found in *e*, one within the protein coding sequence (CDS), and the other within the 3' untranslated region (3'UTR) (Fig 3B). We propose that when *mir-193* was disrupted in the black/dark wing regions in the m4 mutant, *e* was de-repressed in those regions, and started converting dopamine to NBAD, producing a lighter color, and shunting dopamine away from melanin pigment production (Zhang et al., 2017, Matsuoka and Monteiro, 2018).

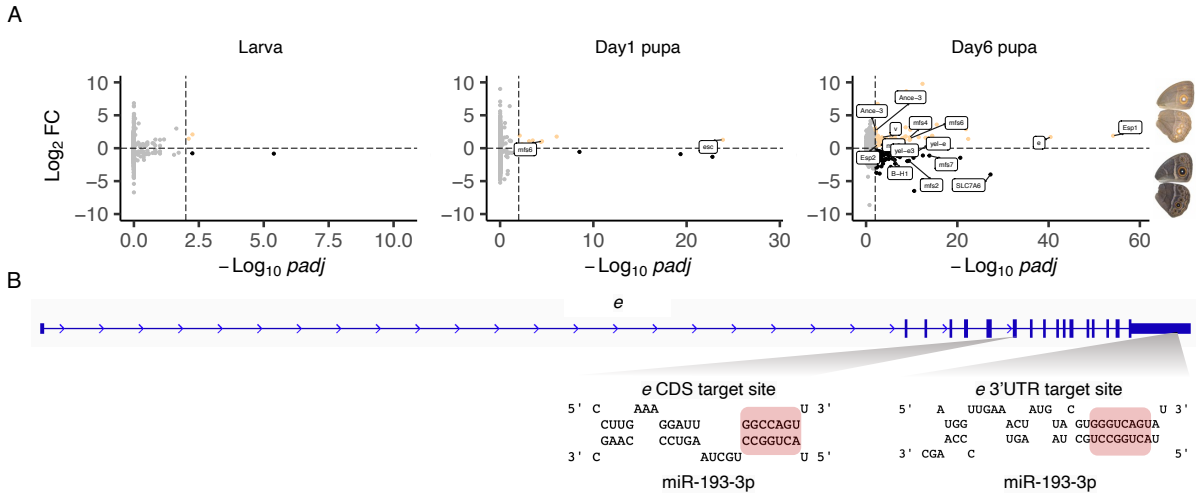

Fig 3. *mir-193* directly/indirectly regulates multiple pigmentation genes during the late pupal stage. (A) Differentially expressed genes (padj < 0.01) across m4 homozygous and WT wing tissues at three developmental stages. Genes previously associated with butterfly wing color pigmentation pathways were highlighted. **(B)** *ebony* (*e*) is predicted to be one of the direct targets of *miR-193-3p*, the guide strand of *mir-193*, with two predicted binding sites. Full sequence complementarity between the miRNA seed region (nucleotides 2-8 from the 5' terminus of the guide strand) and the mRNA target sites are highlighted in red.

4.4. Discussion

Genotype-phenotype association analysis is usually used to assign variation in a genomic region to variation in a phenotypic trait. In many cases, the causative mutations and/or chromosomal rearrangements are frequently found in intergenic regions, and it is usually assumed that these regions harbour cis-regulatory elements (CREs) that regulate the expression of flanking genes. When the causative locus is a poorly annotated non-coding element that acts in *trans*, as in the miRNA or the host lncRNA identified in this study, a direct link between local genetic variations and flanking protein-coding genes can lead to misleading conclusions. This happened with the *cortex* locus where a large number of intraspecific wing color pattern polymorphisms in lepidopterans, in most cases involving black and white, or dark and bright color patterns, have been mapped to this hotspot locus and been attributed to *cortex* regulation (Wang et al., 2022, Hanly et al., 2022, Livraghi et al., 2021, Nadeau et al., 2016, Hof et al., 2016).

