# MOLECULAR MECHANISMS UNDERLYING VENATION AND COLOR PATTERNING IN *BICYCLUS ANYNANA* BUTTERFLIES

**TIRTHA DAS BANERJEE**
(*B.Tech, NIT Durgapur*)

**A THESIS SUBMITTED FOR THE DEGREE OF**

**DOCTOR OF PHILOSOPHY**

**DEPARTMENT OF BIOLOGICAL SCIENCES**
**FACULTY OF SCIENCE**
**NATIONAL UNIVERSITY OF SINGAPORE**

**2020**

**Supervisor:**
**Professor Antónia Monteiro**

**Examiners:**
**Professor Thorsten Wohland**
**Associate Professor Christoph Wolfram Winkler**
**Professor Jose Félix de Celis Ibeas, Universidad Autonoma de Madrid**

## DECLARATION

I hereby declare that the thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.

This thesis has also not been submitted for any degree in any university previously.

Tirtha Das Banerjee

Date: 27th September 2020

## ACKNOWLEDGMENTS

More than half of the credit of all my work presented here goes to my supervisor Dr. Antónia Monteiro, who has also been one of my best friends throughout my PhD journey. She gave me her unconditional support intellectually, financially, and emotionally. Whenever I needed her, she was ready to assist me and helped me explore new and challenging horizons in the field. I would have never been able to achieve the results and tell amazing stories out from them if she was not around. She patiently and carefully analyzed my goals and worked alongside me to provide all the help she can give. I could have never asked for a better supervisor than her.

I would like to express my gratitude towards Dr. Timothy Saunders and Dr. Christopher Winkler for being in my thesis advisory committee and providing me constructive comments at the time I needed them the most. Their expertise in the field has helped me look at my work more critically. Their suggestions helped me design new experiments and perform an in-depth analysis of my results. I would also like to thank Dr. Thorsten Wohland, Dr. Christoph Wolfram Winkler and Dr. Jose Félix de Celis Ibeas, Universidad Autonoma de Madrid for examining my thesis.

I am thankful to all my lab members who are like my family. Mrs. Tan Lu Wee, our lab manager for her support during my entire PhD work. Jocelyn Wee, for being such a great lab mate and friend. She helped me numerous

times with my experiments, maintaining my animals, analysing results, and examining my research papers. Dr. Anupama Prakash, for her wisdom and her help with many of my experiments. Suriya Murugeshwaran Narayanan, who helped me numerous times with my experiments and bioinformatics analysis; and provided many useful comments on my work. Dr. Emilie Dion and Dr. Heidi Connahs for their intellectual guidance and support at every step during my PhD journey. Dr. Yuji Matsuoka for his expertise in the field of CRISPR-Cas9. Dr. Thomas Werner for sharing his expertise in in-situ hybridization. Kwi Shan for her help with confocal microscope and antibody staining's. I would like to thank my current and former lab members; Tian Shen, Dr. Ian Chan, Dr. Hong Ru, Yi Peng, Galen Tong, Juan Olvido, Gowri V., Jun Yan, Dr. Dantong Zhu, Yi Xi Loo, Ling Sheng Loh, Sharran Rajendran, Lin Niang, Yi Ting Ter, Dean Tin Shuen Yew, Zohara Rafi, Athmaja, Dr. Eunice Tan, Li Xian Pui, Jeremy Pang, Johnathan Peh Jun Jie, Dr. Arjen van't Hof, Tricia Loo, Nesibe Ozsu, Khai Ann Hiew, Dr. Mainak DasGupta, Tan Yue Qian Clarissa, Christopher Chan, and Dr. Shivam Bhardwaj for being such great friends and helping me whenever I needed any assistance. I would also like to express my gratitude to Dr. Li Diaqin and his lab member Dr. Zeng Hua, Tammy Ho, Rachel, Dr. Chia-chen Chang, Min Tan, Wong Boon Hui, Bernetta Kwek Zi Wei, Jihea Choi, and Yu Long and for being such great friends and for their help with experiments.

I would like to express my gratitude to Dr. Fred Nijhout, Duke University; Dr. Kennith McKenna, Duke University; Dr. Ethan Bier, UC San Diego; and Dr. Rosa Barriro, CIC bioGUNE for their expert guidance during the preparation

of the second chapter of my thesis entitled “Molecular mechanisms underlying simplification of venation patterning in holometabolous insects”. Dr. Nipam Patel, Marine Biological Laboratory, MA; Dr. Robert Reed, Cornell University; Dr. Nicholas Tolwinsky, Yale-NUS college for antibodies used in this thesis. I am thankful to Ms. Tong Yang, CBIS DBS, for her support with the confocal microscopes. I would like to thank Dr. Sufia K. Kazy, NIT Durgapur; and Dr. Pinaki Sar, IIT Kharagpur for recognising my interest in the field of research and giving me the opportunity to work under them during my undergraduate years. I would like to thank Dr. Thorsten Wohland for his teachings on optical physics that helped me during my confocal microscopy sessions and Dr. Roman Carrasco for his inputs on the understanding of the R programming language.

Last but not least I would like to express my deepest gratitude towards my family members for raising me to be an independent and hardworking person, and their guidance to be kind and generous to everyone around me. They gave me their unconditional love and support whenever I needed during this whole journey and without their support, I would have never been able to reach this place I am right now.

## TABLE OF CONTENTS

## SUMMARY

The diversity of patterns on seashells, on the skin of fishes, the flanks of zebra, and the wings of butterflies is captivating and mesmerizing. Much pattern diversity is due to variation in how cells perceive their relative position to different sources of morphogens, the type of morphogen, and how cells interpret morphogen concentration gradients. The present thesis explored this Positional Information model of pattern formation in two specific types of biological patterns using *Bicyclus anynana* as a model organism: the arrangement of veins (venation) and the eyespot markings made of concentric rings of color on the wings.

Ancestral insects had more veins than current-day insects, and current-day butterflies have more complex venation than flies, however, variation in the molecular mechanisms that led to these distinct venation patterns are unknown. I investigated vein patterning mechanisms in *B. anynana* butterflies and compared these mechanisms with those previously described in *Drosophila melanogaster* flies. I used immunostaining and *in-situ* hybridization to study the expression domains of genes such as *engrailed (en), invected (inv), decapentaplegic (dpp), spalt (sal), wingless (wg), armadillo (arm), aristal-less (al), optix (opx), blistered (bs),* and *rhomboid (rho);* and CRISPR-Cas9 and drug antagonist to study the function of *dpp, sal, wg, arm,* and *opx*. The results illustrate that in developing butterfly wings there are additional morphogen sources relative to fly wings, which set up additional domains of downstream gene expression such as *sal* and *opx,* based on different concentration thresholds. This is important as the boundaries of

expression of these morphogen-response genes lead to the formation of veins in flies. Mutating or downregulating the morphogen gene, *dpp*, or mutating the downstream gene, *sal*, that responds to the morphogen, alters venation patterns in butterflies. In addition, while only a few of these *sal*-expression boundaries lead to veins in flies, additional *sal* boundaries are used to set-up veins in butterflies. Downregulating Wg signaling and mutating *opx* doesn't produce any defect in venation. I concluded that derived insects such as flies lost *dpp* and *sal* expression domains, and as a result, lost veins at the *sal* boundaries. Further simplification in venation occurred due to the loss of vein developmental programs at the boundaries of the *sal* gene domains and loss of vein maintenance genes such as *rho* and activation of vein suppressor gene *bs* in the provein cells.

The next step in my research was to explore the possible co-option of the positional information network involved in venation patterning to pattern the rings of eyespots in butterflies. The rings of the eyespots are believed to be patterned by a positional information system by a morphogen secreted from the center of the eyespot. I aimed at two unexplored eyespot genes that are involved in the positional information system mentioned above: *en* and *opx*. First, I explored the gene *opx* which has also been implicated in the development of orange and red pigments in other butterfly species. Immunostaining of Opx showed expression in the orange ring where *opx* controls both the pigmentation and development of scale lamina in between the cross-ribs of the orange scales. Furthermore, I explored the involvement of Opx with other genes such as Dpp and Sal. The results can be explained by the presence of a similar positional information system as that

observed in the anterior-posterior patterning of veins where Dpp likely acts as a central morphogen activating *sal* at a high concentration threshold in the black scale region and *opx* at a low concentration threshold. Sal then represses the activation of *opx* in the black scales and as a result, *opx* is expressed in a halo-like orange pattern. In the next study I focused on *en* and its family members that are proposed to be expressed in the orange ring and the center of eyespots. I found three copies of *en*-family member genes that are differentially expressed in the eyespots of *B. anynana*. The first copy, *en* is expressed in the eyespot center and in the orange ring, while the other two copies *inv* and *inv-like* are expressed only in the eyespot center. Comparing the presence and orientation of these three copies in other Lepidoptera with *B. anynana* I concluded that these genes have duplicated much earlier than the appearance of the first eyespot, and hence these genes are not retained in the genome due to their involvement with the development of this novel trait in evolution, the eyespots.

In the final chapter, I provide a conclusion of my work and propose future areas of exploration. The work presented here is nowhere near completion. Further comprehensive work is necessary to enhance our understanding of these amazing areas of evo-devo. I propose additional research and experiments to accelerate our knowledge of the processes that pattern our biological world.

## List of figures

## List of tables

## Chapter 1: Introduction: How nature patterns our world?

Stripes of zebra, spots of leopards, petals of flowers, markings on seashells, arrangement of animal bones, feathers of birds and their vibrant colors, and the patterns on the wings of butterflies are some of the fascinating examples in nature that has both intrigued and inspired scientists for centuries to explore the underlying molecular mechanism.

### Biological patterning in nature

One of the most fundamental topics in developmental biology is pattern formation, or the process by which cells acquire different identities depending on their relative spatial positions within a tissue. Colorful markings on sea-shell (Meindhardt, 2012), pigment patterns in zebras and giraffes (Ball, 2015), limbs of mammals (Raspopovic et al., 2014), skin markings on pufferfish (Sanderson et al., 2006), wing venation in insects (Comstock, J.H. Needham, 1898), and eyespots in butterflies (Monteiro, 2015) are some of the remarkable examples of the complex patterns and traits observed in nature that result from pattern formation processes. Many of these traits play important roles in sexual signaling, survival, and camouflage, yet, the molecular mechanisms underlying their development are still not fully understood.

Two popular theories that describe pattern formation in nature are Alan Turing's Reaction-Diffusion (RD) and Lewis Wolpert's Positional Information (PI) models (Green and Sharpe, 2015).

**Reaction-Diffusion**

The great mathematician and code breaker Alan Turing in 1952 proposed a model to explain the development of complex color patterns in organisms that relied on two substances interacting and diffusing through tissues (Turing, 1952) which was later rediscovered and elaborated by subsequent scientists. Turing described two diffusible molecules (called morphogens: a substance capable of morphogenesis) that interacted with each other to produce periodic ripples of high and low concentration giving rise to regular spatial patterns from a starting uniform state. A. Gierer and H. Meinhardt, 20 years later, independently proposed models similar to Alan Turing's RD (Gierer and

Meinhardt, 1972). In the substrate-depletion model, an activator consumes a substrate to autoactivate itself and produce periodic patterns where the activator and substrate are out-of-phase (**Fig. 1A**) (Gierer and Meinhardt, 1972). In the Activator-Inhibitor model, a short-range activator autoactivates itself and a long-range inhibitor. The inhibitor diffuses faster than the activator resulting in periodic patterns where both activator and inhibitor are in-phase (**Fig. 1B**). The theory of RD has been used to explain the patterns in the shells of molluscs (Meindhardt, 2012), patterning of *Drosophila* embryos [10,11], complex skin pigmentations (Sanderson et al., 2006) and insect venation (Yoshimoto and Kondo, 2012). Some more examples where RD has been implemented are mentioned below.

x----------x----------x

**Examples of Reaction-Diffusion**

Activator-Inhibitor**:** Experimental data have demonstrated the involvement of the activator-inhibitor model of RD in several case studies: In zebrafish left-right asymmetry during embryogenesis, where short range Nodal acts as an activator of itself and of the long range inhibitor Lefty, which supresses the expression of Nodal (Müller et al., 2012); In patterning palatal ruggae (folds on the anterior part of palatal mucosa) where Fibroblast Growth Factor (FGF) acts as short range activator and Sonic hedgehog (Shh) act as long range inhibitor (Economou et al., 2012); in the development of hair follicles in mice skin, where Wingless signaling (WNT) acts as short range activator and Dickkopf (DKK) as a long range inhibitor (Sick et al., 2006); and in avian feather bud formation where FGF signaling act as short-range inhibitor and

Bone Morphogenic Protein (BMP) signaling act as long range inhibitor (Jung et al., 1998).

Substrate-Depletion**:** In addition, the substrate-depletion model of RD has been implicated in a few other studies: In lung branching where Shh acts as short-range activator and FGF as substrate (Menshykau et al., 2012); and in a more sophisticated set of experiments, researchers have shown a three-molecule RD system equivalent to the substrate-depletion model proposed by Gierer and Meinhardt, 1972 (Gierer and Meinhardt, 1972) which involves a feedback loop between WNT and BMP signaling and Sox9 (Raspopovic et al., 2014). In mouse limb development, the BMP and WNT signaling pathways along with the transcription factor Sox9 have been shown to form a Turing RD system where activated BMP and WNT signaling pathways are expressed out of phase while Sox9 is out of phase with active WNT signaling (Raspopovic et al., 2014). Inhibition of BMP signaling resulted in loss of digits, while WNT inhibition resulted in loss of interdigit tissue. Inhibition of both BMP and WNT resulted in broader and fewer digits. Simulation experiments have confirmed these results and precisely replicated digit patterning.

**x----------x----------x**

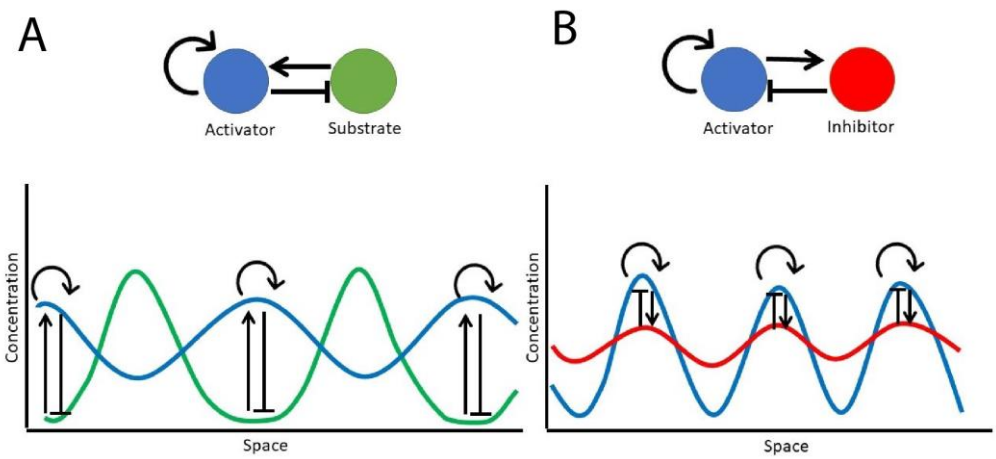



**Figure 1.1: Two models of Reaction diffusion.** (**A**) Activator-substrate: In this model an activator consumes a fast diffusing substrate to autoactivate itself and produce periodic patterns where the activator and substrate are out-of-phase. (**B**) Activator-inhibitor: In this model a short-range activator autoactivates itself and a long-range inhibitor. The inhibitor diffuses faster than the activator resulting in periodic patterns where both activator and inhibitor are in-phase.

## Positional Information

Lewis Wolpert in 1969 and 1971 proposed an alternative theory of Positional Information (PI) to explain how complex patterns appear from simple asymmetries initially present in an embryo, rather than from a uniform state (Wolpert, 1969; Wolpert, 1971). He proposed that different cell fates are determined based on the differences in concentration of morphogens diffusing across tissues in a position-specific manner (Wolpert, 1971), i.e., the cells that receive higher morphogen concentrations, closer to the source of morphogen, follow a differentiation process that is different compared to cells that receive lower concentrations, positioned further away from the morphogen source (**Fig. 1.2**). This theory has been used explain the different cell fates in *Xenopus* embryos (Green and Smith, 1991), segmentation in *Drosophila* embryos

(Green and Sharpe, 2015) and *Drosophila* wing venation (Bier, 2000; De Celis, 2003), among multiple other traits.

