The role of *araucan* in nymphalid wing color pattern formation

Honors Thesis

Presented to the College of Arts and Sciences

Cornell University

in Partial Fulfillment of the Requirements for the

Biological Sciences Honors Program

by

Kate Siegel

May 2022

Supervisor: Dr. Robert Reed, PhD

Abstract

Despite the great beauty and diversity of butterfly wing color patterns, there is still much that remains unknown about how these color patterns are determined. Certain master regulator genes are known that can determine the scale type or pigment across a variety of color pattern elements, and their expression is determined through the combined activity of upstream pre-patterning genes. For one such key gene, *optix*, most of its regulatory network is still unknown, though candidates have been proposed. This study aims to take a closer look at the functions of one candidate upstream gene, *araucan*, a proximally expressed transcription factor that plays a role in wing vein specification in D*rosophila*. Using CRISPR-Cas9 gene editing, mutations were induced in the *araucan* gene in two nymphalid butterfly lineages, *Junonia coenia* and two co-mimetic species of *Heliconius*. Examining wings for phenotypes revealed some *optix*-typical mutations such as bright blue structural iridescence across *J. coenia* wings, as well as vein reductions in *H. erato* that suggest the function of *araucan* may be partially conserved relative to *Drosophila*. An unexpected novel phenotype affecting the iridescence of the eyespot center, or focus, also appeared, introducing a potential role for *araucan* in eyespot color patterning.

Introduction

Butterflies, in the order Lepidoptera, are known as some of the most beautiful and striking animals, with a huge variety of colors patterns across their wings. Color patterns are formed by tiny individual scales on the wing surface, which can generate beautiful colors through both pigmentation and structural iridescence. These unique color patterns play important roles in key functions such as mating, aposematism, or camouflage and mimicry. However, there is still much that is unknown about how pigments are spatially organized and deposited, and how scale

structures, such as those that produce iridescence, are regulated. A better understanding of these unique systems could provide valuable insight into how such rapid diversification could occur, and better our knowledge of how such differential gene regulatory networks function. Some key genes and their regulatory networks have already been discovered [18], for instance defining where blacks [12, 14], yellows [14, 22], or eyespots [23] will form, but how these genes are regulated, how they function, and how far their functions are evolutionarily conserved, is still largely unexplored. The additional factor of structural color [11], responsible for striking visuals such as shimmering blue butterflies, as well as UV iridescence invisible to humans [19], adds another layer of complexity to understanding how scale identity is determined. Although the process of producing structural iridescence is understood to be a product of changes in laminar thickness of the scale [20], there is again little known about how this is regulated or directed across the wing, especially in conjunction with pigmentation.

A key gene for understanding the conjunction of pigmentation and structural iridescence is *optix* [16], whose roles were functionally validated through CRISPR-Cas9 knockouts in a variety of nymphalid species [25]. *optix* was shown to act as a single switch from melanic (black and grey) to ommochrome (red, orange, and brown) pigmentation. Additionally, it acts as a switch between normal and specialized elongated conjugation scale shapes, and interestingly, in the species *Junonia coenia*, can act to repress blue iridescence, based on scale structure regardless of pigment. Notably, all three functions can operate independently and simultaneously, and this single gene is utilized like a paintbrush across species to create drastically different red color patterns. However, the upstream prepatterns that define *optix* regions of expression are unknown. There have been multiple gene expression studies examining different wing regions at different stages of development in an attempt to potentially relate certain genes to particular color pattern

elements [7, 8]. Various expression datasets were cross referenced with ATAC-seq peaks and protein binding motifs to generate a shortlist of candidate genes likely to be upstream regulators of *optix* expression (M. Chatterjee, *in prep*). This list of candidate genes included *araucan*, a member of the Iroquois gene family and transcription factor first identified in *Drosophila melanogaster* [5]. In *Drosophila*, *araucan* and its closely related genes *caupolican* and *mirror* play a role in the dorsoventral compartmentalization of the eye and wing, as well as proneural and provein prepatterning [6, 9]. This vein-related function is interesting, as some *Heliconius* butterflies used here have a hindwing ray pattern parallel to the veins, and could lead to interesting new phenotypes in *araucan* knockouts.

In this study, I more closely examined the role of *araucan* in nymphalid butterflies, primarily the Common Buckeye, *J. coenia*, and two species of tropical *Heliconius*. By inducing CRISPR-Cas9-mediated targeted mutations within the exon of the *araucan* gene, which were validated through genotyping, the resulting phenotypic changes provided insight into the potential functions of the gene. The observed effects included structural iridescent clones typical of *optix* loss, as well as vein disruptions and unique changes in the iridescence of eyespot foci without disrupting eyespot formation. This collection of phenotypes provides some suggestions for regulatory networks that *araucan* may be involved in, and the role it plays across species.

Materials and Methods

Butterfly Rearing

The *Junonia coenia* and *Vanessa cardui* wild type butterfly colonies were kept in the lab in a 27°C growth chamber on a 16-hour light cycle and were fed sugar water, while the *Heliconius erato lativitta* and *Heliconius melpomene aglaope* colonies were kept in a greenhouse enclosure