In this study, however, we demonstrate that one of the two miRNAs, *mir-193*, embedded in the *cortex* locus, is likely the major effector gene of the entire locus. This is supported by 1) the dramatic *mir-193* CRISPR phenotypes produced at high frequency in two species; 2) the observation that these drastic color changes, induced by small deletions on the miRNAs locus, did not trigger any expression changes in adjacent protein-coding genes previously proposed to be the melanic color regulators, including *cortex*; and 3) by *mir-193* having multiple direct targets previously associated with coloration including *ebony*, a well-known insect melanin pathway gene.

We show that the two miRNAs are embedded in one of the introns of a gigantic lncRNA, are co-expressed in the same cells, and co-localize together in the same nuclear regions of single cells. Therefore, genetic variations across the entire locus, potentially carried by CREs, could potentially control the spatiotemporal expression of this lncRNA, and its derived miRNAs, to control polymorphic wing color patterns via the color regulator *mir-193*. Also, understanding the evolution of the multiple

mir-193 direct targets across species is essential to elucidate how *mir-193*, became a central color regulator in Lepidoptera.

Moreover, our results suggest that the function of *mir-193* in promoting melanic wing coloration is conserved across at least 90-million years of evolutionary history, as *B. anynana,* a nymphalid, and *P. canidia,* a pierid, diverged around that time. Therefore, we propose that *mir-193* is the major effector gene of the *cortex* locus that drives a 100-million-year adaptive evolution of melanic wing color variations in Lepidoptera, including moth species such as *B. betularia* and *B. mori* (Ito et al., 2016, Hof et al., 2016). Future studies could investigate both the expression and regulation of *mir-193* in all Lapidoptera species where wing color polymorphisms were previously mapped to this locus, to test this hypothesis.

In conclusion, our study identified a small non-coding RNA that potentially mediated the repeated evolution of intraspecific lepidopteran wing color pattern polymorphisms. It was recently discovered that small non-coding RNAs are key regulators of flower color polymorphism and also speciation loci (Liang et al., 2023). This study adds another piece of evidence that small non-coding RNAs also underly adaptive evolution in animals. Since there is a burst of miRNA innovation at the base of Lepidoptera (Quah et al., 2015, Ma et al., 2021), these novel miRNAs may serve as evolutionary raw materials to create a splendid gamut of morphological diversity within this order, one of the most species-rich on earth. Our current study, and future investigations on non-coding RNAs can also shed light on the long-standing hypothesis that it is the complexity of swiftly evolving non-coding components of the genome (cis-acting regulatory DNA elements and trans-acting non-coding RNAs), rather than the relatively static proteome, that drives organismal complexity (Mattick, 2001, Taft et al., 2007).

Materials & Methods

Insect husbandry

Lab populations of *B. anynana* and *P. canidia* butterflies were reared 27ºC, with a 12:12 day: night cycle and 60% relative humidity. *B. anynana* larvae were fed young corn leaves and adults were fed mashed banana. *P. canidia* larvae were fed potted cabbage leaves and adults were fed artificial nectar.

CRISPR-Cas9 genome editing

Embryonic CRISPR-Cas9 knock out experiments were performed for the two miRNAs within the *cortex* locus, *mir-193* and *mir-2788*, in *B. anynana*, following the established protocol (Banerjee and Monteiro, 2018) (with a video protocol). For each miRNA, two sgRNAs were designed to target the Drosha or Dicer processing sites of the miRNAs, which is effective in blocking mature miRNA biogenesis. Specificity of sgRNA target sequences was checked by blasting against the *B. anynana* v1.2 genome in Lepbase. Only those with a unique hit were selected. The template single guide DNAs (sgDNAs) were produced by PCR, using NEB Q5 High-Fidelity DNA Polymerase. sgDNAs were transcribed into sgRNAs using NEB T7 RNA Polymerase. sgRNAs were purified by ethanol precipitation. Size and integrity of sgDNAs and sgRNAs were checked using gel electrophoresis.