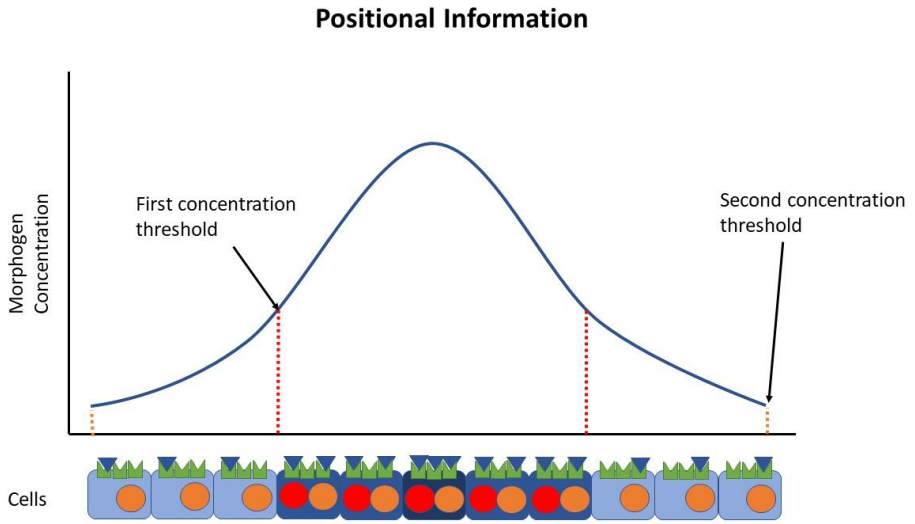



**Figure 1.2. A model of Positional information.** In this model cells at the center (here a single cell) produces a morphogen which diffuses around the surrounding cells. The cells which receive high concentration of the morphogen activates two genes (red and orange). When the concentration falls below this threshold only the orange gene is activated.

A recent study proposed how both reaction-diffusion and positional-information might work together to achieve patterning during embryogenesis (Tewary et al., 2017). In this study researchers have proposed a two-molecule activator inhibitor Reaction-diffusion system which involves BMP4 (activator) and Noggin (inhibitor) which results in the periodic activation of pSMAD1 which then activates other downstream genes such as *CDX2*, *BRA* and *SOX2* based on different concentration thresholds i.e., via a positional information mechanism.

**Venation patterning in insects**

A favourite model system for studies of pattern formation is the flat, two-dimensional insect wing composed of two layers of single-celled sheets supported by veins (Bier, 2000; De Celis, 2003). Insect wings originated in Pterygotes around 400 mya and allowed these insects to more easily disperse, avoid predators, and catch prey. The venation patterns on the wings of insects play a variety of functions in developing as well as adult insects. Veins are ectodermal hollow tubes that develop in between the dorsal and ventral wing epidermis during larval and pupal wing development, and are reinforced by chitin during later developmental stages before the eclosion of adults (Blair, 2007; Waddington, 1940). Veins are involved in providing structural support to the wing and the arrangement of veins (venation patterning) is important for insect flight as the flight muscles are located at the base of the wings (Combes, 2003). Veins act as vessels for blood cells (hemolymph) involved in nourishing developing wings (Chintapalli and Hillyer, 2016) and live cells in the adult wing, such as pheromone secretory cells (Dion et al., 2016), and even play an auditory function (Sun et al., 2018). Tracheal tissue which is a part of the respiratory system of insects develops inside the veins and is used to transport oxygen to different regions of the wing (Quinlan and Gibbs, 2006). The wing veins also play roles in color pattern development (Koch and Nijhout, 2002), most exorbitantly in butterflies and moths, where many of the colorful markings which play vital roles in sexual and natural selection are vein-dependent. Finally, because veins are extremely conserved within a species, but variable across species, they are used extensively in the identification of insect species (Kaba et al., 2017).

Unlike most other insects, the molecular mechanisms underlying venation patterning in *D. melanogaster* have been studied in great detail and follow the positional information paradigm (**Fig. 1.3**; reviewed in Bier, 2000; Blair, 2007; De Celis, 2003; Garcia-bellido and Celis, 1992). Bone Morphogenic Protein (BMP) signaling is an important event that begins the process of vein formation in flies. The BMP signaling ligand Dpp (Decapentaplegic) diffuses from the anterior-posterior (A-P) boundary of the *D. melanogaster* wing blade, set up in early embryogenesis via the segmentation genes, and activates genes such as *spalt* (*sal*), *aristaless* (*al*), *optix*, and *optomotor-blind* (*omb*) based on different concentration thresholds. Veins then develop on the expression boundaries of these genes (Al Khatib et al., 2017; Blair, 2007; Martín et al., 2017). This process leads to the formation of 5 complete longitudinal veins spanning the proximal and distal edge of wing blade numbered as R1 (L1), R2+R3 (L2), R4+R5 (L3), M1 (L4) and CuA1 (L5) (numbering is based on Comstock-Needham system; Comstock, J.H. Needham, 1898; see**Fig. 1.4D**). There are also two incomplete longitudinal veins Sc (L0) and A1+CuA1 (L6), and one incomplete anal vein A2 (**Fig. 1.4D**). Some of the key genes involved in the development, maintenance, and patterning of the wing veins include *engrailed* (*en*) (Guillén et al., 1995; Blair, 1992) , *hedgehog* (*hh*) (Ingham and Fietz, 1995), *dpp* (Lawrence and Struhl, 1996), *sal* (Sturtevant et al., 1997), *omb* (Grimm and Pflugfelder, 1996), *brinker* (*brk*) (Campbell and Tomlinson, 1999a), *wingless* (*wg*) (Yang et al., 2013), *al* (Martín et al., 2017), *optix* (Martín et al., 2017) and *rhomboid (rho)*(Lunde et al., 1998), which will be explored in a comparative context with butterflies in chapter two.

In addition to the longitudinal veins described above, there is also a marginal vein that outlines the fly wing. Wnt signaling plays a role in marginal vein formation (Blair, 2007). One of the Wnt molecules, *wg* (also called *wnt1*) (Martin and Reed, 2014), is expressed along the wing margin in the developing larval wing disc (Yang et al., 2013). *wg* likely results in the activation of *rho* (the earliest marker of vein cells (Lunde et al., 1998)) in the adjacent cells of the dorsal-ventral (D-V) margin (Alexandre et al., 1999) which likely activate epidermal growth factor receptor (EGFR) signaling resulting in marginal vein development (Blair, 2007). Though we understand venation and vein development in great detail in this highly derived Dipteran species (see below for detailed description) (Misof and et al., 2014), it is possible that many of the molecular mechanisms involved might be derived and not representative of venation patterning in other insect lineages.

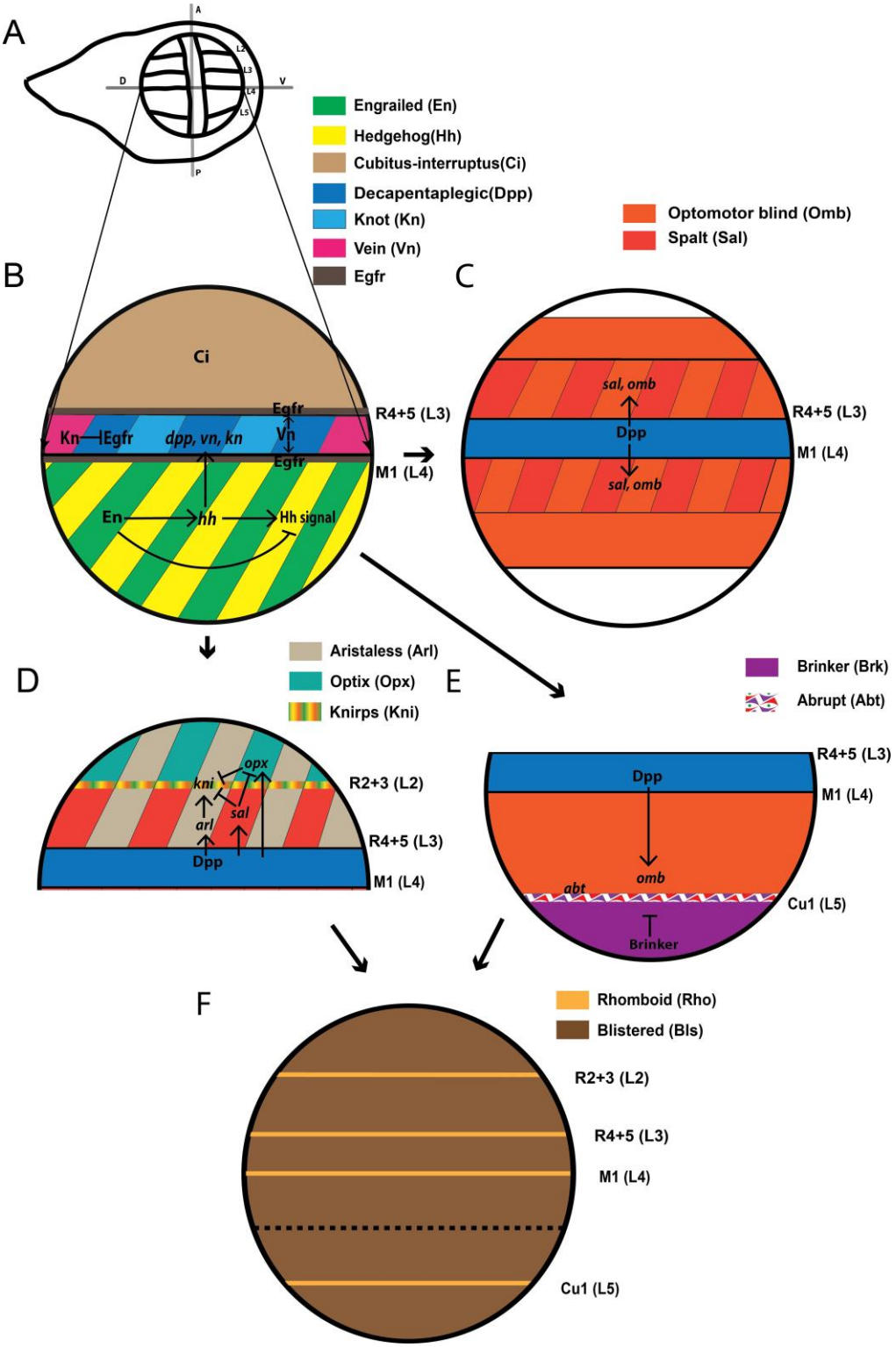



**Figure 1.3. Molecular mechanism involved in venation patterning in *Drosophila melanogaster*. (A)** Larval wing disc of *D. melanogaster*. During the larval stage, the wing is divided into two populations of immiscible cells belonging to the Anterior (A) and Posterior (P) compartments. The boundary where these two-populations meets is referred to as the Anterior-Posterior (A-P) boundary (marked by the gray line). **(B)** Venation patterning is initiated by the transcription factors En and Inv in the posterior compartment that activate

expression of *hh* while suppressing Hh signaling. Hh is a short-range diffusible morphogen. A small amount of Hh diffuses into the anterior compartment where the presence of Ci activates the BMP ligand *dpp*. Hh also activates the genes *vein* and *knot* overlapping the expression of *dpp*. Knot inhibits Egfr signaling at the R4+5 (L3) and M1 (L4) intervein cells. The veins R4+5 (L3) and M1 (L4) form at the anterior and posterior boundary of the *dpp* and *vein* expression domain due to activation of Egfr signaling via Vein protein. **(C)** Dpp protein then acts as a long-range morphogen activating both *spalt* (*sal*) and *optomotor-blind* (*omb*) at high concentrations, and only *omb* when the concentration falls below the *sal*-inducing threshold. (**D**) The vein R2+3 (L2) forms by the interaction of Al, Opx and Sal. Dpp activates all three transcription factors at different concentration thresholds. Al activates the R2+3 (L2) vein specific gene *knirps* while Opx and Sal suppresses it. (**E**) The vein Cu1 (L5) forms at the boundary of Omb and Brinker where the Cu1 (L5) specific gene *abrupt* is expressed. (**F**) The final step of venation patterning involves expression of Rho in the vein cells and Bs in the intervein cells.

**x----------x----------x**

**Detailed description of the mechanisms involved in *D. melanogaster* wing venation (Positional Information)**

The earliest marker of imaginal wing disc formation is the expression of scalloped in clusters of cells located in the embryonic ectoderm. The genes *en* and its paralog *invected* (*inv*) are expressed in a group of posterior cells and are responsible for establishing the anterior-posterior (A-P) wing boundary. En represses the expression of another segment polarity gene called *cubitus interruptus* (*ci*), whose expression is restricted to the anterior compartment of the wing disc (Schwartz et al., 1995; Zecca et al., 1995). Ci, in-turn, supresses *en* in the anterior compartment (Eaton and Kornberg, 1990). Due to the complementary expression pattern of *en* and *ci*, a thin stripe of cells just above the A-P boundary respond to Hh diffusing from the posterior compartment (Bier, 2000) where it activates the BMP ligand *dpp* (Ingham and Fietz, 1995), the epidermal growth factor receptor (EGFR) signaling ligand *vein* (*vn*)

(Schnepp et al., 1996) and transcription factor Knot (Kn) (Mohler et al., 2000). The veins R4+R5 (L3) and M1 (L4) form on the boundary of the expression domains of these genes (Bier, 2000; Blair, 2007). Hh signaling controls the area of *dpp* signaling by controlling the expression of *master of thickveins* (*mkv*) which downregulates the receptor of Dpp, Thickvein (Tkv) (Funakoshi et al., 2001). Dpp is a secreted morphogen (Akiyama and Gibson, 2015) which activates the transcription factors *sal* and *omb* at different threshold concentrations (Strigini and Cohen, 1999a). High concentration of Dpp activates *sal* in a stripe of tissue surrounding the A-P boundary (Zecca et al., 1996). Dpp is also involved in the activation of *al* in the anterior compartment and *opx* in the upper anterior compartment (Martín et al., 2017). Opx is expressed in the upper anterior compartment as a result of repression from sal (Martín et al., 2017). The complementary expression of *opx* and *sal* creates a perfect environment where the gene *al* activates *knirps* (*kni*) along the vein R2+R3 (L2) (Bier, 2000; Martín et al., 2017) which later activates the pro-vein marker gene *rhomboid* (*rho*) in the R2+R3 (L2) pro-vein (Blair, 2007). *omb* is activated in a much broader domain of tissue in response to low concentrations of Dpp (Cook et al., 2004). Omb along with Sal and the transcription factor Brinker (Brk) determine the position of the CuA1 (L5) vein (Cook et al., 2004). Brk is a repressor of Dpp targets and Dpp in turn represses *brk* expression (Campbell and Tomlinson, 1999a). The loss of Brk in cells posterior to CuA1 (L5) results in ectopic vein formation (Cook et al., 2004). During the later stages of development, the transcription factor Abrupt (Ab) are expressed in the CuA1 (L5) vein, and their loss leads to partial loss of the CuA1 (L5) vein (Cook et al., 2004). Two transcription factors of Iroquois

Complex (Iro-C), *araucan* and *caupolican* are expressed in the L3 and L5 veins (Gömez-Skarmeta et al., 1996). Rho is expressed in all the pro-vein cells from larval wing to late pupal wing development (Guichard et al., 1999; Lunde et al., 1998). Rho activates *argos* (*aos*) (Biehs et al., 1998) along the pro-vein cells where *aos* acts as an inhibitor of Egfr signaling by binding with its ligand *spitz* (*spi*) (Schweitzer et al., 1995). Rho is also likely involved in the activation *spi* and it's homolog *keren* (*krn*) which are involved in early maintenance of LVs (Blair, 2007). *blistered* (*bs*), a mammalian ortholog of Serum Response Factor (SRF) is expressed and is essential for intervein cell fate during larval and pupal wing development (Roch et al., 1998). *bs* mutation results in ectopic expression of *rho* and results in development of veins in intervein tissues (Roch et al., 1998). During the later stages of development *dpp* is expressed along the vein cells (Ralston and Blair, 2005) and along with its paralog *glass bottom boat* (*gbb*) is essential of LVs maintenance (Bangi and Wharton, 2006; Blair, 2007). The two cross-veins (CVs), i.e., the Anterior Cross Vein (ACV), connecting R4+R5 - M1 (L3-L4), and the Posterior Cross Vein (PCV), connecting X and Y, form much later in development at around 19hrs after pupation as marked by phosphorylated mothers against decapentaplegic (pMAD), a signal transducer of the BMP signaling pathway (Conley et al., 2000). Dpp along Gbb are involved in PCV formation(Ray and Wharton, 2001). BMP binding protein Short gastrulation (Sog) and Crossveinless (Cv) help in the transport of a Dpp and Gbb dimer into the CV area (Shimmi et al., 2005). The morphogen *wingless* (*wg*) is also present in the CVs (Phillips and Whittle, 1993) but its removal has only minor effects on CV formation (Blair, 2007).

x----------x----------x

Recently, evidence that Dpp transport might be involved in venation patterning came from the sawfly *Athalia rosae* (Hymenoptera; Tenthredinidae) (Matsuda et al., 2013). Sawflies diverged around 340 mya from a common ancestor with Coleoptera, Lepidoptera, Trichoptera, and Diptera (Misof and et al., 2014), and have more complex venation than *D. melanogaster* (see **Fig 1.3**). In *D. melanogaster* pupal wings, a Decapentaplegic-Glass bottom boat (Dpp-Gbb) dimer expressed along the longitudinal veins is transported to the area of the cross veins with the help of the Short gastrulation-Crossveinless (Sog-Cv) transporter and is required for cross-vein development (Shimmi et al., 2005). In sawflies, *dpp* expression is observed all over the wing during pupal wing development. However, pMAD (the signal transducer of Dpp) expression is restricted to the areas where future veins will form indicating BMP signaling is active only in pro-vein cells. Knocking down of Cv using RNAi resulted in sawflies with complete loss of veins indicating that the transport of BMP ligands is pivotal for vein development. So far, there are no further functional studies on other signaling pathways in sawflies in regard to vein development.