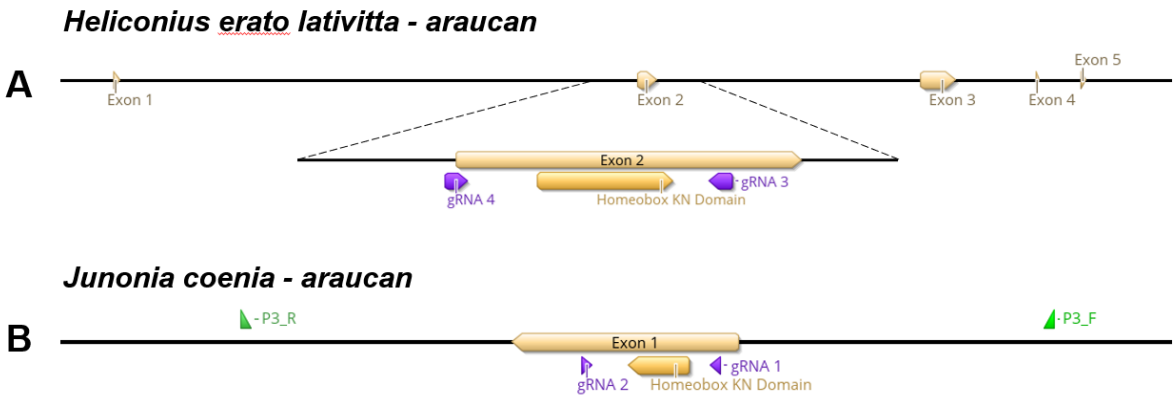
feeding on *Lantana* nectar. Both *Heliconius* species were used since their near-identical comimetic wing patterns, despite distant relation within the genus, could provide insight into conserved or diverging coloration regulation [4, 21]. Eggs were collected by providing *Plantago lanceolata* leaves to *J. coenia* and *V. cardui*, and *Passiflora biflora* leaves to the *Heliconius* colonies, which doubled as their larval food source. Following CRISPR-Cas9 injection, the surviving potentially mutant larvae, called crispants, were raised in the growth chamber, as above, in dishes and boxes of 5 – 20 individuals, being fed a prepared block of diet including important nutrients. Upon emerging from their pupae, butterflies were frozen at -20°C to keep the wings undamaged until they could be removed and examined for unusual phenotypes under the microscope, and the bodies were stored at -20°C for potential eventual DNA extraction.

CRISPR-Cas9 Injections

The injections performed to induce mutations rely on a mosaic knockout system. Cas9 is used to produce a double-stranded break at a chosen location, and the cell's imperfect repair process may accidentally generate small insertions or deletions (indels), disrupting the function of the subsequent portion of DNA. However, though the egg begins as a single cell, cells in the embryo divide exponentially over time, meaning that the gene copies within the cells around the injection site will be affected, and the mutation induction will only be successful in a portion of these. This results in an embryo containing some wild type cells and some mutant allele cells, the ratio of which can be adjusted through sgRNA concentration and the time between egg collection and injection. Too early or too high in concentration and the injections may be lethal, but too late or low in concentration and the size and frequency of visible mutations may decrease. As the embryo grows and develops, cell division produces an uneven distribution of mutated copies of the gene, only found in certain regions derived from one of the original mutagenized cells during

injection. Therefore, when knocking out color pattern genes, the whole wing will not be altered, but clearly delineated regions of an alternate phenotype, known as clones, may be visible. This produces what is known as a mosaic individual, since it will have a visual patchwork of two different phenotypes. These are also helpful in identifying phenotypes resulting from mutation, since clonal regions will be irregular and asymmetric across the body, whereas abnormalities due to natural variation will be symmetrical between the two wings.

In order to help determine the function of the *araucan* gene, the CRISPR-Cas9 gene editing system was used to induce mosaic mutations in *J. coenia* embryos. Targeted single guide RNAs (sgRNAs) were designed in accordance with the protocol developed by Zhang & Reed [24]. An annotated reference genome for *J. coenia* and *H. e. lativitta* was found on LepBase, and the *araucan* gene sequence was analyzed to find N_{20}NGG motifs. Ideal candidates would be located in an exon, and were checked to have at least 60% GC content and BLASTed to ensure no off-target binding could occur. Pairs of sgRNAs were chosen to allow not only for indels at the cut site, but for potential loss of the entire region between the two guides. The *H. e. lativitta* guide RNAs are in exon 2, separated by 216 bp (Fig 1A). The *J. coenia* guides are in exon 1 with 244 bp between them (Fig 1B). Both sets of guides target the conserved homeobox KN domain. The sequences of all guides are listed (Fig 1C, D), with corresponding PAM sites underlined.



C *Heliconius erato lativitta*
sgRNA 1: 5'-TCGTCGTCCGTCTTGTTCTTGGG-3'
sgRNA 2: 5'-TTGTCAACAGATATGGTGCTGGG-3'

D *Junonia coenia*
sgRNA 1: 5'-CTGGGTACGATTTAGCAGCCAGG-3'
sgRNA 2: 5'-CTCCTGCTCCTTGTCCTCGTCGG-3'

Primer 1: 5'-ATACTAAACCCTGCACCCGC-3'
Primer 2: 5'-ACGTACTCACAAGCGAACGT-3'

Figure 1. The locations of sgRNAs, colored purple, and primers, colored green, in relation to *araucan* exons and conserved homeobox KN domains. (A) The location of the *H. e. lativitta* guides, relative to the exon and functional domain, and relative to the gene as a whole when zoomed out. (B) The location of the *J. coenia* guides and PCR primers within the exon and gene. (C) The sequences of the *H. e. lativitta* and the (D) *J. coenia* guides, with corresponding PAM sites underlined at the end. The PCR primers targeting the region of interest in *J. coenia araucan*, used for genotyping.

For egg collection, a *Plantago lanceolata* leaf was placed into the wild type *J. coenia* and *V. cardui* colony and left there for an hour. Afterwards, the deposited eggs were carefully removed with a paintbrush and treated with 5% benzalkonium chloride for one minute to soften the shell, before being rinsed with water. For *H. e. lativitta* and *H. m.aglaope*, egg collection was done using a *Passiflora biflora* plant, and the egg shells were soft enough to not need chemical treatment. They were all then lined up along narrow strips of double-sided tape on a microscope slide. In the meantime, the corresponding two sgRNAs were combined with Cas9 and water to

achieve an approximate guide concentration of 200 ng/µL for each, and an approximate Cas9 concentration of 500 ng/µL. The resulting volume of 2 µL was kept on ice until injection.

Borosilicate glass capillary tubes were pulled into needles and were carefully loaded with a small volume of the sgRNA/Cas9 solution. Using a microinjector, careful pressure was used to inject around 2 nl of this mixture into each egg, within 2 - 3 hours after beginning egg collection.

Genotyping

Genotyping was utilized for *J. coenia* mutants to examine the region targeted by the sgRNAs to determine whether successful mutations were induced. DNA was extracted from the thorax of potential mutants, and was purified using the Monarch Genomic DNA Purification Kit. PCR primers were designed using Geneious software, ideally several hundred bases up- and downstream of the sgRNAs, to allow for low quality reads at either end (Fig 1B, D). PCR was performed using these and the product was run through gel electrophoresis. The amplified region between the primers was 1617 bp long, but if the entire sgRNA-targeted region were cut out, the size difference would be noticeable on the gel when compared to indel mutants or wild type DNA. The bands were cut from the agarose gel and the DNA purified out using the Monarch DNA Gel Extraction Kit, and would ideally be sent for Sanger sequencing. However, this step repeatedly failed, and the amplified regions were instead cloned into *E. coli* vectors and sequenced in this manner.