For microinjection of sgRNAs, corn leaves were placed in WT *B. anynana* adult cages around 2-3pm. Leaves were left in the cages for 1h and eggs were collected immediately and placed onto 1mm-wide double-sided strips of sticky tape in petri dishes. We used a final concentration of 500ng/uL IDT Alt-R S.p. Cas9 Nuclease V3 (with 1 x Cas9 buffer), and 250ng/uL sgRNA in the injection mixture. Different sgRNAs for the same miRNA were not mixed but injected independently in different batches. The injection mixture was incubated at 37 ºC for 10 min prior to injection. 0.5 uL food dye was added per 10 uL mixture to facilitate the visualization of the injected solutions. Microinjection was performed using glass capillary needles. Wet cotton was placed inside the petri dishes with injected eggs, and the eggs were stored at 27ºC for *mir-193*, and at either 27ºC and 17ºC for *mir-2788*. Newly hatched larvae were feed cut corn leaves in small containers for 1 week, then shifted to corn plants in cages to complete

development. Newly emerged butterflies were frozen at -20 ^{o}C ~6h after adult emergence. Phenotypic changes were inspected manually.

In-vivo T7 endonuclease assays were performed to confirm the cutting efficiency for each sgRNA, as described below. Genomic DNA was extracted from ~10 injected eggs 4 days after sgRNA injection, using E.Z.N.A Tissue DNA Kit. For each miRNA, one pair of CRISPR genotyping primers was designed to amplify a 300~600bp region spanning the target sites of sgRNAs, using PCRBIO Taq Mix Red. PCR products were purified using ThermoFisher GeneJET PCR Purification Kit. 200ng of PCR products were denatured and re-hybridized with NEB Buffer 2 in a total volume of 20uL. The product was treated with 1uL NEB T7 endonuclease incubated at 37°C for 15min. Another tube using wild-type genomic DNA as template and processed the same way was used as a control. Both treatment and control tubes were run on a 2% agarose gel. All sgRNAs showed efficient cutting of the target sites.

Similar embryonic CRISPR-Cas9 knock out experiments were performed for *mir-193* in *P. canidia*. For microinjection, *P. canidia* eggs were gathered using a small plastic container lined with a piece of parafilm and topped with a fresh cabbage leaf. This setup encouraged adult butterflies to deposit most of their eggs onto the parafilm, enhancing hatchlings' survival rates by preventing egg dehydration. The container was left inside the cage for approximately 4 h to optimise the number of eggs collected. Once a sufficient number of eggs were obtained, both the leaf and the parafilm were transferred to a Petri dish for injection. Eggs were injected with CRISPR/Cas9 protein-crRNA-tracrRNA complexes. *Synthetic* crRNA guides targeting *mir-193* were ordered from Integrated DNA Technologies (IDT). 1µl of crRNA was incubated with 1µl of tracrRNA and 8µl of Nuclease-Free Duplex Buffer at 95°C for 5 mins and then left to cool down at room temperature. 1.2µl of 10X Cas9 buffer and Cas9 protein were then added to the mixture to make a final concentration of ~800ng/µl. After injection, moistened cotton balls were then placed into each Petri dish to maintain high humidity levels and to prevent egg desiccation. Hatchlings were promptly transferred to potted cabbage plants and raised until they reached adulthood.

Genotyping mosaic mutants via NGS

F0 mosaic mutants generated from CRISPR knock out experiments were genotyped via next generation sequencing (NGS). For *mir-193*, genomic DNA was extracted from five mosaic mutant hindwings with different melanin levels within the eyespot area, from four male mosaic mutants. For *mir-193*, genomic DNA was extracted from both hindwings of a male mosaic mutant with paired phenotypes, one wing showed reduced melanin levels, while the other wing appeared WT-like. DNA extraction was performed using E.Z.N.A Tissue DNA Kit. For *mir-193*, a 362bp region around the sgRNA cutting site was amplified from genomic DNA with NGS genotyping primers, while for *mir-2788*, a 3820bp region was amplified. PCR was performed using NEB Q5 High-Fidelity DNA Polymerase and PCR products were purified using ThermoFisher GeneJET PCR Purification Kit. PCR products were checked using gel electrophoresis and NanoDrop, and were sent for NGS.