Venation in other insect lineages has remained poorly understood. In *Locusta migratoria* (Orthoptera) the Dpp target gene *sal* has been shown to be involved in venation patterning, as a RNAi –mediated knock down resulted in venation defects in the posterior hindwing (Wang et al., 2017). In the ant species *Pheidole morrisi* (Hymenoptera) the genes *en*, *wg* and *sal* have similar wing expression domains as observed in *D. melanogaster* (Abouheif and

Wray, 2002). In the scuttle fly *Megaselia abdita*, *en*, *sal*, *omb*, *blistered* (*bs*), *rho* and *delta* seem to have similar expression patterns as *Drosophila* (Gantz, 2015). Kni which is involved in the development of the vein R2+R3 (L2) in *D. melanogaster* has two strips in *M. abdita*, likely involved in the development of R2+R3 and M1 veins (Gantz, 2015). In Lepidoptera, which includes butterflies and moths, wing veins have been studied mainly for their role in color pattern development (Koch and Nijhout, 2002; Reed and Gilbert, 2004; Schachat and Brown, 2015). However, the molecular mechanisms leading to venation patterning in lepidopteran, have remained unexplored.

Simulation work on the Hemipteran species *Orosanga japonicus* indicates that RD along with PI might be involved in venation patterning in this species (Yoshimoto and Kondo, 2012). Insects of the order Hemiptera diverged from Hymenoptera and Diptera around 370 mya (Misof and et al., 2014). *O. japonicus* has variable venation patterns among individuals and between left and right wings of the same individual where longitudinal veins bifurcate randomly as they run from the proximal to the distal axis of the wing. Researchers found that RD can generate most of the bifurcation patterns, however the position of the bifurcations were almost random. They also used a PI model, where a morphogen is released from the center of the wing, along a line, similarly to Dpp diffusion along the A-P axis of *D. melanogaster* and found that this model can accurately position the veins but cannot generate the bifurcation patterns observed. Therefore, they combined both models: first creating asymmetries on the wing blade by the release of a morphogen along the A-P axis, and then applied RD diffusion on the initial asymmetries generated via PI. The combination of PI and RD accurately generated *O.*

*japonicus* venation patterns (Yoshimoto and Kondo, 2012). The theory of PI upstream of RD has also been proposed for mouse limb patterning where FGF signaling (PI) along the distal limb margin acts upstream of a RD patterning system (involving WNT-Sox9-BMP) to pattern the digits along the proximodistal axis. In the proximal limb region the concentration of Fgf is low which produces a RD pattern with a smaller wavelength, and in the distal limb region the concentration of Fgf is high which produces a RD pattern with longer wavelength (Green and Sharpe, 2015).

A computer simulation study proposes that venation patterning in dragonflies and damselflies involves the process of reaction-diffusion (Hoffmann et al., 2018a). Dragonflies and damselflies have an ancestral venation pattern with highly complex cross-veins that forms polygons. Using high resolution images of dragonfly and damselfly wings researchers tried to simulate the formation of secondary cross-veins by adding inhibitory signals in between the primary longitudinal veins. They discovered that cross-veins develop at the boundaries where the inhibitory signals are at its minimum and proposed that a process of reaction-diffusion is involved in the formation of cross-veins (Hoffmann et al., 2018a).

**Color patterning in butterflies**

Butterflies display a wide variety of color patterns on their wings. Some examples include bright red and yellow patches on the wings of *Heliconius* butterflies (Livraghi et al., 2020; Martin et al., 2014), spots and stripes of orange and black on the wings of milkweed butterflies (*Dannus plexipus*) (Sekimura and Nijhout, 2017), bright blue iridescent color on the wings of the

common buckeye (*Junonia coenia*) (Thayer et al., 2020), and the colorful concentric rings on the wings of squinting bush brown (*Bicyclus anynana*) (Monteiro, 2015). These colorful patterns have been shown to play important roles in mimicry in *Heliconius* and *Papilio* butterflies (Kunte et al., 2014; Lewis et al., 2019), in predator avoidance by either showing that the butterfly is aposematic and unpalatable or by distracting the predator to attack less vulnerable regions such as the wing margins (Martin et al., 2014; Prudic et al., 2015). Eyespots are involved in this latter process as well as in mate signaling during courtship displays (Robertson and Monteiro, 2005). Many butterfly wing patterns show also incredible diversity in between dorsal and ventral surfaces (Prakash and Monteiro, 2018); males and females (Prakash and Monteiro, 2016); and dry and wet seasonal forms (Monteiro et al., 2015).

Many genes involved in determining butterfly color patterns have been recently identified. Example includes *doublesex* which has been linked to the patterning of *Papilio polytes* (Kunte et al., 2014), *optix* which has been linked to the formation of red color pattern and mimicry in Heliconius butterflies (Reed et al., 2011), *spalt* and *BarH-1* involved in the wing pigmentation of *Pieris* and *Colias* butterflies (Stoehr et al., 2013; Woronik et al., 2018a; Woronik et al., 2018b), and *aristal-less* involved in the patterning of transversal bands in *Junonia coenia* (Martin and Reed, 2010).

### *Bicyclus anynana* as a model system

The African squinting bush brown butterfly *Bicyclus anynana* has remained at the forefront of butterfly evo-devo research for over three decades. *B. anynana* has a short life cycle, can be easily bred in the laboratory, has its whole

genome sequenced, and has established tools for genetic and molecular studies such as immunofluorescence, *in-situ* hybridization, CRISPR-Cas9, DNA knock-ins, and RNA interference. The main attraction towards *B. anynana* is the presence of a series of concentric colorful rings on the wings called eyespots.

**The eyespot of *Bicyclus anynana***

The eyespots are a trait novel to the nymphalid lineage of butterflies and have been intensively studied for over the past three decades, as eyespot formation is related to a variety of fundamental questions in biological sciences. Over a dozen of genes have been shown to be involved in the development of eyespots such as *distal-less* (Carroll et al., 1994; Monteiro et al., 2013), *spalt* (Brunetti et al., 2001; Monteiro et al., 2006), *engrailed/invected* (Keys et al., 1999; Monteiro et al., 2006), *antennapedia* (Matsuoka and Monteiro, 2019; Saenko et al., 2011), *notch* (Beldade and Peralta, 2017; Reed and Serfas, 2004), *ultrabithorax* (Matsuoka and Monteiro, 2019; Monteiro and Prudic, 2010; Weatherbee et al., 1999), *cubitus-interruptus* (Keys et al., 1999; Monteiro and Prudic, 2010), *ecdysone receptor* (Bhardwaj et al., 2018; Koch et al., 2003), *apterous* (Prakash and Monteiro, 2018), *wingless* (Monteiro et al., 2006; Özsu et al., 2017), *decapentaplegic* (Connahs et al., 2019), and *doublesex* (Prakash and Monteiro, 2019); and over 180 genes have been implicated to be involved in eyespot development (Özsu and Monteiro, 2017).

A reaction-diffusion mechanism has been proposed to be involved in setting up the center of the eyespot during larval wing development (Connahs et al.,

2019). By studying the gene expression patterns of Wnt (Armadillo: The signal transducer of Wnt signaling) and BMP (Dpp: the morphogen of BMP signaling) along with CRISPR-Cas9 data on *dll,* researchers have proposed that a three molecule substrate depletion reaction-diffusion system is involved in setting up the center of the eyespots during the fifth instar larval wing development. It has been shown that the BMP module acts as a substrate for the Wnt module which is expressed in the center of the eyespot together with Dll. In this study the simulation experiments accurately replicated the observed *dll* CRISPR mutant phenotypes. This study, however, provided no experimental evidence that the molecules involved in the reaction-diffusion system are interacting with each other.

A positional-information mechanism has been proposed to be involved in the formation of the rings of the eyespots (Nijhout, 1978; Beldade and Peralta, 2017). Fred Nijhout ((Nijhout, 1978; Nijhout, 1980) has proposed that a morphogen secreted from the center of the eyespots creates a concentration threshold around it. Different downstream genes such as *engrailed* (expressed in the orange ring) and *spalt* (expressed in the black scales) are activated in response to the concentration gradient (Brunetti et al., 2001). A recent study showing that inhibition of Wg by RNAi results in reduction of the overall size of the eyespot suggests that Wingless (Wg) is the likely morphogen (Özsu et al., 2017). The results, however, need further validation with respect to the downstream genes affected and if there are any other candidate morphogens involved in the differentiation of the rings of the eyespots.

Much research has also been carried out about the evolution of eyespots (Bhardwaj et al., 2020; Oliver et al., 2012; Oliver et al., 2014; Prakash and

Monteiro, 2018). By comparing the protein expression data of Engrailed/Invected, Spalt, Antennapedia, Notch and Distal-less in 21 nymphalid butterflies and mapping protein expression to a phylogenetic tree researchers have proposed that eyespot originated around 70 mya via co-option of a pre-existing network (Oliver et al., 2012; Kawahara et al., 2019). It has been proposed that eyespots originated from a simpler trait such as a spot on the wings of butterflies (Oliver et al., 2014). Furthermore, eyespots likely originated first in the ventral surface to deflect predators and gained expression later on the forewing and dorsal wing surfaces where eyespots play a potential novel role in sexual signaling (Huq et al., 2019; Oliver et al., 2014). Many of the genes studied for their association with eyespot evolution, however, have multiple gene duplicates. Gene duplicates, their involvement with the development of eyespots, and how they evolve over evolutionary time period has remained unexplored.

**Research questions**

My thesis aims at understanding two specific forms of patterning on the wings of butterflies, venation patterning and eyespot patterning.

Vein variation across species lies in their number and arrangement on the wing blade, from hundreds of mesh-like veins in the order Odonata (Damselflies and Dragonflies) to highly reduced veins in Diptera (such as *D. melanogaster*), a likely derived state (**Fig. 1.3**). The fossil record of *palaeodictyopterids* (Prokop and Ren, 2007) and *protorthoptera* (Kukalova-Peck, 1978) indicates that ancestral insects had a highly complex network of longitudinal and cross

veins (LVs and CVs) which was modified into a reduced and robust venation pattern in modern insects such as butterflies, moths, and flies with enhanced efficiency to sustain powered flight (Combes, 2003). Although venation patterns are extremely variable across insects, our understanding of the molecular mechanisms underlying these patterns has remained limited.

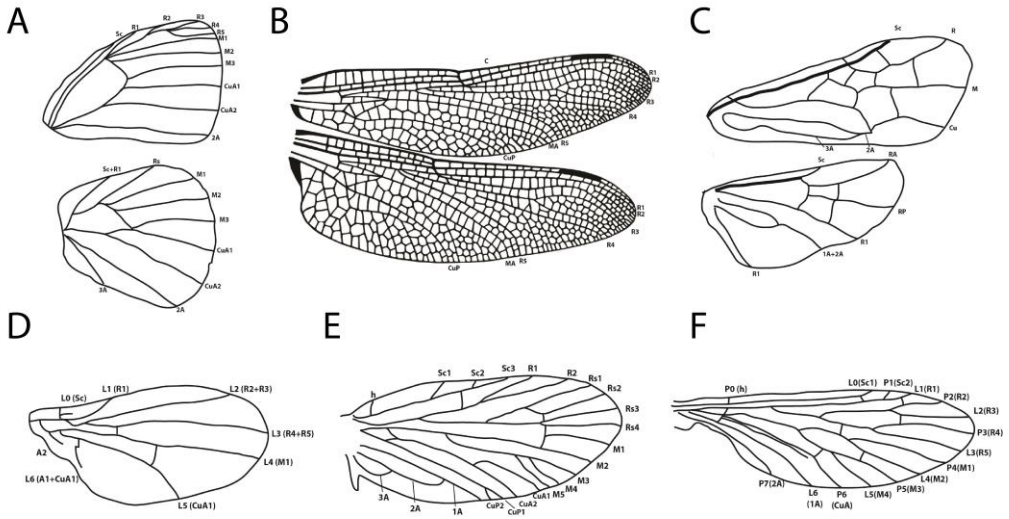



**Figure 1.4**: **Insect wing venation (based on Comstock-Needham venation system)** (Comstock, J.H. Needham, 1898). (**A**) *Bicyclus anynana* (Lepidoptera); (**B**) Dragonfly (Odonata); (**C**) *Athalia rosae* (Hymenoptera) (Redrawn from (Shimmi et al., 2014)); (**D**) *Drosophila melanogaster* (Diptera) (Redrawn from (Stark et al., 1999)); (**E**) Lepidopteran ancestral forewing venation ground plan (Redrawn from (Schachat and Brown, 2015)). (**F**) Ancestral Dipteran venation pattern (Redrawn from (Bier, 2000)) C: Costa, Sc: Subcosta, R: Radius, M: Media, MA: Media Anterior, Cu: Cubitus, CuA: Cubitus Anterior, CuP: Cubitus Posterior, A: Anal. P: Paraveins (indicate ancestral veins that are lost in modern Diptera).

The first main objective of my endeavour during this work was to understand the molecular mechanisms underlying these complexities using a system where we observe more complex venation. *B. anynana* with a more complex venation system compared to *D. melanogaster* and ample genetic and molecular biology tools at its disposal is a perfect candidate for such studies. I aimed at exploring the expression and function of ten orthologous genes

known to play role in venation patterning in *D. melanogaster* in chapter 2 and analysed what differences result in the appearance of a more complex venation in *B. anynana*.

The origin of novel traits is of great importance to our overall understanding of the process of evolution. Eyespots in butterflies are one of such novel marvels that has provided nymphalid butterflies higher chances of survival in nature compared to other species of butterflies. Several genes; and an array of mechanism and signaling pathways have been explored with respect to their involvement in eyespot patterning. However, there are large gaps in our understanding on how such a visually simple yet extremely complex (with respect to the molecular players involved) produces the magnificent rings of colors and have evolved over the past 70 mya.

The main objective of my next chapter (chapter 3) was to explore a possible mechanism that leads to the formation of the colorful rings of the butterflies. I aimed at exploring the expression and function of a few candidate genes which includes *optix* (which has been found to be expressed in the orange rings of butterflies), *spalt* (which is expressed in the black scale region of the eyespots) and one of the possible morphogens Dpp. I proposed how a system of positional information with Dpp as the source morphogen can activate genes such as *spalt* and *optix* based on different concentration thresholds and can hence lead to the formation of concentric rings of colors.

In my final data chapter (chapter 4) I was interested in exploring the evolution of the gene *engrailed (en)* and its association with eyespots. *en* has been proposed to be involved in a positional information system where *en* and its

paralogs are likely expressed in different regions of the eyespot. I aimed at finding the number of copies present in the genome of *B. anynana* and looked at the expression pattern of each of the *en*-like genes using *in-situ* hybridization. Comparison with similar copies of *en* in other lepidopteran species led to an interesting finding that these duplicates are not retained in the genome of butterflies due to their involvement with eyespot patterning. I also proposed a model where the network that is involved in a positional information system in venation patterning likely got co-opted in the eyespot centers which led to the expression of all the *en* paralogs in the eyespot center at the same time, and only later one of the copies gained expression in the novel orange ring.

## Chapter 2: Molecular mechanisms underlying simplification of venation patterning in holometabolous insects

Banerjee, T. Das, and Monteiro, A. (2020). Molecular mechanisms underlying simplification of venation patterns in holometabolous insects. Development dev.196394.

### Abstract

How mechanisms of pattern formation evolve has remained a central research theme in the field of evolutionary and developmental biology. An example of such a mechanism is the precise but varied way that insects pattern their wings with a network of veins. The mechanism of wing vein differentiation in *Drosophila* is a classic text-book example of pattern formation using a system of positional-information, yet very little is known about species with a different number of veins tweak this mechanism and how insect venation patterns evolved. Here, we examine the expression pattern of genes previously implicated in vein differentiation in *Drosophila* in two butterfly species with more complex venation, the African squinting bush brown *Bicyclus anynana* and the Asian cabbage white, *Pieris canidia*. We also test the function of these genes with CRISPR-Cas9 and drug inhibitors in *B. anynana*. We identify both conserved as well as new domains of *decapentaplegic (dpp), engrailed (en), invected (inv), spalt (sal), optix (opx), wingless (wg), armadillo (arm), blistered (bs),* and *rhomboid (rho)* gene expression in butterflies, and propose how the simplified venation in *Drosophila* might have evolved via loss of *dpp, sal* and *optix* gene expression

domains, silencing of vein inducing programs at Sal-expression boundaries, and changes in gene expression of vein maintenance genes.