Ligation and transformation of the vector was performed in accordance with the protocol for the pGEM-T Easy Vector Kit. DNA from three potential mutants, as well as from a wild type *J. coenia*, was added to the provided vectors in a ratio of approximately 2:1 insert:vector, as well as the provided buffers. The next day, *E. coli* were heat shocked and incubated with the vector product, and the final transformed product was spread across LB agar plates containing

ampicillin, as well as X-Gal and IPTG. The following day, several white colonies from each plate were collected and replated to grow further overnight. A PCR was performed using the same primers to try to determine which colonies were transformed with mutant inserts. Colonies of interest were incubated and pelleted, before using the QIAgen Miniprep Kit to isolate plasmids. These were sent to the Cornell Biotechnology Core Facility for sequencing with universal M13 primers, the results of which were BLASTed against LepBase and NCBI databases to determine whether the returned insert was in fact a portion of *araucan*. Finally, they were aligned against the wild-type sequence using Geneious and Synthego ICE software to identify any successful mutations.

Results

Injections

A mixture of Cas9 and sgRNAs targeting the *araucan* gene were injected into freshly laid eggs of three different nymphalid species to try to induce mutations and observe the resulting phenotypes. The 1778 total *Vanessa cardui* eggs injected had a 19% hatch rate, which was already somewhat low, but unfortunately only three of these survived to adulthood and none of them displayed any mutations. Since the wild types also exhibited high mortality at the time, possibly due to a disease, the low survival rate is unlikely to be related to any induced mutations. It was decided that they were too difficult to work with at the time, and they will not be included in any following sections. Of the 459 injected *Heliconius* eggs from both species, 153 hatched, giving a 33% hatch rate, and 50 survived to adulthood. Of these 50, there were eight with apparent mutant phenotypes. Lastly, of the 596 injected *J. coenia* eggs, 141 hatched giving a 24% hatch rate, which can largely be accounted for by abnormally low hatch rates when first

learning how to use the microinjector on the hard-shelled *J. coenia* eggs. 92 of the larvae survived until adulthood, of which up to 39 had potential mutations.

Observed Phenotypes

Heliconius erato lativitta & Heliconius melpomene aglaope

Examining the wings of *araucan* knockouts in *Heliconius* revealed several mutant phenotypes related to vein formation, as well as some apparently related to scale development. Both co-mimetic species have a proximal base of black scales, interrupted by a large patch of red/orange and white/yellow in the middle of the forewing, and long parallel rays of red/orange on the hindwing. These hindwing rays are located along the midline between veins, which are thought to inform positioning of the ray color patterns, likely by repressing expression of some pre-pattern gene [10]. In three of the *H. e. lativitta* mutants, the A2 vein abruptly ends before reaching the wing margin (Fig 2A-A"). Soon after the termination of the vein, the two rays on either side curve inward and join at the center, since they are no longer being inhibited along the vein. In one specimen, the vein briefly reappeared partway down its natural path, before again prematurely disappearing. In this instance, the two rays that had briefly joined at the center separated again. Interestingly, although none of the *H. m. aglaope* crispants demonstrated any vein loss, at least one seemingly gained a brief vein portion at a bifurcation in the forewing (Fig 2B). There was a potential second specimen of this type, but it could not be confidently identified as a vein rather than a strong fold within the dark black scales. There were also several mutations relating to issues in scale formation and pigmentation. One *H. e. lativitta* mutant had a clearly bounded clone of darker or more vibrant color in the muted conjugation scales of the forewing (Fig 2C). The shape of the scales is the same, but appears to be a darker base pigment. There were also large regions of strangely formed scales in multiple specimens from either

species. Some scales appeared washed out, with reduced pigment deposition, and in some patches, the cover scales were all individually curled up instead of lying flat, creating a rough texture and revealing ground scales (Fig 2D).

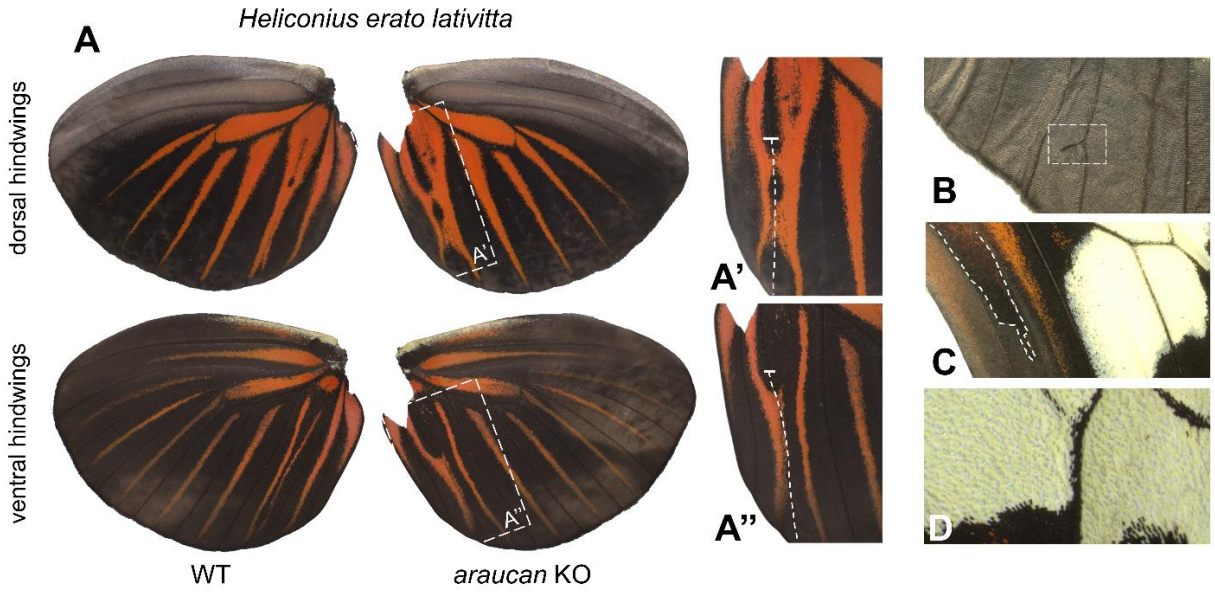


Figure 2. Observed phenotypic effects from *araucan* knockouts in *Heliconius* wings. (A) *H. erato lativitta* hindwings experienced a premature termination of the second hindwing vein, which disrupts the ray pattern formation. (A', A'') In a closer view, the dotted white line indicates the portion of the vein that is lost. (B) *H. melpomene aglaope* forewing shows a small bifurcation in the third vein, boxed in white. (C) A darker potential knockout region in *H. e. lativitta* conjugation scales, outlined in dotted white. (D) Asymmetric curled and/or discolored patches of scales appearing in various locations in multiple specimens of both species.