For *mir-193*, Illumina NovaSeq short-read amplicon sequencing was performed for the five mutant wing samples to generate 1M 250bp paired-end reads for five samples in total. For *mir-2788*, Pacbio Sequel II long-read amplicon sequencing was performed for the two wing samples to generate 7K-10K 4kb HiFi reads per sample. Quality check, library construction, and sequencing were all carried out by Azenta, Singapore. For Illumina short reads, adaptors were trimmed using Cutadapt. Clean reads were aligned to the reference sequence of the amplified region from the NCBI *B. anynana* v2 genome. Mutant alleles generated by CRISPR were called using CRISPResso2. Pacbio long reads were aligned to the reference sequence of the amplified region from the NCBI *B. anynana* v2 genome using NGMLR and alignments were visualized in Geneious prime. Structural variants and mutant alleles were called using Cas-Analyzer and Sniffles.

Generation of miRNA mutant lines

Homozygous mutant lines were generated for both *mir-2788* and *mir-193* in *B. anynana*. In general, F0 mosaic female mutants with wing phenotypes from CRISPR knock out experiments were crossed with WT males. Some of these F0 mosaic females might carry one or several mutant alleles in their germline

to pass to their offspring. F1 offspring were screened via hemolymph extraction and Sanger sequencing, as described below. F1 heterozygotes carrying mutant alleles were crossed with either other F1 heterozygotes carrying the same mutant allele, if any, or with WT individuals. Subsequent generations were screened in the same way, and crossed until mutant homozygous lines were generated. The *mir-2788* mutant line was maintained in a completely homozygous status. For *mir-193*, the four mutant lines were maintained in a way that in each generation, mutant homozygous females (with visible wing phenotypes) were crossed with heterozygous males (WT-like), since *mir-193* mutant alleles were recessive and homozygous mutants have various degrees of wing hinge weakness. WT-like heterozygous males were needed to perform proper courtship displays for successful mating.

Genotyping mutant lines via hemolymph extraction and Sanger sequencing

For hemolymph extraction, around 10uL hemolymph was extracted from 5th instar larvae and suspended in 200uL Saline-Sodium Citrate buffer. Hemocytes were collected by centrifugation, then resuspended and incubated in 20uL digestion buffer (1.1mL 1M Tris-HCl of pH6.8 in 50mL deionized water) with 2uL proteinase K (New England Biolabs) at 37°C for 15min. Upon heat inactivation of proteinase K at 95°C for 3min, 3uL of the cell lysate containing genomic DNA was used for PCR using the same genotyping primers from the CRISPR knock out experiments. PCR was performed using PCRBIO Taq Mix Red. PCR products were purified using ThermoFisher GeneJET PCR Purification Kit, and were sent for Sanger sequencing by 1st Base (Axil Scientific Pte Ltd, Singapore). Zygosity of the amplicons were checked using the Synthego ICE Analysis tool (v3) (Fig. 4.S4).

qPCR

To quantify the expression of miRNAs and the lncRNA, all wings were sampled from females for precise staging and avoid dealing with variation between sexes. Staging of females was performed as described before. For *mir-193* and *mir-2788* mutants, wings were sampled at Day1 (15%) pupal stage, from both homozygous mutants and WT individuals. To generate a time series expression profile for the guide strand of *mir-193*, *miR-193-3p*, wings were sampled from WT animals at 60% wanderer stage,

50% pre-pupal stage, and Day1-6 (15% - 92%) pupal stages. One forewing and one hindwing from the same individual were dissected and pooled as one biological replicate. Each condition consists of 3-4 biological replicates. Fresh tissues were preserved in RNAlater and stored at -80 ºC.