## Introduction

Current venation patterns in several insect groups appear to be simplified versions of more complex ancestral patterns. The fossil record indicates that ancestral holometabolous insects, such as *Westphalomerope maryvonneae*, had highly complex vein arrangements which evolved into simpler venation with enhanced efficiency to sustain powered flight in modern representatives of Diptera and Lepidoptera (Nel et al., 2007). To identify these simplifications, Comstock and Needham in the 1900s developed a system of vein homologies across insects (**Fig S2.1 and S2.2**). The system nomenclature recognizes six longitudinal veins protruding from the base of the wings called Costa (C), Sub-costa (Sc), Radius (R), Media (M), Cubitus (Cu) and Anal (A)(Comstock and Needham, 1898). These veins can later branch into smaller veins, and additional complexity is added with cross-veins connecting two or more longitudinal veins. Every longitudinal vein across insects, however, can be identified using this nomenclature. Vein simplifications over the course of evolution have happened either via fusion of veins or disappearance of particular veins (Bier, 2000; De Celis and Diaz-Benjumea, 2003; Garcia-bellido and Celis, 1992; Stark et al., 1999), but the molecular mechanisms behind these simplifications remain unclear.

Molecular mechanisms of vein pattern formation have been primarily investigated in the model vinegar fly *Drosophila melanogaster,* where a

classic system of positional information takes place (**Fig. S2.3)**. Here, the wing is initially sub-divided into two domains of gene expression, an anterior compartment expressing *cubitus-interruptus* (*ci*), and a posterior compartment, expressing *engrailed (en)* and *invected (inv)*. In the posterior compartment *en* and *inv* activate the short-range morphogen *hedgehog* (*hh*) and restrict the activation of *ci* to the anterior compartment. In the anterior compartment *ci* encodes the protein involved in the transduction of Hh signaling (Cheng et al., 2014; Guillén et al., 1995). Hh diffusing to the anterior compartment establishes a central linear morphogen source of the protein Decapentaplegic (Dpp) at the posterior border of the anterior compartment, and genes like *aristaless* (*al*), *optix*, *spalt* (*sal*) and *optomotor-blind* (*omb*) respond to a Dpp morphogen gradient in a threshold-like manner, creating sharp boundaries of gene expression that provide precise positioning for the longitudinal veins (Barrio and De Celis, 2004; Martín et al., 2017; Sturtevant et al., 1997). Veins differentiate along these boundaries, along a parallel axis to the Dpp morphogen source via activation of vein-specific genes such as *knirps* (*kni*), *knirps-related* (*knil*), and *abrupt* (*abt*)(Blair, 2007; De Celis, 2003; Martín et al., 2017). Vein cell identity is later determined by the expression of genes such as *rhomboid (rho)*, downstream of the aforementioned genes (Guichard et al., 1999; Sturtevant et al., 1997). Conversely, intervein cells will later express *blistered* (*bs*) which suppresses vein development (Fristrom et al., 1994; Roch et al., 1998). The final vein positions are then determined by the cross-regulatory interaction of *rho* and *bs*.

The mechanisms underlying venation patterning in other insect lineages have remained poorly understood, and so far, gene expression patterns and functions for the few genes examined in beetles (order: Coleoptera), ants (order: Hymenoptera), and scuttle flies (order: Diptera) seem to be similar to those in *Drosophila* (Abouheif and Wray, 2002; Gantz, 2015; Tomoyasu et al., 2005; Wang et al., 2017).

Venation patterning in butterflies (order: Lepidoptera) has been examined in a few mutants in connection with alterations of color pattern development (Koch and Nijhout, 2002; Schachat and Brown, 2015) and more directly via the expression pattern of a few genes during larval development. Two of the species in which a few of the venation patterning genes have been studied in some detail are the African Squinting Bush Brown butterfly, *Bicyclus anynana* and the common Buckeye butterfly, *Junonia coenia*. In both species, En and/or Inv were localized in the posterior compartment using an antibody that recognizes the epitope common to both transcription factors (Keys et al., 1999; Monteiro et al., 2006). The transcript of *inv* in *Junonia*, however, appears to be absent from the most posterior part of the wings (Carroll et al., 1994), whereas the transcript of *hedgehog (hh*), a gene that is up-regulated by En/Inv in *Drosophila* (Tabata et al., 1992) is uniformly present in the posterior compartment of both species (Keys et al., 1999; Saenko et al., 2011). There is little knowledge of the expression domains of the other genes, including the main long-range morphogen *dpp* and its downstream targets (e.g., *sal, aristaless (al), optomotor blind (omb),* and *optix*) with respect to venation patterning. Studies on *optix* have been mostly focused on understanding the

mechanisms underlying color pattern development during the late pupal stages (e.g., its role in the development of red pattern elements in *Heliconius* butterflies)(Jiggins et al., 2017; Reed et al., 2011; Zhang et al., 2017a). Al has also been proposed to play a role in color pattern development in *J. coenia* and *Heliconius* butterflies, in particular in the coloring of transversal bands (Martin and Reed, 2010; Westerman et al., 2018). A recent report proposed the presence of a second *dpp*-like organizer at the far posterior compartment in butterflies (Abbasi and Marcus, 2017). This report, however, showed no direct gene expression or functional evidence and has been debated by other researchers (Lawrence et al., 2017). Currently there is also no functional evidence of altered venation for knock-out phenotypes for any of these genes in Lepidoptera.

In the present work, we explore the expression of an orthologous set of genes to those that are involved in setting up the veins in *Drosophila* in two butterfly species: *Bicyclus anynana* and *Pieris canidia*. We subsequently perform CRISPR-Cas9 and drug inhibition experiments to test the function of some of these genes in venation patterning in *B. anynana*.

## Materials and Methods

### Animal husbandry

*B. anynana* butterflies were reared at 27°C in 12:12 day: night cycle. The larvae were fed young corn leaves and the adults mashed bananas.

CRISPR-Cas9

Knock-out of *sal*, *optix* and *dpp* was carried out using a protocol described previously (Banerjee and Monteiro, 2018). Briefly, for *sal* and *dpp* single guides were designed targeting exon 1 of *sal* and *dpp*, while for *optix* two guides were designed targeting exon 1 of *optix* (see Supp file for sequence and primers **Table S2.1**). A total of 863 embryos for *sal*, 1973 embryos for *dpp*, and 1509 embryos for *optix* were injected with each containing 300 ng/µl of guide (for *optix* both guides were used at the same time) and 300 ng/µl of Cas9 protein (NEB; Cat. no.: M0641) mixed together in equal parts (total volume = 10 µl) with an added small amount of food dye (0.5 µl) (**Table S2.2-S2.4**). The hatchlings were transferred into plastic cups and fed young corn leaves. After pupation, each individual was assigned a separate emergence compartment (a plastic cup with lid). Once eclosed, the adults were frozen at -20°C and imaged under a Leica DMS1000 microscope using LAS v4.9 software. Descaling of the adult wings for imaging was done using 100% Clorox solution (Patil and Magdum, 2017). Mutant individuals were tested for insertions or deletions via an *in-vitro* endonuclease assay on the DNA isolated from the wings and then sequenced.

*In-situ* hybridization

Fifth instar larval wings were dissected based on a previously described protocol (Banerjee and Monteiro, 2020a) in cold PBS and transferred into 1X PBST supplemented with 4% formaldehyde for 30 mins. After fixation, the

wings were treated with 1.25 µl (20 mg/ml) proteinase-K (NEB, P8107S) in 1 ml 1X PBST and then with 2 mg/ml glycine in 1X PBST. Afterward, the wings were washed three times with 1X PBST, and the peripodial membrane was removed using fine forceps (Dumont, 11254-20) (in preparation for *in situ* hybridization). The wings were then gradually transferred into a pre-hybridization buffer (see **Table A1** for composition) by increasing the concentration in 1X PBST and incubated in the pre-hybridization buffer for one hour at 60°C. The wings were then incubated in hybridization buffer (see **Table A1** for composition) supplemented with 100ng/µl of probe at 60°C for 16-24 hrs. Subsequently, wings were washed five times with preheated pre-hybridization buffer at 60°C. The wings were then brought back to room temperature and transferred to 1X PBST by gradually increasing the concentration in the pre-hybridization buffer and they were later blocked in 1X PBST supplemented with 1% BSA for 1 hr. After blocking, wings were incubated in 1:3000 anti-digoxygenin labeled probe diluted in block buffer. To localize the regions of gene expression NBT/BCIP (Promega) in alkaline phosphatase buffer (See **Table A1** for composition) was used. The wings were then washed, mounted in 60% glycerol, and imaged under a Leica DMS1000 microscope using LAS v.4.9 software.

Immunostainings

Fifth instar larval wings were dissected based on a previously described protocol (Banerjee and Monteiro, 2020a) in cold PBS and immediately transferred into a fixation buffer supplemented with 4% formaldehyde (see

**Table A2** for composition) for 30 mins. Afterward, the wings were washed with 1X PBS and blocked for one to two days in block buffer (see **Table A2** for composition) at 4°C. Wings were incubated in primary antibodies against En/Inv (1:20, mouse 4F11, a gift from Nipam Patel, (Patel et al., 1989)), Sal (1:20000, guinea-pig Sal GP66.1, (Oliver et al., 2012)), Arm (1:1000, rat Arm), and Rho (1:1000, rabbit Rho) at 4°C for one day, washed with wash buffer (see **Table A2** for composition) and stained with secondary antibodies anti-mouse AF488 (Invitrogen, #A28175), anti-rat AF488 (Invitrogen, #A-11006), anti-rabbit AF488 (Invitrogen, #A-11008), and anti-guinea pig AF555 (Invitrogen, #A-21435) at the concentration 1:500. The wings were then washed in wash buffer, mounted on an in-house mounting media (see **Table A2** for composition), and imaged under an Olympus fv3000 confocal microscope, Zeiss Axio Imager M2.

Drug treatment and qPCR analysis

Fourth instar larval were injected with 1mM Dorsomorphin (BMP inhibitor) (Yu et al., 2008) and iCRT3 (Wnt inhibitor) (Lee et al., 2013) in between the second and third thoracic legs. DMSO (the solvent) was kept as control. Few of the individuals were dissected at late fifth instar larval wings based on the protocol described in (Banerjee and Monteiro, 2020a) for qPCR analysis in a Bio-Rad qPCR thermocycler (three individuals in each biological replicate) and the rest were allowed to develop till adulthood. A total of four biological replicates of larval wing were tested against expression of *spalt* with three technical replicates. FK506 and UBQL40 were kept as control (Arun et al.,

2015). Raw Cq data are provided in **Table S2.5**. Adult individuals were descaled in 100% Clorox solution (Patil and Magdum, 2017) and imaged under Leica DMS1000 microscope.

**Results**

Staging of the butterfly wings

Comparative analysis between two distantly related butterfly species, with different larval and pupal development times, as well as with *Drosophila*, with three larval instars instead of the five observed in butterflies requires an initial consideration of what might be comparable wing developmental stages for vein differentiation. Age of larvae (e.g., days since last molt) is a poor predictor of the development state of the wings in butterflies (Reed et al., 2007). Wing development of butterflies in the larval stage involves the growth and expansion of a flattened wing disc, with a ventral and dorsal layer of cells on each side of the disc. This arrangement of cells is unlike that of *Drosophila* where the wing blade during the larval stage has a single sheet of cells which later folds to become the dorsal and ventral surface. Wing discs in both *Bicyclus* and *Pieris* remain very small until the final (fifth) instar. The same is true for larval wing discs of *Drosophila* (Matsuda and Affolter, 2017). Based on the onset of the expression of multiple homologous genes, described below, we estimated that venation patterning starts during the final instar of both butterfly and fly wing discs. A previous study described the developmental stages of butterfly wing discs during the final instar using the pattern of tracheal growth as it invades the lacunae, the space between the dorsal and ventral epidermal layers, and zone of differentiation of the future veins (Reed et al., 2007). This study uses a numbering system from 0.00 (wing at the late

fourth to beginning of the fifth instar) to 4.00 (wings at the end of the fifth instar, the wandering stage, just before the pre-pupal stage) (Reed et al., 2007). Along with the tracheal invasion data, which starts at stage 1.00, we used the cellular arrangement and the shape of the wing margin as indicators of developmental stages prior to stage 1.00 (**Fig. 2.1**), as mentioned below. We focused most of our analyses on larval wings between developmental stages 0.00 and 1.00, the time we estimated venation patterning is taking place in lepidopteran wings (**Fig. 2.1**).

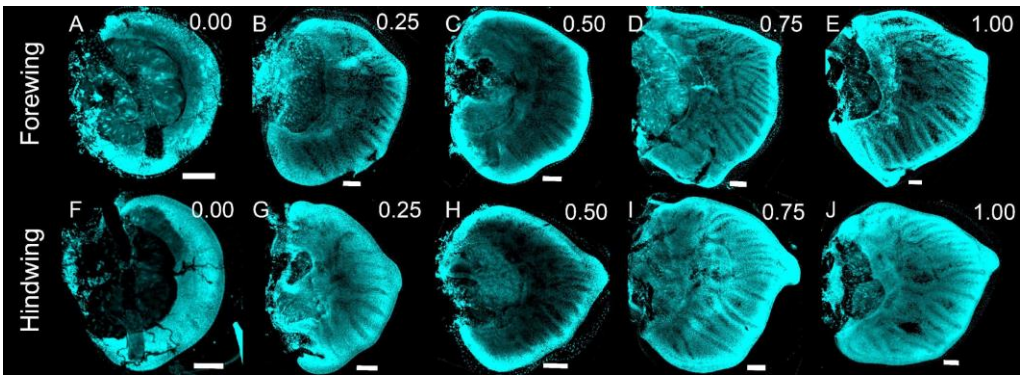



**Figure 2.1. DAPI staining and staging of early wing development of *Bicyclus anynana* based on Reed et al.** (Reed et al., 2007)**, and on cellular arrangements and wing shape changes.** (**A-E**) Forewings, (**F-J**) Hindwings. (**A, F**) Stage 0.00 is the earliest stage at which wing staining is possible at the end of fourth instar. The wing disc is crescent shaped, surrounding a mass of central cells, epidermal cells are homogenously distributed and it's difficult to distinguish forewings and hindwings based on morphology. (**B, G**) Stage 0.25 is when the wing starts to develop a slight bulge at the distal edge. Cell density is higher along the wing margin and lower along the position of the future veins. These areas, also called lacuna, represent areas through which the tracheal system will invade, in between the dorsal and ventral surfaces, and which will differentiate into veins. (**C, H**) Stage 0.50. Cell density increases in the wing margin and the fore and hindwings can be distinguished on the basis of their shape. (**D, I**) Stage 0.75. The wing is much larger compared to the previous stage and exhibits higher density of cells in the wing margin and flanking the lacuna. (**E, J**) Stage 1.00. The cells continue to condense around the veins/lacuna and tracheal invasion starts to happen. Scale bar: 100 µm.

Expression of *engrailed* and *invected* transcripts and proteins

We first examined the expression pattern of En and Inv at both the transcript and protein levels in *B. anynana* and in *P. canidia* fifth instar larval wings. We used an antibody (4F11) that recognizes the epitope of both proteins and confirmed that En and/or Inv expression is found throughout the posterior compartment in both forewings and hindwings in *B. anynana* (Keys et al., 1999; Patel et al., 1989), and also in *P. canidia.* However, a sharp drop in expression levels is observed posterior to the A2 lacuna in both species (**Fig. 2.2A-D, Fig S2.6E, G and I**). We will use "vein" instead of lacuna from here on, for simplicity, even though at this stage we only observe a pattern of longitudinal gaps between the two layers of wing epidermis and not proper vein tissue. We hypothesized that the low En/Inv posterior expression could either be due to lower levels of transcription or translation of En and/or Inv, or due to the absence of either of the two transcripts in the area posterior to the A2 vein. To test these hypotheses, we performed *in-situ* hybridization using probes specific to the transcripts of *en* and *inv* in *Bicyclus* (see Supp file for sequences). *en* was expressed homogeneously throughout the entire posterior compartment on both the forewing and the hindwing, but *inv* was restricted to the anterior ~70% of the posterior compartment (**Fig. 2.2E-H**). Hence, the low levels of En/Inv protein expression appear to be due to the absence of *inv* transcripts in the most posterior region of the posterior compartment.

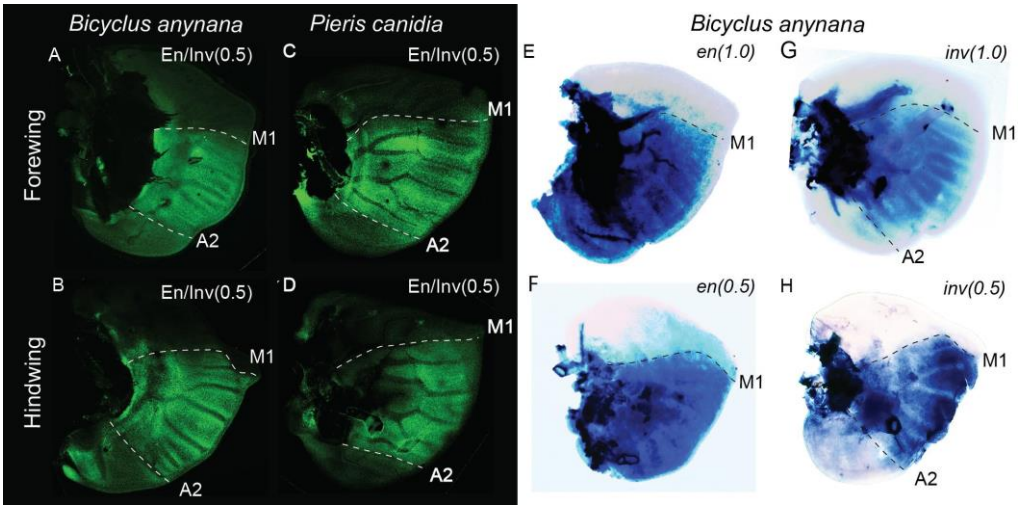



**Figure 2.2. Expression of Engrailed and Invected proteins in *Bicyclus anynana* and *Pieris canidia* and expression of mRNA transcripts in *Bicyclus anynana*. (A)** Expression of En/Inv proteins in the forewing, and **(B)** hindwing of *B. anynana* and **(C)** forewing and **(D)** hindwing of *P. canidia* is strong between the M1 and A2 veins, and weaker posterior to the A2 vein. **(E)** Expression of *en* mRNA transcripts in the forewing, and **(F)** hindwing in *B. anynana* is almost homogeneous across the posterior compartment. **(G)** Expression of *inv* in the forewing, and **(H)** hindwing in *B. anynana* is detected around 70% of the posterior compartment from the M1 vein to the A2 vein.