Junonia coenia

Examining emerged crispant butterfly wings revealed three distinct types of phenotypic mutations, largely related to structural iridescence. There is a lot of natural variation in *J. coenia* wing patterns, and they can have naturally occurring blue iridescent scales, though the lab colony generally does not. Regardless, CRISPR-induced clones can usually be found by looking for asymmetry between wings, and for clearly delineated differential regions, indicating that they are likely not caused by natural variation. However, there was still a lot of uncertainty in what could fall under natural variation. The anterior edge of the dorsal forewing is naturally slightly

iridescent, and often a few blue scales come into the region between the two orange bands. However, a number of mutants demonstrated greatly increased patches of blue in this region, not mirrored on the other wing. (Fig 3B-B'). Very often, this iridescence would extend further down the top edge and become more vibrant, as well as continue along the veins down the forewing. However, most of these altered regions did not demonstrate a complete shift to iridescence, with brown scales throughout, making it impossible to identify any clone boundaries. In the hindwing, clones were even more inconsistently affected, with many brown scales throughout, although a general outline could still be drawn around the clones (Fig 3B, B"). The shape and size of these mosaic regions varies, but they seemed to most commonly be found above and in between the two eyespots on the dorsal hindwing, never beyond, and rarely near the base of the wing. On both wings, iridescence only ever occurred dorsally, only affecting the dark brown scales and never cutting into a pattern element.

Several *J. coenia* mutants also demonstrated a change in pigmentation. In at least four specimens, a clearly defined and filled-in clone was identified in the ventral forewing, where the tan, cream, or orange pigment was lightened and muted, turning to a pale grey or tan color (Fig 3A-A'). In some instances, this could easily be confused with poorly formed wings or damaged regions where ground scales are revealed, but in some specimens the cover scales were still distinct, and parts of pattern elements were still identifiable despite the muted coloration, confirming that the scales were intact, but with altered pigmentation.

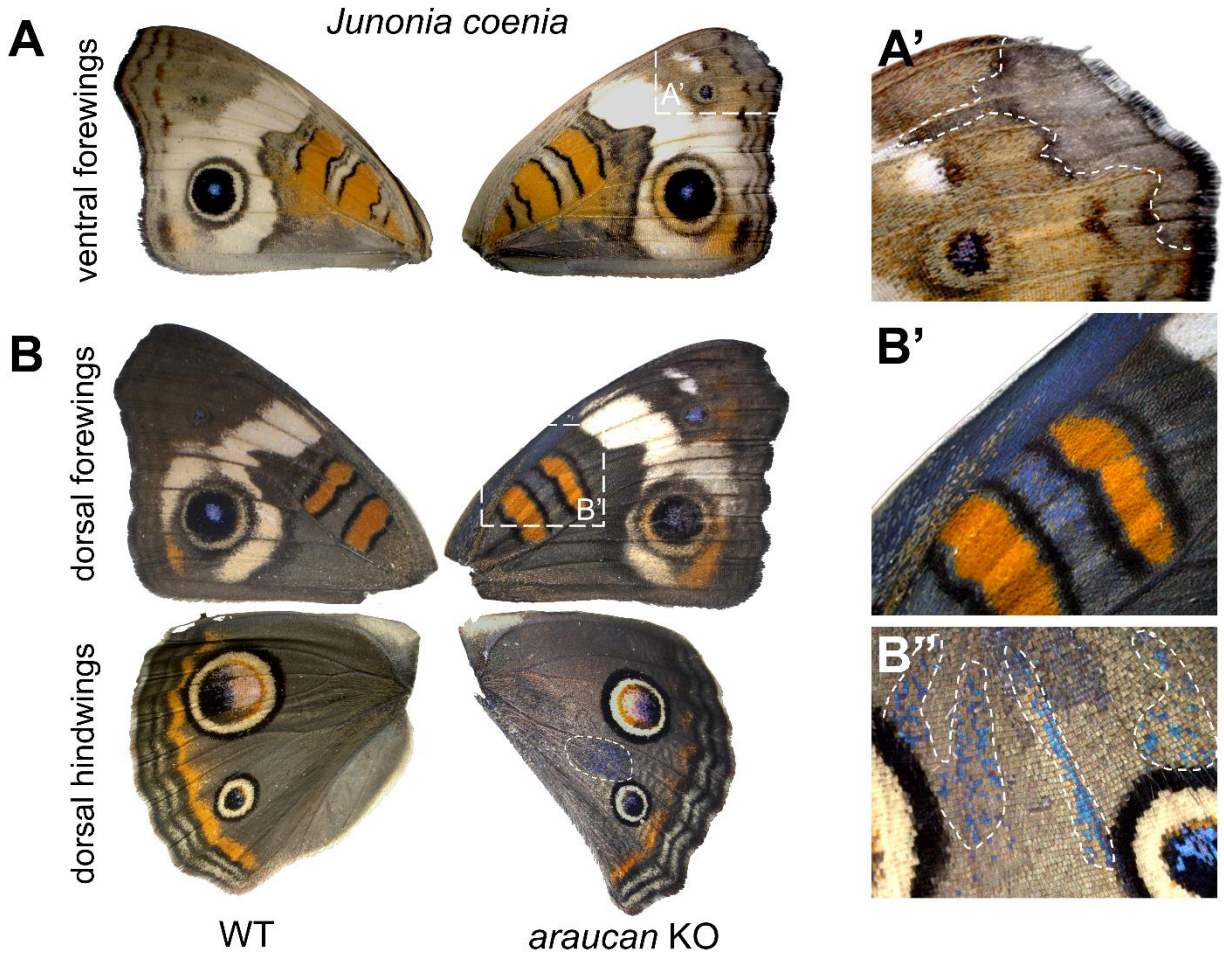


Figure 3. Observed phenotypic changes in *araucan*-knockout *J. coenia* wings. (A-A') mKO ventral forewing scales become a faded tan or grey compared to darker or more vibrant WT scales. (B-B') Expansion of structurally iridescent blue scales in the dorsal forewing, especially between the orange patches and along the veins. (B") In the dorsal hindwing, mosaic iridescent clones form anywhere in the base brown scales.