Total RNAs including small RNAs were extracted using miRNeasy Tissue/Cells Advanced Mini Kit. To quantify a specific mature miRNA strand, cDNAs were reverse-transcribed from 1ug total RNAs using stem-loop (SL) primers designed for each miRNA, together with an SL primer for the small nuclear RNA (snRNA) U6 as an internal control (Chen et al., 2005). To quantify the lncRNA, cDNAs were reverse-transcribed from 400ng total RNAs using oligo-dT. *Rps18* was used as an internal control. Reverse transcription was performed using RevertAid RT Reverse Transcription Kit. To quantify mature miRNAs, cDNA products from the reverse transcription reaction were used for qPCR using miRNA-specific qPCR forward primers, and a universal qPCR reverse primer. To quantify the lncRNA, qPCR primers for the lncRNA were used. qPCR was performed using KAPA SYBR FAST Universal Kit. The qPCR thermal cycling program was set as 40 cycles of 95ºC for 10s and 60ºC for 30s. For each biological replicate, three technical replicates were included, and expression data was analyzed using the $2^{-\Delta\Delta Ct}$ method. Standard curves were generated for qPCR primer pairs to ensure high primer efficiencies. Melting curves were also generated after each qPCR run and checked to ensure unique amplification. All SL primers and qPCR primers were listed in Table S4.

References

ABOOBAKER, A. A., TOMANCAK, P., PATEL, N., RUBIN, G. M. & LAI, E. C. 2005. Drosophila microRNAs exhibit diverse spatial expression patterns during embryonic development. *Proceedings of the National Academy of Sciences,* 102**,** 18017-18022.

ALTUVIA, Y., LANDGRAF, P., LITHWICK, G., ELEFANT, N., PFEFFER, S., ARAVIN, A., BROWNSTEIN, M. J., TUSCHL, T. & MARGALIT, H. 2005. Clustering and conservation patterns of human microRNAs. *Nucleic acids research,* 33**,** 2697-2706.

BANERJEE, T. D. & MONTEIRO, A. 2018. CRISPR-Cas9 mediated genome editing in Bicyclus anynana butterflies. *Methods and Protocols,* 1**,** 16.

CHANG, H., YI, B., MA, R., ZHANG, X., ZHAO, H. & XI, Y. 2016. CRISPR/cas9, a novel genomic tool to knock down microRNA in vitro and in vivo. *Scientific reports,* 6**,** 22312.

CHANG, T.-C., PERTEA, M., LEE, S., SALZBERG, S. L. & MENDELL, J. T. 2015. Genome-wide annotation of microRNA primary transcript structures reveals novel regulatory mechanisms. *Genome research,* 25**,** 1401-1409.

CHEN, C., RIDZON, D. A., BROOMER, A. J., ZHOU, Z., LEE, D. H., NGUYEN, J. T., BARBISIN, M., XU, N. L., MAHUVAKAR, V. R. & ANDERSEN, M. R. 2005. Real-time quantification of microRNAs by stem–loop RT–PCR. *Nucleic acids research,* 33**,** e179-e179.

EDELMAN, N. B., FRANDSEN, P. B., MIYAGI, M., CLAVIJO, B., DAVEY, J., DIKOW, R. B., GARCÍA-ACCINELLI, G., VAN BELLEGHEM, S. M., PATTERSON, N. & NEAFSEY, D. E. 2019. Genomic architecture and introgression shape a butterfly radiation. *Science,* 366**,** 594-599.

GALLANT, J. R., IMHOFF, V. E., MARTIN, A., SAVAGE, W. K., CHAMBERLAIN, N. L., POTE, B. L., PETERSON, C., SMITH, G. E., EVANS, B. & REED, R. D. 2014. Ancient homology underlies adaptive mimetic diversity across butterflies. *Nature communications,* 5**,** 4817.