Expression and function of *dpp* (BMP signaling)

We explored the presence of transcripts for the BMP signaling ligand Decapentaplegic (Dpp) with the help of *in-situ* hybridization using a probe specific to its transcript (see Supp file for sequence). A band of *dpp* was observed along the A-P boundary (i.e., along the M1 vein) as previously reported in Connahs et al., (2019). However, another expression domain was observed in the lower posterior compartment around the A3 vein (**Fig. 2.3A; Fig. S2.3A-C**). Inhibiting BMP (Dpp) signaling via Dorsomorphin resulted in the reduction of overall wing size (**Fig. 2.3T; Fig. S2.3H**); and incomplete and ectopic development of veins, indicating a role for Dpp in wing growth and vein development (**Fig. 2.3U-W**). Furthermore, disrupting *dpp* via CRISPR-Cas9 resulted in ectopic and missing vein phenotypes emerging from the Cu2

and Cu1 veins (**Fig. 2.3X; Fig. S2.3K**). Inhibition of BMP signaling by Dorsomorphin also resulted in reduced transcription of *spalt* during larval wing development, a known downstream target of Dpp signaling in *Drosophila* wings (Blair, 2007; Szuperak et al., 2011)(**Fig. 2.3K**).

Expression and function of *spalt*

To localize the transcription factor Sal (only one *spalt* gene is present in *Bicyclus* (Nowell et al., 2017)) we performed immunostainings in larval wings of *B. anynana* and *P. canidia* using an antibody previously described (Stoehr et al., 2013). Sal is expressed in four clearly separated domains in both early (0.25) and later (1.00) stages of development (**Fig. 2.3B-G; Fig S2.5D, F and H**). The first domain is anterior to the Sc vein. The second domain spans the interval between the R2 and M3 veins. The third domain is between the Cu2 and A2 veins. No expression is observed in domain posterior to the A2 vein and finally, a fourth domain is present posterior to a boundary between the A2 and A3 veins (**Fig. 2.3J**). These expression domains are also observed in *P. canidia* (**Fig. 2.3H and I**).

To test the function of *sal* in vein development we targeted this gene using CRISPR-Cas9. The phenotypes observed support a role for *sal* in establishing vein boundaries at three of the four domains described above (**Fig. 2.3B-G**). We observed both ectopic and missing vein phenotypes in both the forewing and the hindwing at the domains where Sal protein was present during the larval stage of wing development (**Fig. 2.3N-S; Fig S2.4E-P**). In the forewing, *sal* crispants generated ectopic and loss of vein phenotypes between the R2

and the M3 vein domain (**Fig. 2.3N and P; Fig. S2.4G, K, M and O**) and ectopic veins between the Cu2 and A2 veins (**Fig. 2.3N and Q; Fig. S2.4I-O**). In the hindwing, we observed ectopic veins connecting to the existing Sc vein (**Fig. 2.3O and R; Fig S2.4P**), and missing veins between the Rs and M3 vein (**Fig. 2.3O and R; Fig S2.4J and L**).

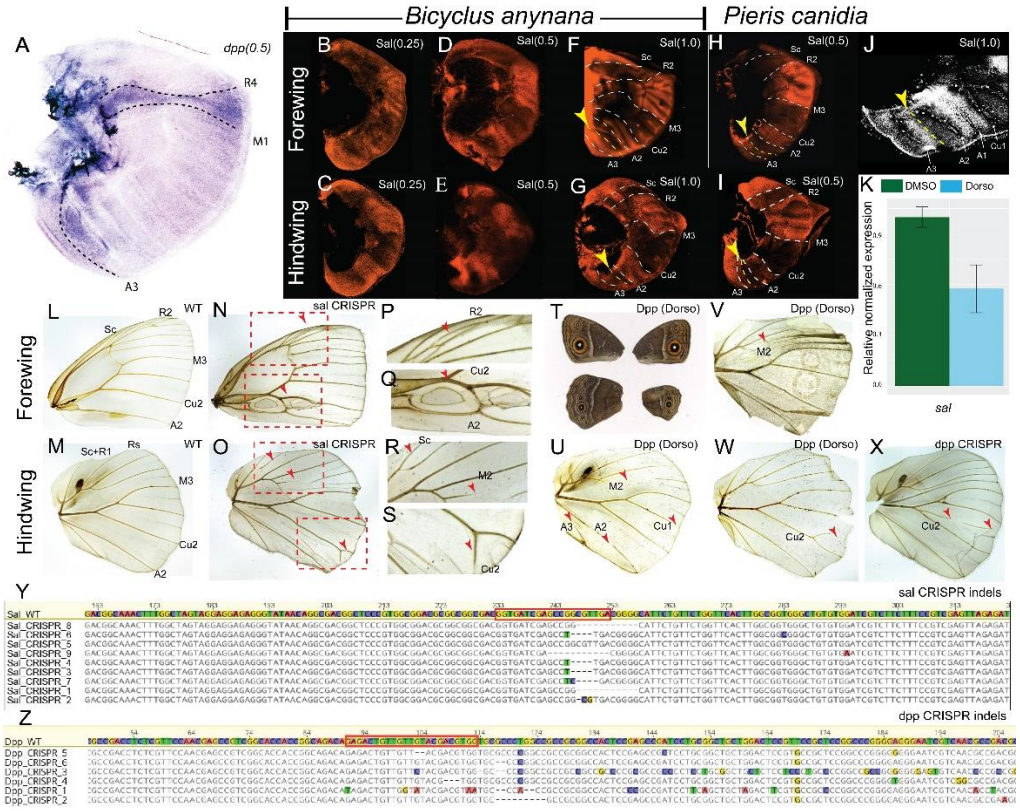



**Figure 2.3. Expression and function of *dpp* and *sal* in *Bicyclus anynana*, and expression of *sal* in *Pieris canidia*. (A)** Expression of *dpp* in the larval wing of *B. anynana* is visible anterior to the A-P boundary (M1) in between the R4 and M1 veins (for vein positioning at the same stage refer to **Fig. S5**) and around the A3 vein. Expression of Sal in forewings at stages **(B)** 0.25, **(D)** 0.5, and **(F)** 1.0, and in hindwings at stages **(C)** 0.25, **(E)** 0.5, and **(G)** 1.0, showing four distinct domains of expression: from the anterior margin to the Sc vein; from the R2 to M3 vein; from the Cu2 to A2 vein; and, from a boundary in between the A2 and A3 veins to the posterior wing margin. Expression of Sal in the larval **(H)** forewing and **(I)** hindwing of *P. canidia*; **(J)** Closeup of Sal expression showing the anterior boundary of the fourth Sal domain (yellow arrowhead and dotted yellow line). **(K)** *sal* qPCR on Dorsomorphin treated wings. The levels of *sal* transcripts are reduced due to BMP inhibition. **(L)** WT adult forewing and **(M)** hindwing. **(N, P and Q)** In

CRISPR-Cas9 *sal* crispants ectopic veins are produced within the boundaries of Sal expression in forewings and **(O, R and S)** ectopic, as well as missing veins, are produced within the Sal expression domains in hindwings. **(T)** Wings of an adult individual treated with Dorsomorphin showing reduction in hindwing size. Dorsomorphin treated wings fail to form complete veins such as **(U)** M2, Cu1, A2 and A3 veins, and **(V)** the M2 vein. **(W)** Dorsomorphin treatment also produces ectopic veins at Cu2. **(X)** A *dpp* crispant produced ectopic and missing vein phenotype at Cu2. Indels at the **(Y)** *sal* and **(Z)** *dpp* CRISPR-Cas9 cleavage sites (red boxes).

Expression and function of *optix*

Optix proteins are present in two domains during early larval wing development. Anteriorly to the R2 vein (Rs for hindwing) and posteriorly to the A2 vein (**Fig. 2.4; Fig. S2.5K, N, P, R, T and V**). Optix is also present in scale cells involved in orange pigment production in later (pupal) stages of wing development (**Fig. S2.6J**). Knocking out *optix* using CRISPR-Cas9 resulted in the loss of these orange scales (**Fig. S2.6E-I**), but no changes in venation were observed (**Fig. S2.6C and D**).

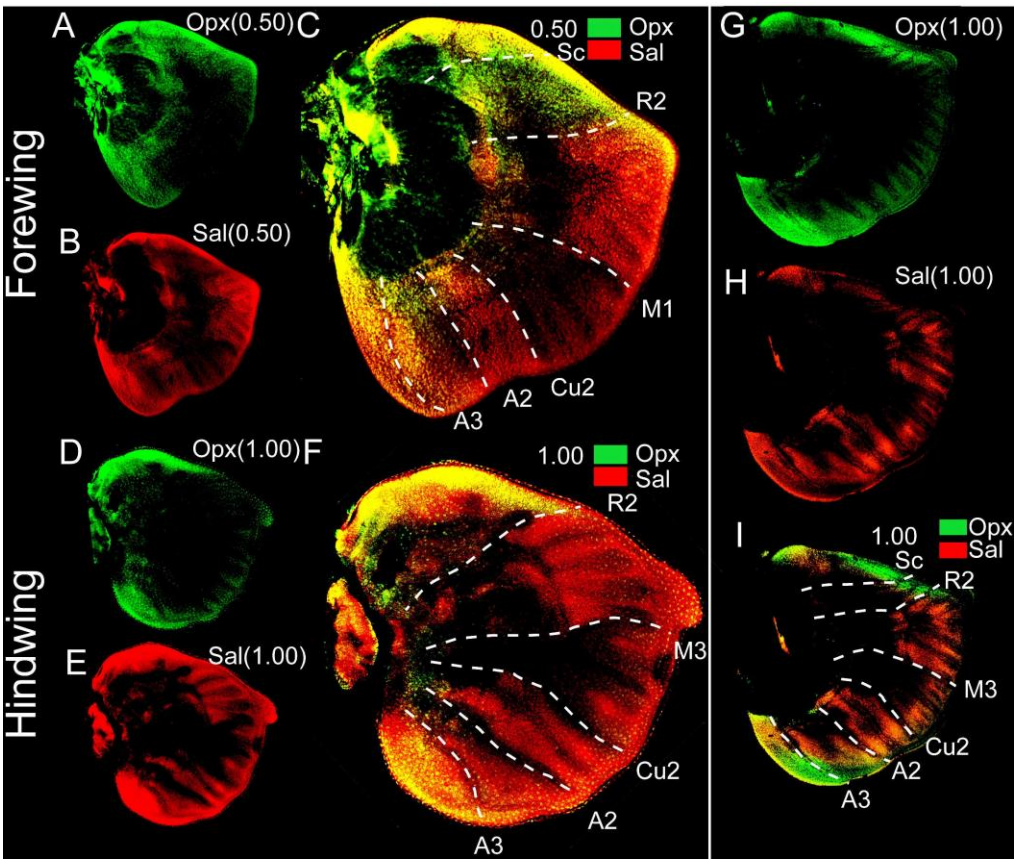



**Figure 2.4. Expression and Optix (Opx) in *Bicyclus anynana*.** Expression of Optix proteins in the **(A and G)** larval forewing and **(D)** hindwing. Optix expression is stronger in two regions: one anterior to the R2 vein and one posterior to the A2 vein (Stage: 0.5 and 1.0). **(B, E and H)** Expression of Spalt and **(C, F and I)** Co-expression of Sal and Opx.

Expression and function of *wg* (Wnt signaling)

During the early larval wing development (0.5) *wg* is expressed along the wing margin (**Fig. 2.5A and B; Fig S2.3D-F**). At an earlier stage (0.25), the signal transducer of Wg signaling, Armadillo (Arm), has a homogeneous expression across the wing disc (**Fig. 2.5C and D**). Later in development, at stage 1.0, Arm shows more focused expression in the wing margin, at the eyespots centers and mid-line connecting these centers to the wing margin, and along the veins (**Fig. 2.5E and F**). Inhibition of Wnt signaling via a small drug

inhibitor, iCRT3(Lee et al., 2013), led to the reduction of wing size, however, no defects in the veins were observed (**Fig. 2.5G-I; Fig. S2.3M and N**).

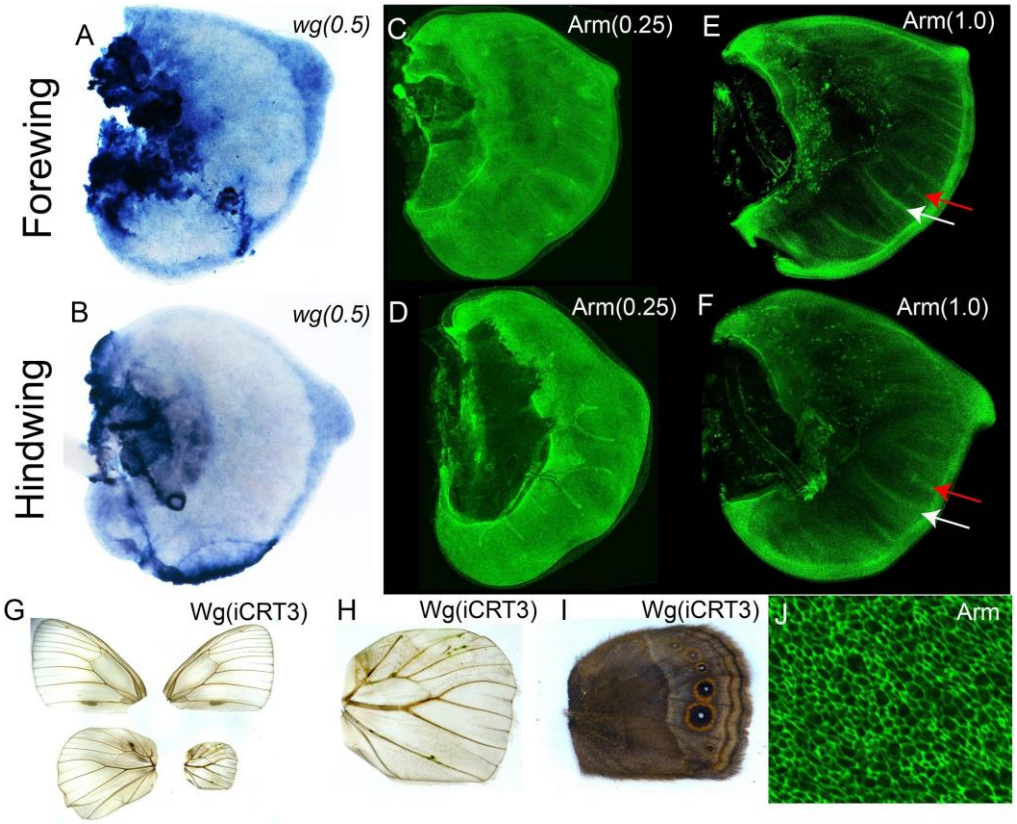



**Figure 2.5. Expression of *wingless* (*wg*), and Armadillo (Arm); and function of Wnt signaling in *Bicyclus anynana*. (A and B)** *wg* is expressed in the wing margin at the 0.5 stage. **(C-F)** Expression of Arm during different developmental stages. During **(C)** early forewing, and **(D)** early hindwing development Arm is more homogeneously expressed, while during later stages of development in both the **(E)** forewing and **(F)** hindwing expression is stronger in the wing margin, along the veins (white arrow) and in the eyespot centers (red arrow). **(G)** Individual treated with the Wnt inhibitor iCRT3 shows reduction in wing size but **(H)** no effect is observed in the venation. **(I)** Same wing as in H, before descaling. Eyespots became reduced in proportion to wing size due to Wnt inhibition. **(J)** Zoomed image of Arm expression in the larval wing at the 0.25 stage.

Expression of *blistered* (*bs*) and Rhomboid (Rho)

To localize vein and intervein cells, we performed *in-situ* hybridization of the intervein marker and vein suppressor gene *bs*, and antibody stains against the

vein marker and vein promotor gene Rho. In the larval forewing and hindwing of *B. anynana*, *bs* is expressed in the intervein cells and is absent in the vein cells and cells around the wing margin (**Fig. 2.6A-D, Fig. S2.5A-C**).

An antibody raised against *B. anynana* Rho showed expression along the veins and around the wing margin **(Fig. 2.6E, F, K, L and M)**. The expression of Rho is complementary to that of *bs*.