Additionally, several *J. coenia araucan*-knockout mutants demonstrated a novel, unexpected eyespot phenotype. The scales located at the center, or focus, of the *J. coenia* eyespot do not have pigment, but have a strictly structural bright blue iridescent color. In these mutant specimens, the focus of the forewing eyespot changed the color of its iridescence. The outcomes varied, including a grey-blue, tan, yellow-orange, red, pink, and several pinkish-purples, but in all cases, only the focus was affected (Fig 4A). None of the surrounding black scales ever gained iridescence or changed color, and notably, the overall shape of the eyespot always remained consistent and symmetrical to the other wing. However, in some cases the pigment in the area

surrounding the eyespot appeared slightly faded or washed out, and the edge of the outer brown eyespot ring appeared slightly blurred (Fig 4B). This eyespot change was non-mosaic and always affected the entire wing, with both the dorsal and ventral large eyespots changing to the same color, as well as the minor eyespots, if not lost to natural variation in that individual. In two specimens, the mutation appeared in both forewings, changing all eight eyespots to the same color. However, there was never any change observed in the hindwing eyespots. The shape of the individual focus scales and their lack of pigment also appeared unchanged, as the iridescent effect is still strongly visible and the scales are still translucent (Fig 4C).

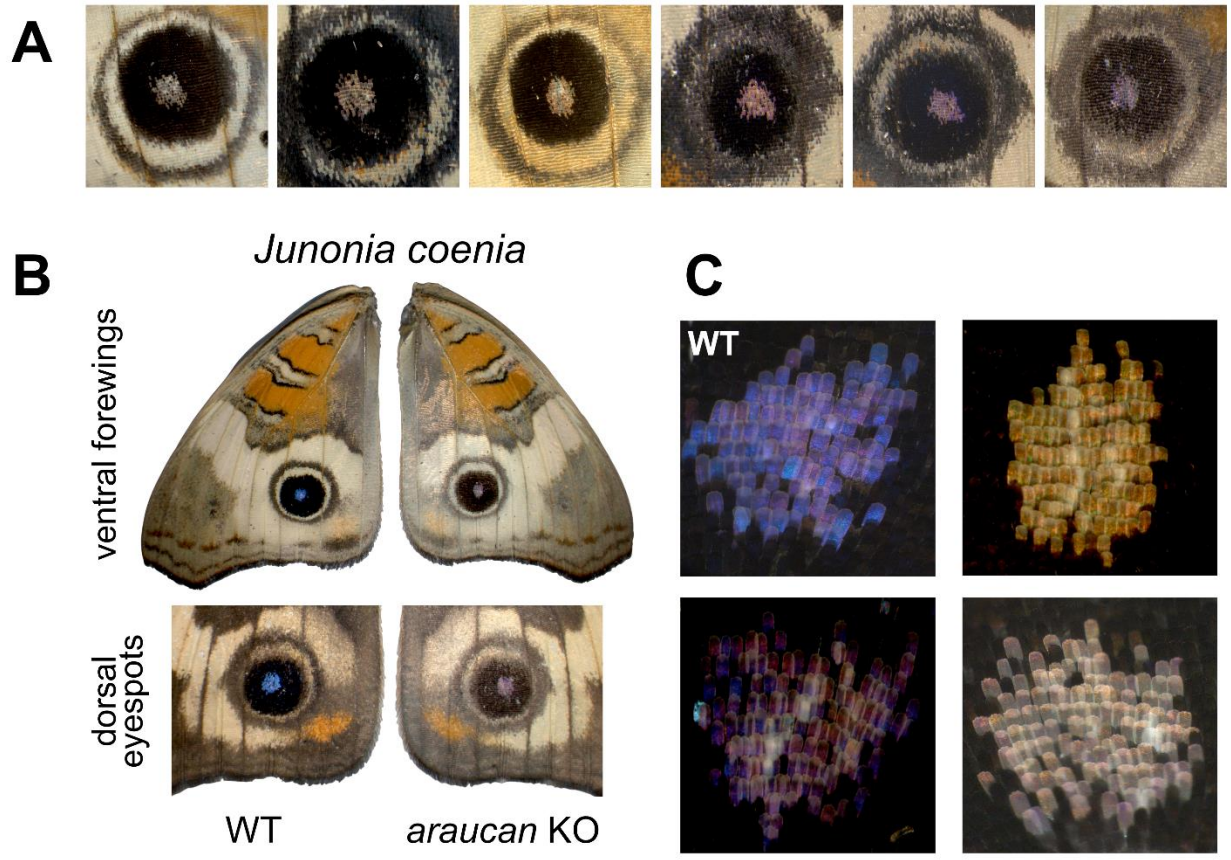


Figure 4. Observed eyespot phenotypes in *araucan*-knockout *J. coenia* wings. (A) The iridescence in the mutant eyespot foci appears in a variety of tan, orange, and pink shades, but the formation and coloration of the rest of the eyespot is unaffected. (B) In some eyespot mutants, the pigment near the eyespot appears slightly faded or washed out. (C) Although the color changes, it still appears iridescence-based, and does not affect the shape of the focus scales or the color of the surrounding black scales, when compared to the top left wild-type image.