GRIFFITHS-JONES, S., GROCOCK, R. J., VAN DONGEN, S., BATEMAN, A. & ENRIGHT, A. J. 2006. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic acids research,* 34**,** D140-D144.

HA, M. & KIM, V. N. 2014. Regulation of microRNA biogenesis. *Nature reviews Molecular cell biology,* 15**,** 509-524.

HANLY, J. J., LIVRAGHI, L., HERYANTO, C., MCMILLAN, W. O., JIGGINS, C. D., GILBERT, L. E. & MARTIN, A. 2022. A large deletion at the cortex locus eliminates butterfly wing patterning. *G3,* 12**,** jkac021.

HOF, A. E. V. T., CAMPAGNE, P., RIGDEN, D. J., YUNG, C. J., LINGLEY, J., QUAIL, M. A., HALL, N., DARBY, A. C. & SACCHERI, I. J. 2016. The industrial melanism mutation in British peppered moths is a transposable element. *Nature,* 534**,** 102-105.

IIJIMA, T., KAJITANI, R., KOMATA, S., LIN, C.-P., SOTA, T., ITOH, T. & FUJIWARA, H. 2018. Parallel evolution of Batesian mimicry supergene in two Papilio butterflies, P. polytes and P. memnon. *Science advances,* 4**,** eaao5416.

ITO, K., KATSUMA, S., KUWAZAKI, S., JOURAKU, A., FUJIMOTO, T., SAHARA, K., YASUKOCHI, Y., YAMAMOTO, K., TABUNOKI, H. & YOKOYAMA, T. 2016. Mapping and recombination analysis of two moth colour mutations, Black moth and Wild wing spot, in the silkworm Bombyx mori. *Heredity,* 116**,** 52-59.

JORON, M., FREZAL, L., JONES, R. T., CHAMBERLAIN, N. L., LEE, S. F., HAAG, C. R., WHIBLEY, A., BECUWE, M., BAXTER, S. W. & FERGUSON, L. 2011. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature,* 477**,** 203-206.

KAWAHARA, A. Y., STORER, C., CARVALHO, A. P. S., PLOTKIN, D. M., CONDAMINE, F. L., BRAGA, M. P., ELLIS, E. A., ST LAURENT, R. A., LI, X. & BARVE, V. 2023. A global phylogeny of butterflies reveals their evolutionary history, ancestral hosts and biogeographic origins. *Nature Ecology & Evolution*, 1-11.

KUNTE, K., ZHANG, W., TENGER-TROLANDER, A., PALMER, D., MARTIN, A., REED, R., MULLEN, S. & KRONFORST, M. 2014. Doublesex is a mimicry supergene. *Nature,* 507, 229-232.

LIANG, M., CHEN, W., LAFOUNTAIN, A. M., LIU, Y., PENG, F., XIA, R., BRADSHAW, H. & YUAN, Y.-W. 2023. Taxon-specific, phased siRNAs underlie a speciation locus in monkeyflowers. *Science,* 379, 576-582.

LIVRAGHI, L., HANLY, J. J., VAN BELLGHEM, S. M., MONTEJO-KOVACEVICH, G., VAN DER HEIJDEN, E. S., LOH, L. S., REN, A., WARREN, I. A., LEWIS, J. J. & CONCHA, C. 2021. Cortex cis-regulatory switches establish scale colour identity and pattern diversity in Heliconius. *Elife,* 10, e68549.

MA, X., HE, K., SHI, Z., LI, M., LI, F. & CHEN, X.-X. 2021. Large-scale annotation and evolution analysis of MiRNA in insects. *Genome biology and evolution,* 13, evab083.