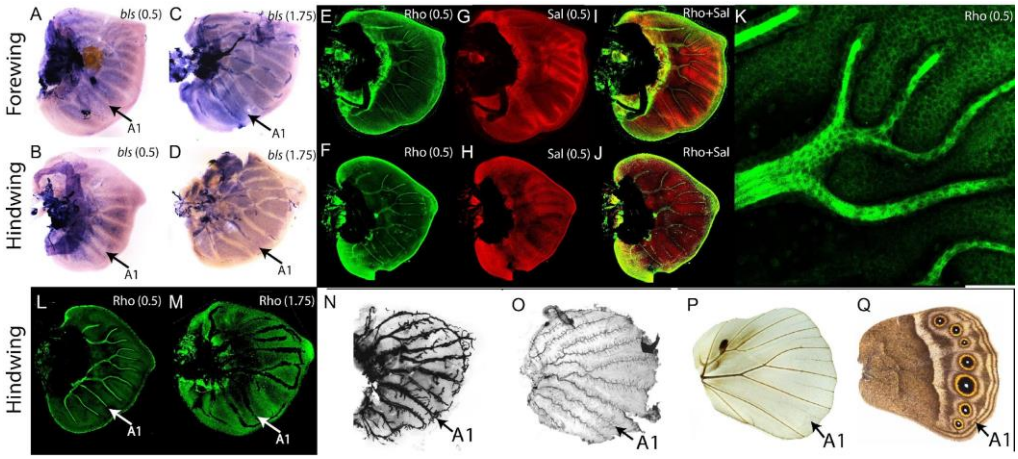



**Figure 2.6. Expression of *blistered* (*bs*) and Rhomboid (Rho); and loss of the A1 vein in *Bicyclus anynana*. (A)** Expression of *bs* in the intervein cells during early larval forewing and **(B)** hindwing (0.5). **(C and D)** During later stages of development *bs* expression is observed in the A1 vein, which will disappear in the pupal stage. **(E, F)** Expression of Rho in the forewing and hindwing. Rho is expressed along the veins and the wing margin. (**G, H**) Expression of Sal. (**I, J**) Co-staining of Rho and Sal. (**K**) High magnification image of Rho. Rho (a seven transmembrane protein receptor) is expressed in the plasma membrane. (**L**) Expression of Rho at stage 0.5 shows expression of Rho in the vein A1. (**M**) Expression of Rho at stage 1.75. The tracheal cells along the veins at this stage are auto fluorescent and are removed during image acquisition. Lower amount of signal is obtained from the region around the A1 vein. **(N-Q)** Disappearance of the A1 vein in *B. anynana* during the pupal to adult wing transition. **(N)** Larval Wing; **(O)** Pupal wing; **(P)** Adult wing with scales removed; **(Q)** Adult wing with scales. The A1 vein is observed in the larval and pupal stages, but it disappears in the adult stage.

## Discussion

A positional-information mechanism like that observed in *Drosophila* appears to be involved in positioning the veins in *B. anynana* and *P. canidia* but differences exist between flies and butterflies at multiple stages of vein patterning (**Fig. 2.7**). These differences are highlighted below.

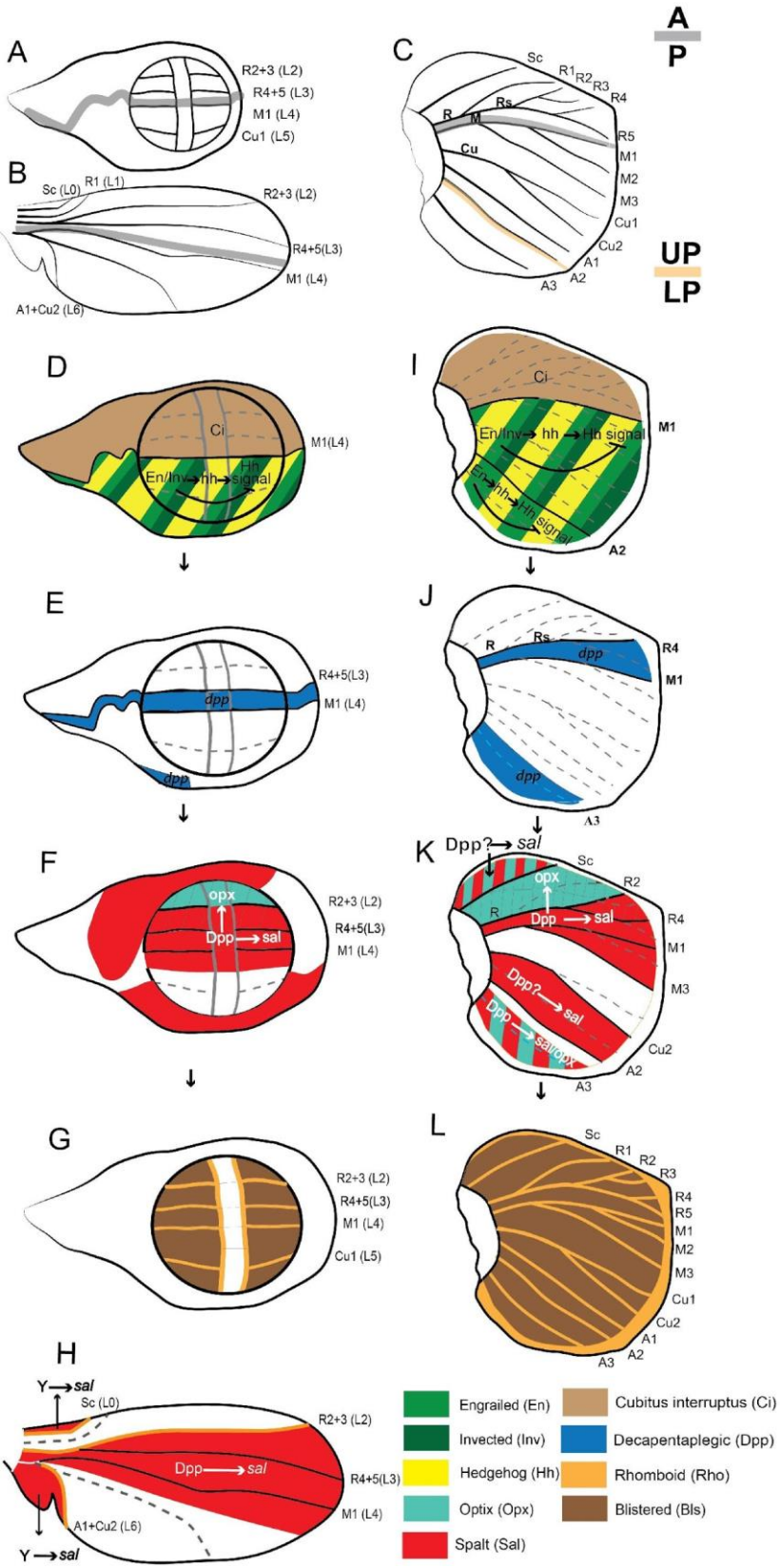



**Figure 2.7. A model for venation patterning in *Bicyclus anynana* and comparison with *Drosophila melanogaster*. (A)** Venation in *D. melanogaster* larval and **(B)** pupal wing; **(C)** Venation in *B. anynana* larval forewing; The Anterior-Posterior (A-P) boundary is marked by the thick grey line. The

boundary between the Upper-Posterior sector (UP: from the M1 to A2 vein) and the Lower-Posterior sector (LP: from the A2 vein till the posterior wing margin) of the wing is marked by the thick orange line. **(D-H)** Venation patterning in the *D. melanogaster* wing (for details refer to the main text). **(I)** We propose that in *B. anynana,* venation patterning is initiated by En and Inv expressed in the posterior compartment. En/Inv or En activates Hh in the posterior compartment, while suppressing Hh signaling. **(J)** A small amount of Hh diffuses into the anterior compartment where, due to the presence of the Hh signal transducer Ci, it activates *dpp* in a thin stripe of cells between the R4 and M1 veins. The boundary between *dpp*-expressing and non-expressing cells potentially sets up the position of the R4 and M1 veins (as it does in *Drosophila* **(E**)). A second domain of *dpp* is activated straddling the A3 vein via a Hh independent mechanism. **(K)** In *B. anynana*, Sal is expressed in four distinct domains, and Opx in two distinct domains in response to Dpp signaling. The three most anterior domains of Sal are involved in induction of veins at the domain boundaries (Sc, R2, M3, Cu2 and A2). **(L)** In *B. anynana*, as in *D. melanogaster* **(G)**, Rho is expressed along the veins and *bs* in the intervein cells.

**The early wings of *B. anynana* are subdivided into three gene expression domains instead of two as in *D. melanogaster***

One of the earlier steps in vein patterning in *D. melanogaster* is the separation of the wing blade into distinct compartments via the expression of En/Inv in the posterior compartment **(Fig. 2.7D)** (Guillén et al., 1995). *In-situ* hybridizations against the separate transcripts of *en* and *inv* in *B. anynana*, showed that *en* is expressed across the whole posterior compartment (in the whole region posterior to the M1 vein) and continues till the pupal stage (Banerjee et al., 2020), as in *D. melanogaster* (Blair, 1992), whereas *inv* is expressed only in the most anterior region of the posterior compartment, anterior to the A2 vein. This presumably leads to the higher En/Inv protein levels observed in the upper posterior compartment, and lower protein levels in the lower posterior compartment **(Fig. 2.7I)**. While the *en in-situ* results are new, the *inv* expression is consistent with that observed in a previous study of

*J. coenia* (Carroll et al., 1994). The *inv* expression pattern in butterflies is, thus, distinct from that of *D. melanogaster* where *inv* is expressed homogeneously throughout the posterior compartment (Cheng et al., 2014). These differences in expression of *en* and *inv* between *D. melanogaster* and *B. anynana* essentially set up two main domains of gene expression in fly wings but three in butterfly wings: An anterior domain with no *en* or *inv* expression, a middle domain with both *en* and *inv*, and a posterior domain with *en* but no *inv*.

**Two *dpp* signaling domains are established in *B. anynana* whereas a single domain is present in the wing pouch of *D. melanogaster***

The next step in venation patterning in *D. melanogaster* is the establishment of the main *dpp* organizer along a stripe of cells, in the middle of the wing pouch (Tanimoto et al., 2000). This group of *dpp*-expressing cells is established just anterior to the *en/inv* expressing cells, at the A-P boundary, where the M1+2 (L4) vein will differentiate (Ingham and Fietz, 1995; Tanimoto et al., 2000). The R4+5 (L3) vein differentiates at the anterior boundary of *dpp* expressing cells due to activation of Epidermal Growth Factor (EGF) signaling via the gene *vein* (*vn*) co-expressed along with *dpp* in response to Hh diffusing from the posterior compartment (Bier, 2000; Simcox et al., 1996) . In *B. anynana* we also observe a group of cells expressing *dpp* at the A-P boundary anterior to the M1 vein (**Fig. 2.7J**). This *dpp* expression in *B. anynana* is likely driven by Hh diffusing from the posterior compartment to the anterior compartment where *Cubitus interruptus* (Ci), the signal transducer of Hh signaling, is present (Keys et al., 1999; Saenko et al., 2011). We propose that the R4 vein

forms at the anterior boundary of *dpp* expressing cells, as it does in *D. melanogaster* (Bier, 2000). In *B. anynana* there is a second group of *dpp*-expressing cells straddling the A3 vein (**Fig. 2.7J**). This second *dpp* domain in *B. anynana* is probably activated via a Hh-independent mechanism, since no Ci or Patched (the receptor of Hh signaling) expression is observed in the posterior compartment around the A3 vein in butterflies (Keys et al., 1999; Saenko et al., 2011). In *D. melanogaster,* there is also a group of *dpp*-expressing cells outside the wing pouch, which are activated via a Hh-independent mechanism (Foronda et al., 2009) (**Fig. 2.7E**). These two groups of cells could be homologous.

**Four domains of Sal expression are established in *B. anynana*, whereas a single domain is present in *D. melanogaster* larval wing pouch**

The expression of *dpp* activates the next step in venation patterning in *D. melanogaster*, which involves the activation of *sal* expression some distance away from the signaling center in a single main central domain (Barrio and De Celis, 2004; Bier, 2000; Blair, 2007; Strigini and Cohen, 1999b). Here, the anterior boundary of Sal expression is involved in setting up the R2+3 (L2) vein(Bier, 2000; Blair, 2007; Cook et al., 2004; Sturtevant et al., 1997). In *B. anynana*, Sal is expressed in four separate domains in the larval wing, and functional data (discussed below) indicate that the boundaries delimiting the three most anterior Sal domains set up veins Sc, R2, M3, Cu2 and A2 (**Fig. 2.7K**). Only two of the Sal domains straddle the two *dpp* expression domains (**Fig. 2.7J and K**). This suggests that Dpp might be activating *sal* in two of the domains where *dpp* and Sal are co-expressed and overlap, but either Dpp or

some other morphogen, might activate *sal* in the first and third domains of Sal expression in *B. anynana*. We explored both the expression and function Wg as well as its signal transducer Armadillo (Arm), as possible activators of these additional Sal domains, as discussed in the sections below, but found no supporting evidence for this. In *D. melanogaster*, only one Sal central domain is present in the wing pouch during the larval stage, where venation patterning takes place (circle in **Fig. 2.7F**), but a more anterior and a more posterior Sal expression domain appear during the pupal stage (**Fig. 2.7H**) (Biehs et al., 1998; Grieder et al., 2009; Sturtevant et al., 1997). To our knowledge, no study has yet elucidated which gene drives the expression of Sal in these additional domains in *D. melanogaster* pupal wings.

**Two domains of Optix expression are observed in *B. anynana* while a single domain is present in *D. melanogaster***

In *B. anynana* we observe two expression domains of Optix, one in the upper anterior compartment and one in the lower posterior compartment **(Fig. 2.7K)**. In *D. melanogaster*, however, only one Optix expression domain is observed in the upper anterior compartment in response to low levels of Dpp secreted from the A-P boundary (Martín et al., 2017). In *B. anynana*, both Optix domains are also likely activated by Dpp. The upper anterior domain of Optix (**Fig. 2.7K**) likely responds to low levels of Dpp secreted from the A-P boundary (**Fig. 2.7J**), while the lower posterior domain of Optix (**Fig. 2.7K**) is likely activated by Dpp present around the A3 vein (**Fig. 2.7J**). It is interesting to note that the first and the fourth domains of Sal overlap with the anterior and the posterior domains of Optix, respectively, since Sal has been shown to

repress *optix* in its own domain straddling the A-P boundary in *D. melanogaster* (Martín et al., 2017) (**Fig. 2.7K**). Co-expression of Sal and Optix at the upper anterior and lower posterior compartment might be an ancestral state which got modified at the A-P domains in modern insects. Conserved (from larval stage) and novel expression domains of Optix during the pupal wing stage (Banerjee and Monteiro, in prep) are involved in the development of ommochrome pigments in different areas of the wing (**Fig. S2.7**).

**_sal_ crispants show that three Sal boundaries are involved in positioning veins in _B. anynana_, whereas a single Sal boundary performs this function in _D. melanogaster_**

Sal knockout phenotypes in *B. anynana* led to disruptions of veins in three out of the four Sal expression domains suggesting that, as in *D. melanogaster*, Sal is involved in setting up veins. *sal* crispants displayed: 1) ectopic Sc veins at the posterior boundary of the first Sal expression domain (**Fig. 2.3O and R; Fig. S2.3P**); 2) both ectopic and missing veins in the region of the second Sal domain straddling the A-P boundary, on both forewings and the hindwings, consistent with previous results on *Drosophila* (**Fig. 2.3N, P, O and R** (Organista and De Celis, 2013; Sturtevant et al., 1997)); and 3) ectopic veins in both the forewing and the hindwing in the region of the third Sal domain (**Fig. 2.3Q and S; Fig. S2.4I-O**). The final Sal expression domain in *Bicyclus* is present posterior to a boundary running in between the A2 and A3 vein (**Fig. 2.3J**), and we obtained no crispant with disruptions in veins in this area. Our data provide evidence, thus, that Sal boundaries of expression in domains 1, 2,

and 3, are involved in differentiating veins at those boundaries in *B. anynana*, whereas the boundary of the last Sal domain might not be used to position veins in the most posterior wing region (**Fig. 2.7K and L**).

The presence of both ectopic veins as well as disrupted veins in the domains of Sal expression in *Bicyclus* might be due to the disruption of the vein-intervein network in those regions. In *D. melanogaster*, ectopic and disrupted veins in *sal* knockout mutants lead to ectopic and missing *rho* expression (Sturtevant et al., 1997). A proposed mechanism for how these genes interact involves Sal, Opx, Aristaless (Al) and Knirps. In *D. melanogaster*, a single stripe of Knirps is present along the R2+3 (L2) vein in response to Dpp signaling. Dpp from the A-P margin activates Al throughout the anterior compartment. *sal* is activated in response to a high concentration of Dpp posterior to the R2+3 (L2) vein and *optix* is activated only anterior to the R2+3 (L2) vein in response to presence of Dpp but absence of Sal (Martín et al., 2017). These different expression domains create a perfect environment at the R2+3 (L2) vein where Al activates *knirps*, while Sal and Opx repress *knirps* expression (Martín et al., 2017). Knirps then activates further downstream genes such as *rho* that induces vein development (Lunde et al., 1998). Our *knirps* staining using *in-situ* hybridization and immunostaining (Kosman et al., 1998) didn't produced any positive result but, expression of Sal, Opx and Al (**Fig. S2.5K-Z**) indicate that a similar mechanism might be in place in *B. anynana*, where the absence of Sal and Opx at the R2 vein might lead to activation of a gene similar to *knirps* by Al present homogeneously in the anterior compartment (**Fig. S2.5X-Z**).