Genotyping

Sequencing of the region of interest in mutants was attempted in order to confirm whether a mutation successfully took place at the intended sgRNA cut site locations and properly disrupted function of the *araucan* gene. Although there were several interesting *Heliconius* mutants, they were not able to be genotyped due to primer difficulties and time constraints. Multiple PCR primer pairs were tested, but none of them yielded any results when run on a gel, even for wild type DNA. Instead, *J. coenia* were prioritized as they had functional primers. Due to the lengthier process of insertion into an *E. coli* vector and its subsequent sequencing, only three mutant specimens with varying phenotypes were sequenced, as well as one wild type individual to compare. Using the online Inference of CRISPR Edits (ICE) tool from Synthego, various mutated alleles at the locations of the two gRNAs were identified (Fig 5). Each mutant exhibited at least one common allele at one of the guide locations that produced a frameshift, disrupting gene function. Using the alignment tool within Geneious also indicated that some alleles within specimens A and M experienced a complete loss of the region between the two guides. This is suggested to be common by a gel run on DNA from various mutants, many of which showed a smaller fragment corresponding to a large-scale deletion (Fig 5C). Since *J. coenia* specimens with excessive iridescence, muted pigment, and eyespot focus phenotypes were all confirmed to have a disruptive mutation in the *araucan* gene, other crispant individuals with these phenotypes can also more confidently be classified as mutants.

Figure 5. The alleles present in different *J. coenia* mutants around the intended cut sites in the *araucan* gene. Each line shows a different specimen, with the WT sequence at the top. The (A) sgRNA 1 and (B) sgRNA 2 sequence is colored green, with the corresponding PAM site in orange and underlined. The proposed cut site is represented by a vertical line. To the left of each sequence is the specimen name, the size of the indel, and the proportion of the total allele pool that it represents for that sample. (C) DNA of several potential mutants run on a gel. Relevant sizes are labeled on the ladder, a WT as seen on the right should be 1617 bp. Fragments indicating large mutations are labeled with a *.

Discussion

This study examined the effects of *araucan* on wing pattern formation in two species of *Heliconius* as well as *J. coenia*, distantly related members of the butterfly family Nymphalidae. The CRISPR-Cas9 gene editing system was used to induce mutations in the conserved homeobox domain, and the crispant larvae were raised to adulthood and examined for visible mosaic mutant phenotypes. There were a variety of mutations, differing between the species, some of which were in line with predictions and some unexpected, but which seemed to indicate three separate potential functions for *araucan* (Fig 6).

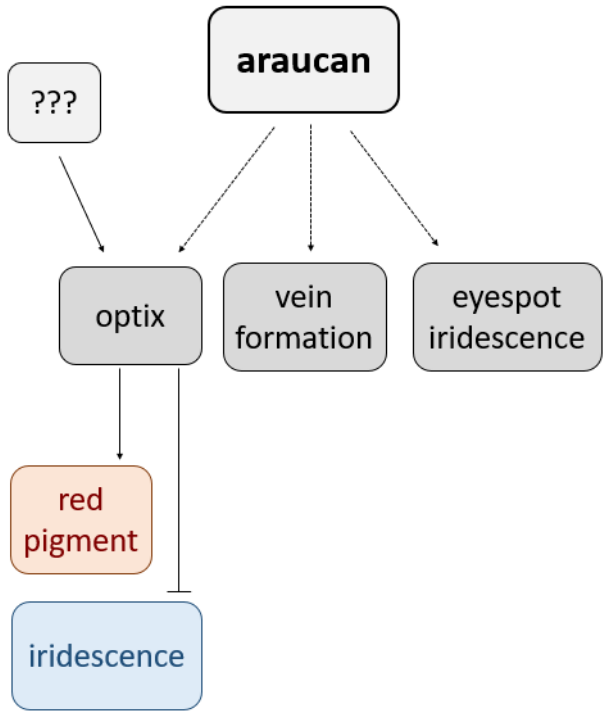


Figure 6. Chart demonstrating the three proposed functions of *araucan* based on the observed mutant phenotypes. Based on the apparent switch from red to grey scales and the appearance of iridescent blue scales in *J. coenia*, *araucan* may be one of several upstream pre-patterning genes for *optix*. The loss of the 2A vein in *H. erato* implies that its function in *Drosophila* as a vein specification gene may be partially conserved. Finally, the change in eyespot focus iridescence indicates that *araucan* may be involved in the process of eyespot coloration, or perhaps determination.

Since I first arrived at the *araucan* gene by looking for upstream regulators of *optix*, I was hoping to see some *optix*-knockout-typical phenotypes, which are threefold: ommochrome pigments switching to melanins, pointed conjugation scales reverting to normal scale shapes, and, in *J. coenia*, normal pigmented scales turning iridescent blue [25]. Some of these were observed to a certain extent. One of the *Heliconius* forewings displays a distinct clone that appears to contain darkened scale pigments, although this does not match the *optix*-typical complete switch from ommochromes to melanins. In *J. coenia*, the pigment mutations seem to involve a reduction of ommochromes, specifically in the ventral forewing, which are either muted in color or substituted with grey, in distinct clones. This is somewhat similar to the ommochrome-melanin switch in *optix* mutants, but again appears incomplete. The blue

iridescence is also very typical of *optix* knockouts, but with a reduced intensity. In all *araucan* mutants, it is near impossible to outline a specific iridescent clone on the forewing, and they could pass for extreme cases of natural variation. The hindwings have more distinctly outlined clones, but are still very sparse, with only half of the scales actually becoming iridescent. In contrast, *optix* mosaics seem to produce clearly defined solid clones of blue iridescence, that indiscriminately cut across pattern elements, overriding orange, white, and tan pigment as well as brown with blue iridescence on any wing surface.

However, despite differences in exact phenotype, many of the *araucan* mutants displayed effects similar to *optix* mutants, just in a toned-down or limited capacity. This is promising for *araucan* as an *optix* pre-pattern gene, as it would surely not be the only regulator of *optix* expression, and other genes upstream are still active in these mutants, potentially limiting the range and intensity of changes in *optix* produced by only *araucan* loss. In particular, this limited effect may be due to some redundancy in function between *araucan* and its closely related gene *caupolican*, which could be tested via dual *araucan-caupolican* knockouts. In order to confirm *araucan* as an upstream positive regulator of *optix*, examining RNA-seq data would hopefully show *optix* expressed at lower levels in *araucan* mutants, and ChIP-seq data could provide insight into the continuation of the *araucan* pathway.