MARTIN, A., PAPA, R., NADEAU, N. J., HILL, R. I., COUNTERMAN, B. A., HALDER, G., JIGGINS, C. D., KRONFORST, M. R., LONG, A. D. & MCMILLAN, W. O. 2012. Diversification of complex butterfly wing patterns by repeated regulatory evolution of a Wnt ligand. *Proceedings of the National Academy of Sciences,* 109, 12632-12637.

MATSUOKA, Y. & MONTEIRO, A. 2018. Melanin pathway genes regulate color and morphology of butterfly wing scales. *Cell reports,* 24, 56-65.

MATTICK, J. S. 2001. Non-coding RNAs: the architects of eukaryotic complexity. *EMBO reports,* 2, 986-991.

MONTEYS, A. M., SPENGLER, R. M., WAN, J., TECEDOR, L., LENNOX, K. A., XING, Y. & DAVIDSON, B. L. 2010. Structure and activity of putative intronic miRNA promoters. *Rna,* 16, 495-505.

NADEAU, N. J., PARDO-DIAZ, C., WHIBLEY, A., SUPPLE, M. A., SAENKO, S. V., WALLBANK, R. W., WU, G. C., MAROJA, L., FERGUSON, L. & HANLY, J. J. 2016. The gene cortex controls mimicry and crypsis in butterflies and moths. *Nature,* 534, 106-110.

QUAH, S., HUI, J. H. & HOLLAND, P. W. 2015. A burst of miRNA innovation in the early evolution of butterflies and moths. *Molecular biology and evolution,* 32, 1161-1174.

REED, R. D., PAPA, R., MARTIN, A., HINES, H. M., COUNTERMAN, B. A., PARDO-DIAZ, C., JIGGINS, C. D., CHAMBERLAIN, N. L., KRONFORST, M. R. & CHEN, R. 2011. Optix drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science,* 333, 1137-1141.

SAENKO, S. V., CHOUTEAU, M., PIRON-PRUNIER, F., BLUGEON, C., JORON, M. & LLAURENS, V. 2019. Unravelling the genes forming the wing pattern supergene in the polymorphic butterfly Heliconius numata. *EvoDevo,* 10, 1-12.

TAFT, R. J., PHEASANT, M. & MATTICK, J. S. 2007. The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays,* 29, 288-299.

TIAN, S. & MONTEIRO, A. 2022. A transcriptomic atlas underlying developmental plasticity of seasonal forms of Bicyclus anynana butterflies. *Molecular biology and evolution,* 39, msac126.

TUNSTRÖM, K., KEEHNEN, N. L., WORONIK, A., GOTTHARD, K. & WHEAT, C. 2024. The genetic basis of a regionally isolated sexual dimorphism involves cortex.

VANKUREN, N. W., MASSARDO, D., NALLU, S. & KRONFORST, M. R. 2019. Butterfly mimicry polymorphisms highlight phylogenetic limits of gene reuse in the evolution of diverse adaptations. *Molecular Biology and Evolution,* 36**,** 2842-2853.

WANG, S., TENG, D., LI, X., YANG, P., DA, W., ZHANG, Y., ZHANG, Y., LIU, G., ZHANG, X. & WAN, W. 2022. The evolution and diversification of oakleaf butterflies. *Cell,* 185**,** 3138-3152.

ZHANG, L., MARTIN, A., PERRY, M. W., VAN DER BURG, K. R., MATSUOKA, Y., MONTEIRO, A. & REED, R. D. 2017. Genetic basis of melanin pigmentation in butterfly wings. *Genetics,* 205**,** 1537-1550.

ZHOU, L., LIM, M. Y. T., KAUR, P., SAJ, A., BORTOLAMIOL-BECET, D., GOPAL, V., TOLWINSKI, N., TUCKER-KELLOGG, G. & OKAMURA, K. 2018. Importance of miRNA stability and alternative primary miRNA isoforms in gene regulation during Drosophila development. *Elife,* 7**,** e38389.