An alternative mechanism for how these genes interact involves Sal activating a hypothetical short-range diffusible protein in the intervein cells and at the same time inhibiting the intervein cells from responding to the signal (Bier, 2000; Sturtevant et al., 1997). A small amount of this diffusible protein moves towards the *sal*-negative cells activating vein inducing signals which include genes such as *rho* (Sturtevant et al., 1997). Knockout of *sal* in clones of cells within a *sal*-expressing domain will create novel or missing boundaries of Sal+ against Sal- cells and will result in ectopic or missing expression of *rho*, activating or inhibiting vein development.

In *B. anynana*, we observe Rho protein expression in the vein cells (**Fig. 2.6E, F, K, L and M**) and *bs* mRNA expression in the intervein cells (**Fig. 2.6A-D**). These expression domains are similar to those observed in *D. melanogaster* (Fristrom et al., 1994; Roch et al., 1998). This indicates that knocking out *sal* most likely results in ectopic or loss of Rho in the *B. anynana* wing, resulting in ectopic and disrupted vein phenotypes (**Fig. 2.3N-S; Fig. S2.4E-P**).

**Inhibition of Dpp signaling results in venation defects likely due to reduced Sal expression**

Inhibition of Dpp signaling using Dorsomorphin resulted in missing and ectopic veins, likely due to reduced Sal expression levels, along with overall reductions of wing size (**Fig. 2.3K and T-W**). Inhibition of Dpp in *Drosophila* has also resulted in similar phenotypes (Bosch et al., 2017). Dorsomorphin has been shown to block the phosphorylation of Mad (the signal transducer of Dpp) and to selectively inhibit BMP (Dpp) signaling (Yu et al., 2008). Injection of Dorsomorphin resulted in lower levels of *sal* gene expression in

whole larval wings (**Fig. 2.3K**). Sal, as discussed above, is necessary for proper positioning of veins. Lower levels of Sal likely lead to missing and ectopic veins in Dorsomorphin treated individuals (**Fig. 2.3U-W**). Knocking out Dpp, using CRISPR-Cas9, produced similar ectopic as well as incomplete vein phenotypes at the Cu1 and Cu2 veins (**Fig. 2.3X; Fig. S2.3J**).

**Wingless (Wg) signaling is not likely involved in the activation of the first and the third Sal domains in *Bicyclus***

To explore new ligands that might be involved in the activation of the first and third Sal domains we studied Wg signaling. *wg* is expressed in the wing margin throughout the fifth instar larval wing development in butterflies (**Fig. 2.5A and B; Fig S2.3D-F** (Martin and Reed, 2010; Martin and Reed, 2014)). Arm, however, is homogeneously expressed during the early larval wing development (**Fig. 2.5C and D**). During later stages Arm becomes expressed in the wing margin, along the veins, and in the eyespot centers (**Fig. 2.5E and F;** (Connahs et al., 2019)). The presence of Arm along the veins indicates that Wnt signaling might be involved in the maintenance of veins after they are set up. The inhibition of Wnt signaling during fourth instar development, using the drug iCRT3 (Lee et al., 2013), however, didn't produce venation defects (**Fig. 2.5G and H**). Wnt inhibition reduced wing size and eyespot size (proportionately) and led to a few defects in the wing margin (**Fig. 2.5G-I; Fig. S2.3M and N**). Similar reduction in wing size without venation defects has also been observed in *D. melanogaster* (Couso et al., 1994). Reduction in both wing size and eyespot size, in a disproportionate degree relative to wing size, has been observed in *B. anynana* after a *wg*-RNAi knock-down was

performed closer to the relevant stages of eyespot differentiation – at the end of the fifth instar and during the earliest stages of pupal development (Özsu et al., 2017). It is possible that the early injections of iCRT3, in fourth instar larvae, reduce wing size but have no subsequent effect on eyespot development, which develop via a reaction-diffusion mechanism during the fifth instar (Connahs et al., 2019).

**Loss of *sal, optix* and *dpp* expression domains likely led to venation simplification in *D. melanogaster***

Insect wing venation has simplified over the course of evolution, but it is unclear how exactly this simplification took place. Insect fossils from the Carboniferous period display many longitudinal veins in their wings compared to modern insects such as *D. melanogaster* or even *B. anynana* (Kukalova-Peck, 1978; Nel et al., 2007; Prokop and Ren, 2007). Many of the differences in venation remaining between *B. anynana* and *D. melanogaster* are due to the additional loss of veins in the posterior compartment in *D. melanogaster* (**Fig. 2.7A-C**). Sal expression domains and crispant phenotypes in *B. anynana* indicate that the third Sal expression domain, present in *B. anynana* but absent in *D. melanogaster*, is involved in the formation and arrangement of posterior veins Cu2 and A2 (**Fig. 2.7F and K**). In *D. melanogaster* there is partial development of the Cu2+A1 (L6) vein and there are no A2 and A3 veins (**Fig. 2.7B**). The partial and missing veins in the posterior compartment of *D. melanogaster* are likely due to the reduction of the third and loss of the fourth Sal expression domains (**Fig. 2.7H**). It is also interesting to note that only one Optix domain is present in the upper anterior compartment in *D. melanogaster*

(**Fig. 2.7F**), while in *B. anynana* we observe two domains, one in the upper anterior and one in the lower posterior compartment (**Fig. 2.7K**). The loss of the fourth Sal and second Optix domain was perhaps a consequence of the partial loss of the second *dpp* organizer (**Fig. 2.7E**), and the reduction of the third Sal domain in *D. melanogaster*. This reduced Sal domain was probably mediated by the delayed expression of a yet undiscovered organizer in this region (Y) that becomes activated only during the pupal stages in *D. melanogaster* (**Fig. 2.7H**).

**Simplification of venation is also achieved via silencing of vein inducing or vein maintenance mechanisms**

Vein number reduction via loss of *dpp*/*sal*/*optix* expression domains is one mechanism of vein reduction across evolution, but a different mechanism of vein reduction appears to take place downstream of the stable expression of these genes. For instance, in *B. anynana*, we observe the development of veins at both the boundaries of the second Sal domain (i.e., veins R2 and M3) (**Fig. 2.7K**), whereas in *D. melanogaster*, only cells abutting the anterior boundary of the homologous Sal expression domain activate the R2+3 (L2) vein (Sturtevant et al., 1997) (**Fig. 2.7F**). Vein activation proceeds via the activation of vein-inducing genes such as *knirps* and *rho*, which does not take place at the posterior boundary of Sal expression in *D. melanogaster* (**Fig. 2.7F and H**) (Sturtevant et al., 1997). In *B. anynana*, veins are also not being activated at the anterior boundary of the fourth Sal expression domain (in between the A2 and A3 veins) (**Fig. 2.7K**). It is still unclear why veins don't form at some boundaries of *sal* expression, but the paravein hypothesis

proposes that loss of a vein-inducing program at these boundaries, resulted in venation simplification in modern insects such as *D. melanogaster* (Bier, 2000).

Further venation simplification might be happening via disruptions of vein maintenance mechanisms, where vein induction is later followed by vein loss. In *D. melanogaster* the maintenance of vein identity involves the stable expression of Rho and the exclusion of Bs from vein cells throughout wing development (Blair, 2007; Fristrom et al., 1994). Disruptions to this mechanism, however, appear to be taking place at the A1 vein during *B. anynana* wing development (**Fig. 2.6N-Q**). The A1 vein is present during larval and early pupal wing development (**Fig. 2.6N and O**) but is absent in adults (**Fig. 2.6P and Q).** In *B. anynana bs* is absent and Rho is present at the A1 vein in young larval wing discs (**Fig. 2.6A, B, and L, Fig. S2.5A;** stage: 0.5). However, as the wing grows the expression of *bs* appears at the A1 vein, while Rho seems to disappear (**Fig. 2.6C, D, and M; Fig. S2.5B and C**; Stage: 1.75). The detection of Rho using immunofluorescence is difficult at stage 1.75 (**Fig. 2.6M**), when *bs* was initially detected in the A1 vein, since tracheal tissues along the veins are auto-fluorescent. No stainings were performed at later stages of development but early onset or the stable expression of *bs* at the A1 vein may result in the disappearance of this vein. It is unclear how the balance between Bs and presumably Rho expression is altered during development in the A1 veins of *B. anynana*, but such a mechanism is likely contributing to the loss of that vein in adults and could be contributing to vein loss, in general, across insects.

In conclusion, we have provided evidence for the presence of three main domains of gene expression in the early wings of butterflies, an anterior, middle, and posterior domain instead of two (anterior and posterior) domains, as observed in flies. We have found the presence of two *dpp* expression domains in butterfly wings. Furthermore, we have described and functionally characterized four domains of Sal expression and two domains of Optix expression in butterflies, whose boundaries map to the development of multiple longitudinal veins in these insects. Two of the Sal domains and both the Optix domains straddle the two *dpp* expression domains, and may be activated by a *dpp* gradient, but Dpp or a different and yet undiscovered ligand (or ligands) is activating the two other Sal domains. The data presented in this study supports a Positional-Information mechanism involved in venation patterning in Lepidoptera as that observed in Diptera. Moreover, the data provide support to the hypothesis of venation simplification in insects via loss of gene expression domains, silencing of vein inducing boundaries (Biehs et al., 1998; Bier, 2000), and disruptions to vein maintenance programs (Blair, 2007). However, the mechanisms proposed in this paper cannot explain every feature of insect venation. Insects with left-right wing differences in their longitudinal vein branching patterns, such and in the hemipteran *Orosanga japaonicus* (Yoshimoto and Kondo, 2012), and cross-vein patterns, such as in the hymenopteran *Athalia rosae* (Huang et al., 2018; Matsuda et al., 2013) and the odonate *Erythremis simplicicolis* (Hoffmann et al., 2018b), most likely pattern their wings using both Positional-Information as well as Reaction-Diffusion mechanisms. Future comparative gene expression studies along with

venation patterning modeling should continue to illuminate the evolution and diversity of venation patterning mechanisms in insects.

**Funding**

This work was supported by a Ministry of Education, Singapore grant MOE2015-T2-2-159, the National Research Foundation, Singapore grant NRF-NRFI05-2019-0006, and the Department of Biological Sciences LHK fund GL 710221. TDB was supported by a Yale-NUS scholarship.

**Authors contribution**

TDB designed and performed the experiments. TDB analyzed the results and wrote the manuscript with input from AM. Both the authors agreed to the final version of the manuscript.

**Acknowledgments**

We would like to thank Jocelyn Wee for rearing *Pieris canidia*, Kenneth McKenna and Fred Nijhout for initial discussions on the paper, Robert Reed for anti-Optix antibody, Arnaud Martin for anti-Aristaless antibody, Nipam Patel for the anti-Engrailed/Invected antibody; and, Timothy Saunders, Christopher Winkler, Jocelyn Wee, Suriya Narayanan Murugesan, Sofia Sigal-Passeck and the comments of three anonymous reviewers that improved the manuscript. We would like to thank DBS-CBIS confocal facility for access to the confocal microscopes. The authors declare no conflict of interest.

**Supplementary Materials**

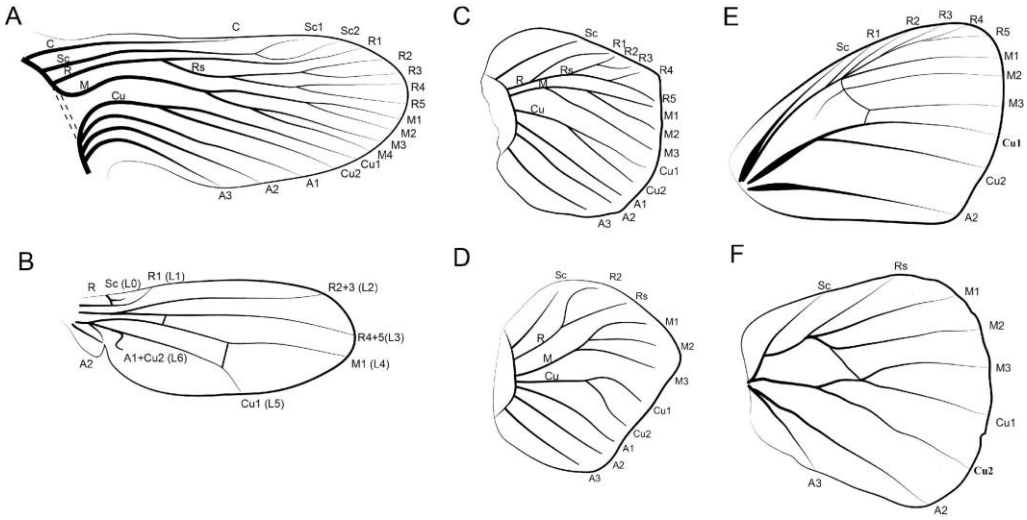



**Figure S2.1. Venation patterns in insects. (A)** Comstock-Needham hypothetic venation of primitive insects (redrawn from (Comstock and Needham, 1898)), **(B)** Wing venation of *Drosophila melanogaster* (redrawn from (Blair, 2007)), **(C)** Larval forewing venation and **(D)** hindwing venation of *Bicyclus anynana* butterflies. Larval wings of *B. anynana* were drawn based on methylene blue staining's (**Fig. S5**). **(E)** Adult forewing and **(F)** hindwing venation of *B. anynana*.

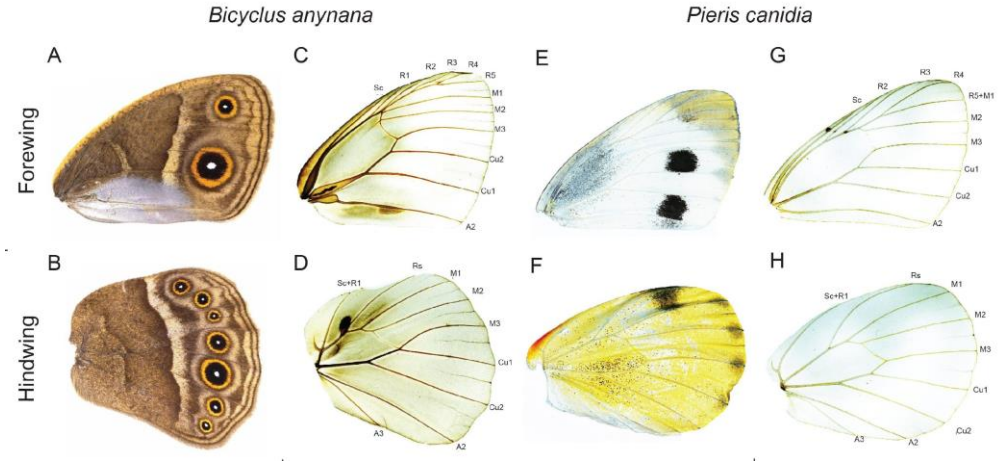



**Figure S2.2. Venation pattern in butterflies. (A and C)** *Bicyclus anynana* forewing, **(B and D)** and hindwing. **(E and G)** *Pieris canidia* forewing, **(F and H)** and hindwing.

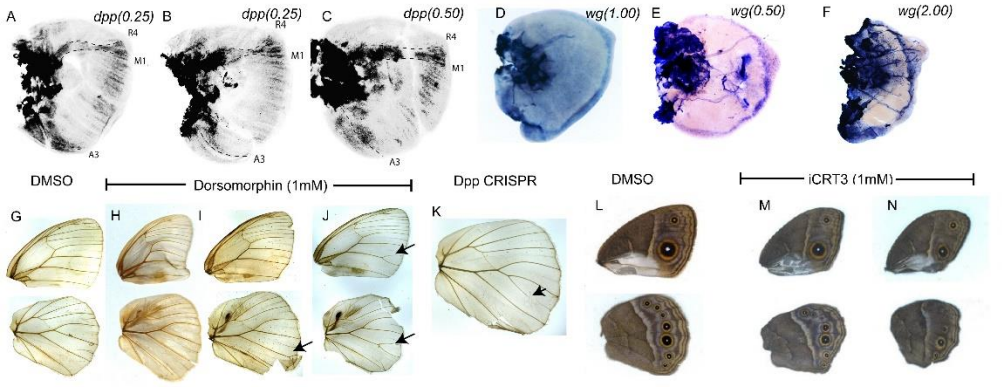



**Figure S2.3**. **Expression of *decapentaplegic* (*dpp*) and *wingless* (*wg*) and the effect of Dorsomorphin, iCRT3 and Dpp CRISPR on the wings of *Bicyclus anynana*.** (**A-C**) *dpp* is expressed in two domains in the larval wings. (**D-F**) Expression of wg in the wing margin. (**G**) Wings of a control individual injected with DMSO. (**H and J**) Dorsomorphin affects the wing size and venation (black arrow). (**K**) Dpp CRISPR individual with ectopic and missing vein (black arrow). (**L**) Wings of a control individual injected with DMSO. (**M and N**) iCRT3 injections reduce wing size relative to DMSO injections.