Another draw towards *araucan* was its role in wing vein specification in *Drosophila*, and the potential relation to the *Heliconius* hindwing rays being parallel to and centered between hindwing veins [5, 8]. Because of this, I was hoping for an effect on either the ray pattern element or the veins themselves. The vein loss was an exciting support for a potential conserved role for *araucan* in specifying wing veins, although it was only ever the same single A2 vein affected. Additionally, it is still unclear how far this function extends, as there was no such effect

in *J. coenia* or even in *H. m. aglaope*. Since only three of the recovered *H. e. lativitta* had vein mutations, it may just be rare, and since the guides were designed for *H. e. lativitta*, they could have been less efficient in *H. m. aglaope*, but for now, this function seems to be experimentally confined to *H. e. lativitta*.

As for the actual mechanism of the *Drosophila* wing venation effect, *araucan* is a direct activator of highly specific "provein" genes [6]. It forms its dorsoventral compartmentalized prepattern for L1, L3, and L5 veins through activation by *decapentaplegic* (*dpp*) and *cubitus interruptus* (*ci*), while being repressed by *wingless* (*wg*) and *engrailed* (*en*). Elsewhere, *dpp* activates *spalt* followed by *optix*, and *spalt* represses *optix*, in order to form the L2 vein [13]. This network seems to be conserved in *Bicyclus anynana*, and potentially *J. coenia* [1]. Interestingly, it is also proposed to have been co-opted for eyespot pigmentation in *B. anynana*, where *dpp* is expressed at the eyespot focus during pupation and diffuses out, again activating *spalt* followed by *optix* but in a radial pattern, creating the black ring followed by the orange [1].

On the other hand, *distalless* (*dll*) is a key gene for initial eyespot formation during late larval wing disc development, and potentially recruits *dpp* as the morphogen to help form the eyespot as described above. Although *dll* mutants result in both eyespot gain and loss, depending on the exon affected, they also produce grey pigmentation clones, and in extreme cases even scale degradation and loss, that look remarkably similar to some of those seen in *araucan* pigmentation mutations [3]. It is proposed that this is in line with the theory of butterfly scales being homologous to *Drosophila* sensory bristles, since *dll* knockouts in *Drosophila* exhibit bristle loss through downregulation of the same proneural genes that *araucan* also activates.

These multiple overlapping regulatory gene networks make it difficult to see exactly how *araucan* fits in. If its expression in the eyespot focus occurs later, during pupal stages, it may be related to *dpp*, since it is already shown to be activated by high concentrations of *dpp* during *Drosophila* vein formation. In this case, its effect could be limited to the focus by repression from *spalt* in the black ring, since *spalt* ChIP-seq data produced by Cornell graduate student Jeannie McDonald indicates a small potential peak at the *araucan* locus. *dpp* levels are also increased in the very posterior *B. anynana* hindwing, directly along the repeatedly lost A2 vein from the *Heliconius araucan* knockouts [2]. *araucan* eyespot expression during late larval stages could implicate *dll*, since this gene drives initial eyespot formation, and they potentially share a function in scale formation and pigmentation. However, this in turn does not relate to *araucan*'s vein functions, though these were also never observed in *J. coenia*.

Many different genes are expressed in the eyespot focus throughout different stages of development, including *spalt*, *notch*, *en*, *ci*, *dll*, *wg*, and *dpp* [15, 17], many of which have been shown in other systems to either activate or repress *araucan* function. Despite this, none of these genes have ever demonstrated mutations affecting only the eyespot focus without disrupting the shape and structure of the larger eyespot, which makes it hard to narrow down exactly where *araucan* lies in these overlapping and co-opted regulatory networks. In fact, the only gene in these networks shown to affect the lower scale lamina and alter iridescence wavelengths is still *optix* [20], which is seemingly only briefly and lowly, if at all, expressed in the eyespot focus.

I have been attempting to observe *araucan* expression patterns in *J. coenia* in both 5th-instar larval wing discs and 68-hour pupal wings through immunostains. Unfortunately, these have been repeatedly unsuccessful, but if completed could potentially provide more detailed insight into where and when *araucan* is expressed, beyond the vague early pupal proximal expression

that is currently known [8]. With more time, RNA-seq and ChIP-seq data could also hopefully provide valuable insight into where exactly *araucan* fits into the regulatory landscape of *optix*, as well as the various vein- and eyespot-determining genes. For now, this study has provided support for *araucan* being an upstream regulator of *optix* in nymphalid butterflies, as well as partially conserving its role in wing venation from *Drosophila*. Even without a deeper understanding of how exactly it controls iridescent changes in the eyespot focus, or how it localizes to the focus in the first place, this still introduces an interesting and novel role for *araucan*, hopefully contributing to future understanding of eyespot regulation.

Acknowledgements

I would like to thank Dr. Robert Reed, Martik Chatterjee, and Dr. Richard Fandiño for their helpful advice and assistance throughout this process, and Martik Chatterjee specifically for his initial instruction and assistance in learning how to design sgRNA and perform CRISPR microinjections. Additionally, I thank my fellow undergraduate students, especially Xin Yi Yu, and all members of the Reed Lab for their support and accommodation while working together through these last two unusual years. Finally, I would like to thank the Cornell OUB's BioSIP Grant and the College of Arts & Sciences Summer Experience Grant for enabling me to pursue my research full-time during the summer.

References

1. Banerjee, T.D., Kwi-Shan, S., & Monteiro, A. (2022). Reuse of a wing venation gene-regulatory network in patterning the eyespot rings of butterflies. *bioRxiv* 2021.05.22.445259; https://doi.org/10.1101/2021.05.22.445259
2. Banerjee, T. D., & Monteiro, A. (2020). Molecular mechanisms underlying simplification of venation patterns in holometabolous insects. *Development*, *147*(23), dev196394. https://doi.org/10.1242/dev.196394

3. Connahs, H., Tlili, S., van Creij, J., Loo, T. Y., Banerjee, T. D., Saunders, T. E., & Monteiro, A. (2019). Activation of butterfly eyespots by Distal-less is consistent with a reaction-diffusion process. *Development*, *146*(9), dev169367. https://doi.org/10.1242/dev.169367
4. Ferguson, L. C., & Jiggins, C. D. (2009). Shared and divergent expression domains on mimetic Heliconius wings. *Evolution & development*, *11*(5), 498-512. https://doi.org/10.1111/j.1525-142X.2009.00358.x
5. Gómez-Skarmeta, J. L., del Corral, R. D., de la Calle-Mustienes, E., Ferrés-Marcó, D., & Modolell, J. (1996). Araucan and caupolican, two members of the novel iroquois complex, encode homeoproteins that control proneural and vein-forming genes. *Cell*, *85*(1), 95-105. https://doi.org/10.1016/S0092-8674(00)81085-5
6. Gómez-Skarmeta, J. L., & Modolell, J. (1996). araucan and caupolican provide a link between compartment subdivisions and patterning of sensory organs and veins in the Drosophila wing. *Genes & Development*, *10*(22), 2935-2945. https://doi.org/10.1101/gad.10.22.2935
7. Hanly, J. J., Wallbank, R. W., McMillan, W. O., & Jiggins, C. D. (2019). Conservation and flexibility in the gene regulatory landscape of heliconiine butterfly wings. *Evodevo*, *10*(1), 1-14. https://doi.org/10.1186/s13227-019-0127-4
8. Hines, H. M., Papa, R., Ruiz, M., Papanicolaou, A., Wang, C., Nijhout, H. F., ... & Reed, R. D. (2012). Transcriptome analysis reveals novel patterning and pigmentation genes underlying Heliconius butterfly wing pattern variation. *BMC genomics*, *13*(1), 1-16. https://doi.org/10.1186/1471-2164-13-288
9. Ikmi, A., Netter, S., & Coen, D. (2008). Prepatterning the Drosophila notum: the three genes of the iroquois complex play intrinsically distinct roles. *Developmental biology*, *317*(2), 634-648. https://doi.org/10.1016/j.ydbio.2007.12.034
10. Koch, P. B., & Nijhout, H. F. (2002). The role of wing veins in colour pattern development in the butterfly Papilio xuthus (Lepidoptera: Papilionidae). *European Journal of Entomology*, *99*(1), 67-72. https://doi.org/10.14411/eje.2002.012
11. Lloyd, V. J., & Nadeau, N. J. (2021). The evolution of structural colour in butterflies. *Current opinion in genetics & development*, *69*, 28-34. https://doi.org/10.1016/j.gde.2021.01.004
12. Martin, A., Papa, R., Nadeau, N. J., Hill, R. I., Counterman, B. A., Halder, G., ... & Reed, R. D. (2012). Diversification of complex butterfly wing patterns by repeated regulatory evolution of a Wnt ligand. *Proceedings of the National Academy of Sciences*, *109*(31), 12632-12637. https://doi.org/10.1073/pnas.1204800109
13. Martín, M., Ostalé, C. M., & de Celis, J. F. (2017). Patterning of the Drosophila L2 vein is driven by regulatory interactions between region-specific transcription factors expressed in response to Dpp signalling. *Development*, *144*(17), 3168-3176. https://doi.org/10.1242/dev.143461
14. Nadeau, N. J., Pardo-Diaz, C., Whibley, A., Supple, M. A., Saenko, S. V., Wallbank, R. W., ... & Jiggins, C. D. (2016). The gene cortex controls mimicry and crypsis in butterflies and moths. *Nature*, *534*(7605), 106-110. https://doi.org/10.1038/nature17961
15. Oliver, J. C., Tong, X. L., Gall, L. F., Piel, W. H., & Monteiro, A. (2012). A single origin for nymphalid butterfly eyespots followed by widespread loss of associated gene expression. https://doi.org/10.1371/journal.pgen.1002893

16. Reed, R. D., Papa, R., Martin, A., Hines, H. M., Counterman, B. A., Pardo-Diaz, C., ... & McMillan, W. O. (2011). Optix drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science*, *333*(6046), 1137-1141. https://doi.org/10.1126/science.1208227
17. Reed, R. D., Selegue, J. E., Zhang, L., & Brunetti, C. R. (2020). Transcription factors underlying wing margin color patterns and pupal cuticle markings in butterflies. *EvoDevo*, *11*(1), 1-10. https://doi.org/10.1186/s13227-020-00155-w
18. Schember, I., & Halfon, M. S. (2022). Common Themes and Future Challenges in Understanding Gene Regulatory Network Evolution. *Cells*, *11*(3), 510. https://doi.org/10.3390/cells11030510
19. Silberglied, R. E., & Taylor, O. R. (1978). Ultraviolet reflection and its behavioral role in the courtship of the sulfur butterflies Colias eurytheme and C. philodice (Lepidoptera, Pieridae). *Behavioral Ecology and Sociobiology*, *3*(3), 203-243. https://doi.org/10.1007/BF00296311
20. Thayer, R. C., Allen, F. I., & Patel, N. H. (2020). Structural color in Junonia butterflies evolves by tuning scale lamina thickness. *Elife*, *9*, e52187. https://doi.org/10.7554/eLife.52187
21. Van Belleghem, S. M., Alicea Roman, P. A., Carbia Gutierrez, H., Counterman, B. A., & Papa, R. (2020). Perfect mimicry between Heliconius butterflies is constrained by genetics and development. *Proceedings of the Royal Society B*, *287*(1931), 20201267. https://doi.org/10.1098/rspb.2020.1267
22. Westerman, E. L., VanKuren, N. W., Massardo, D., Tenger-Trolander, A., Zhang, W., Hill, R. I., ... & Kronforst, M. R. (2018). Aristaless controls butterfly wing color variation used in mimicry and mate choice. *Current Biology*, *28*(21), 3469-3474. https://doi.org/10.1016/j.cub.2018.08.051
23. Zhang, L., & Reed, R. D. (2016). Genome editing in butterflies reveals that spalt promotes and Distal-less represses eyespot colour patterns. *Nature communications*, *7*(1), 1-7. https://doi.org/10.1038/ncomms11769
24. Zhang, L., & Reed, R. D. (2017). A practical guide to CRISPR/Cas9 genome editing in Lepidoptera. In *Diversity and evolution of butterfly wing patterns* (pp. 155-172). Springer, Singapore. https://doi.org/10.1007/978-981-10-4956-9_8
25. Zhang, L., Mazo-Vargas, A., & Reed, R. D. (2017). Single master regulatory gene coordinates the evolution and development of butterfly color and iridescence. *Proceedings of the National Academy of Sciences*, *114*(40), 10707-10712. https://doi.org/10.1073/pnas.1709058114