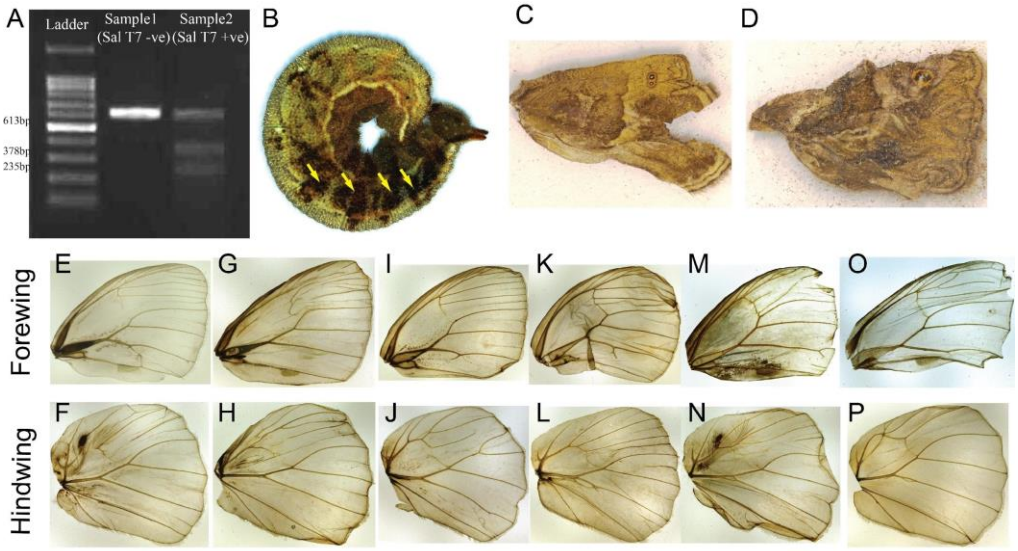



**Figure S2.4. Spalt CRISPR-Cas9 on *Bicyclus anynana* butterflies.** (**A**) T7 endonuclease assay on *sal* guide and Cas9 injected individuals. Sample 2 with T7 endonuclease added shows two shorter DNA bands indicating cleavage of the PCR product. (**B**) Pigmentation defects on the embryos. Spalt has been implicated to be involved in the development of black pigment on the eyespots of *B. anynana* butterflies. (**C and D**) Severe wing patterning defects in some individuals were observed. **(E-P)** Venation defects in *B. anynana* forewings and hindwings.

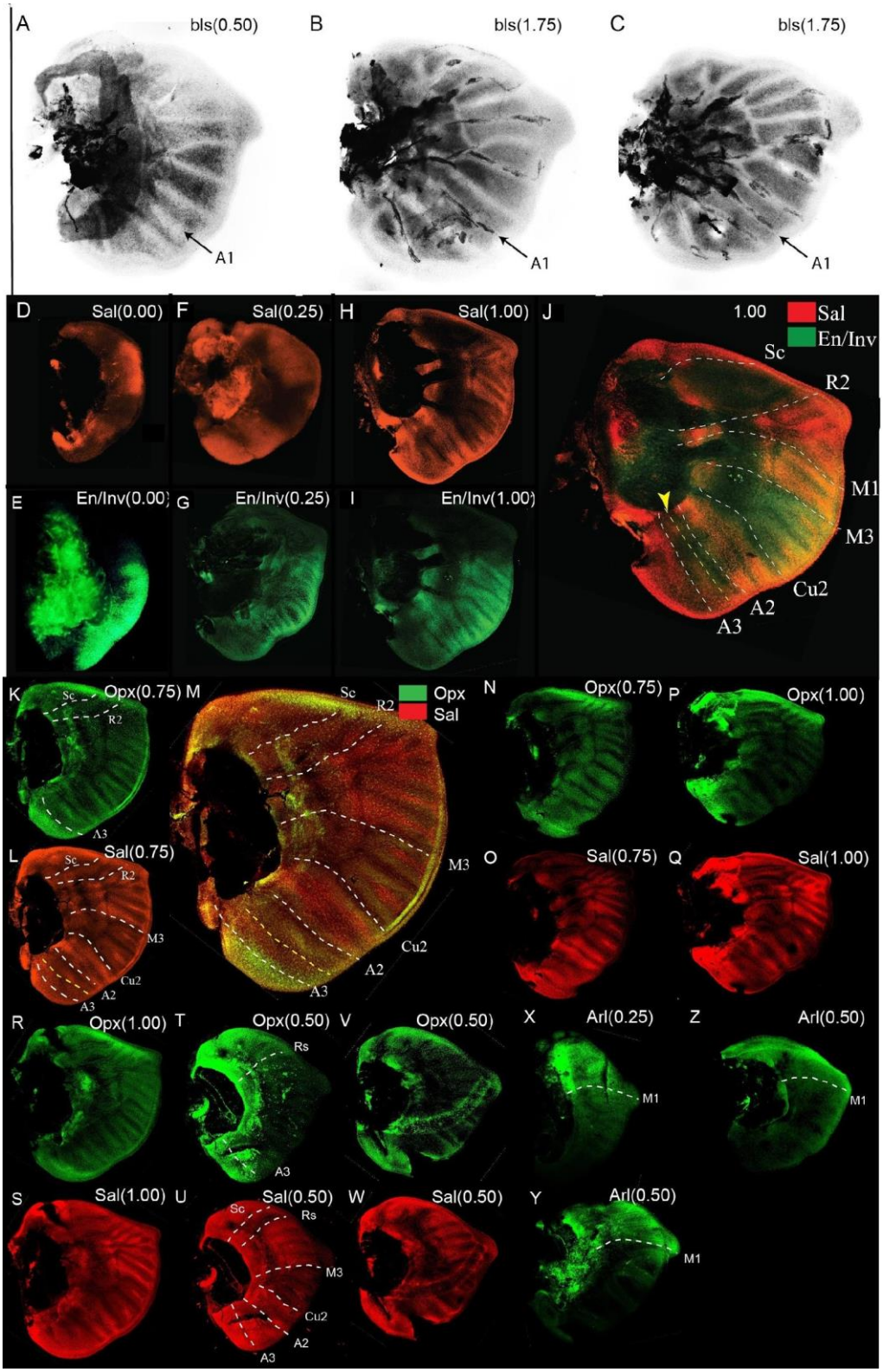



**Figure S2.5. Expression of *blistered* (*bs*), Spalt (Sal), Engrailed /Invected (En/Inv), Optix (Opx) and Aristaless (Al) in *Bicyclus anynana*.** (**A-C**) Expression of *bs*. *bs* is absent at the A1 vein at an early stage. However,

during later stages *bs* has a stronger expression at the A1 vein. (**D, F and H**) Sal staining at different stages of wing growth. (**E, G and I**) En/Inv staining at different stages of wing growth. (**J**) Co-staining of Sal and En/Inv. Opx expression at different stages of wing growth in the (**K, N, P and R**) forewing and (**T and V**) hindwing. Sal expression at different stages of wing growth in the (**L, O, Q and R**) forewing and (**U and W**) hindwing. **(M)** Co-staining of Opx and Sal. (**X-Z**) Al expression at different stages of wing growth.

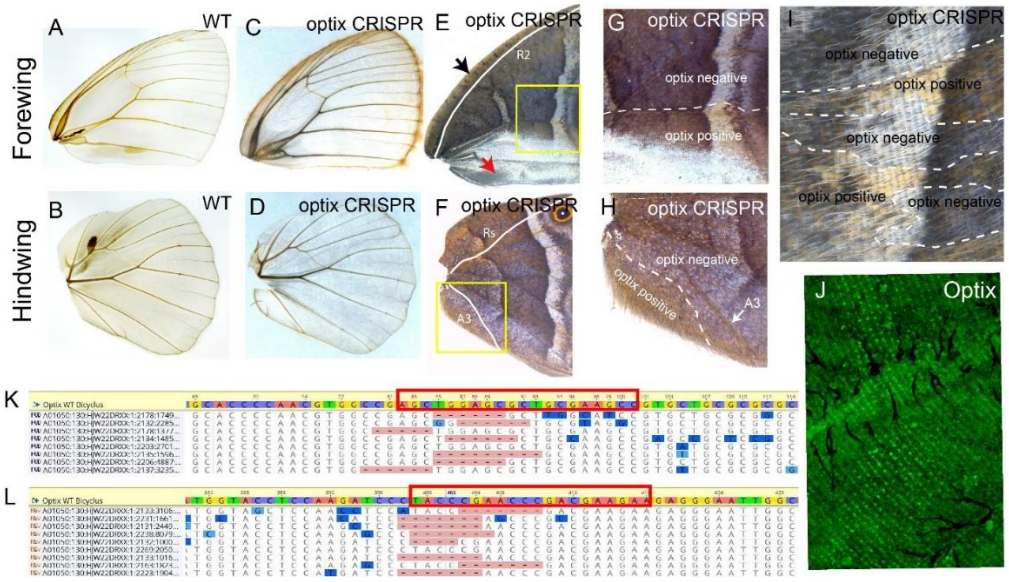



**Figure S2.6. Expression and function of Optix in *Bicyclus anynana*.** WT (**A**) forewing and (**B**) hindwing. Optix CRISPR (**C**) forewing and (**D**) hindwing. No defects in venation are observed. (**E-I**) Optix CRISPR individuals with loss of scales with ommochrome (orange) pigment. (**J**) Optix expression in the pupal wings. (**K and L**) Deletions in the regions targeted for *optix* CRISPR. Black arrow: orange scales in the anterior margin of the forewing overlap the anterior expression of Optix in the larval wing disc. Red arrow: silver scales in the posterior region of the forewing overlap the posterior expression of Optix in the larval wing disc.

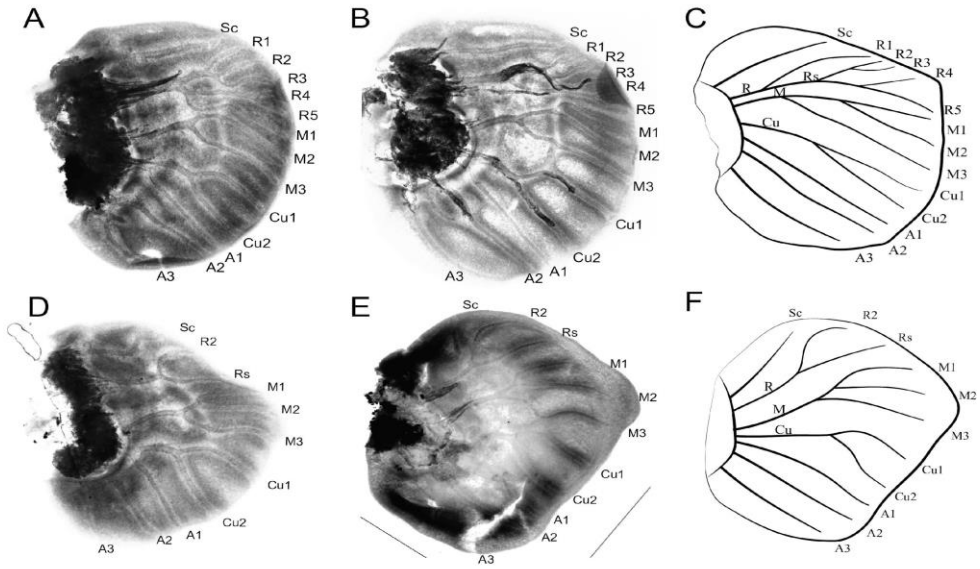



**Figure S2.7. Methylene blue staining of *Bicyclus anynana* larval wings. (A and C)** Forewing stained with methylene blue; **(B and D)** Hindwing stained with methylene blue; Illustration of (**E**) forewing and (**F**) hindwing venation.

**Table S2.1. Primer table**

| Sl. No. | Primer Name | Sequence | Description |
|---|---|---|---|
| **1.** | Dpp_insitu_F | GTTCTTCAACGTAAGCGGCG | Forward primer to amplify *dpp* for in-situ hybridization |
| **2.** | Dpp_insitu_R | CCACAGCCTACCACCATCAT | Reverse primer to amplify *dpp* for in-situ hybridization |
| **3.** | En_insitu_F | TTGAAGACCGTTGCAGTCC | Forward primer to amplify *en* for in-situ hybridization |
| **4.** | En_insitu_R | TAGATTGCTGTTCCCGCTTT | Reverse primer to amplify *en* for in-situ hybridization |
| **5.** | Inv_insitu_F | GGACCAAAGTGACGAAGAGC | Forward primer to amplify *inv* for in-situ hybridization |
| **6.** | Inv_insitu_R | TCCGGCACTCTAGCCTCTAC | Reverse primer to amplify *inv* for in-situ hybridization |

| | | | |
|---|---|---|---|
| **7.** | Bs_insitu_F | CTGACCGGCACCCAAGTGAT | Forward primer to amplify *bs* for in-situ hybridization |
| **8.** | Bs_insitu_R | CGTTGCGGGTGGTGAGACAT | Reverse primer to amplify *bs* for in-situ hybridization |
| **9.** | Sal_CRISPR_Seq_F | GCATCGACAAGATGCTGAAA | Forward primer to amplify *sal* for CRISPR-Cas9 invitro cleavage assay |
| **10.** | Sal_CRISPR_Seq_R | TTCATTTAGGGACGGTGGAG | Reverse primer to amplify *sal* for CRISPR-Cas9 invitro cleavage assay |
| **11.** | Sal_CRISPR_Guide | GAAATTAATACGACTCACTATAGGTGATCGAGCCGGCGTTGAGTTTTAGAGCTAGAAATAGC | Forward primer for guide synthesis to knockout *sal* |
| **12.** | Optix_CRISPR_Guide_1 | GAAATTAATACGACTCACTATAGGGGCTTCGCAGCGCTCCAGCTGTTTTAGAGCTAGAAATAGC | Forward primer 1 for guide synthesis to knockout *optix* |
| **13.** | Optix_CRISPR_Guide_2 | GAAATTAATACGACTCACTATAGGTTCTTCGTCGGGTTCGGGTAGTTTTAGAGCTAGAAATAGC | Forward primer 2 for guide synthesis to knockout *optix* |
| **14.** | Dpp_CRISPR_Guide | GAAATTAATACGACTCACTATAGGGAGACTGTTGTTGTACGACGTGGGTTTTAGAGCTAGAAATAGC | Forward primer for guide synthesis to knockout *dpp* |
| **15.** | CRISPR_Guide_R | AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTG CTATTTCTAGCTCTAAAAC | Reverse primer for guide synthesis CRISPR guides |

| | | | |
|---|---|---|---|
| **16.** | Wg_insitu_F | CAGCAGCTGGATTTTGTCAG | Forward primer to amplify *wg* for in-situ hybridization |
| **17.** | Wg_insitu_R | TATTGTGCCGTTGTCATCGT | Reverse primer to amplify *wg* for in-situ hybridization |
| **18.** | Sal_qPCR_F | TGTATGCCATCGCGTATTGT | Forward primer to amplify *sal* for qPCR |
| **19.** | Sal_qPCR_R | TAGTGGTAAACGCACGACCA | Reverse primer to amplify *sal* for qPCR |
| **20.** | FK506_qPCR_F | AAACTAACCTGCAGCCCTGA | Forward primer to amplify FK506 for qPCR |
| **21.** | FK506_qPCR_R | CAAGACGGAGAAGTTCCACA | Reverse primer to amplify FK506 for qPCR |
| **22.** | UBQL40_qPCR_F | CGGTAAACAATTGGAAGATGG | Forward primer to amplify UBQL40 for qPCR |
| **23.** | UBQL40_qPCR_R | CGAAGTCTGAGGACAAGATGC | Reverse primer to amplify UBQL40 for qPCR |

**Table S2.2. Spalt CRISPR-Cas9 injection table**

| Sl. No. | Concentration | Date | Eggs Injected | Hatchlings | % Hatchlings |
|---|---|---|---|---|---|
| **1.** | 300 ng/µl | 28th Sept 2018 | 302 | 48 | 15.9 |
| **2.** | 300 ng/µl | 10th Oct 2018 | 306 | 25 | 8.2 |
| **3.** | 300 ng/µl | 11th Nov 2018 | 120 | 18 | 15.0 |
| **4.** | 300 ng/µl | 9th Feb 2019 | 135 | 8 | 5.9 |

**Table S2.3. Optix CRISPR-Cas9 injection table**

| Sl. No. | Concentration | Date | Eggs Injected | Hatchlings | % Hatchlings |
|---|---|---|---|---|---|

| | | | | | |
|---|---|---|---|---|---|
| **1.** | 300 ng/µl | 11th March 2020 | 785 | 85 | 10.8 |
| **2.** | 300 ng/µl | 12th March 2020 | 398 | 47 | 11.9 |
| **3.** | 300 ng/µl | 6th June 2020 | 326 | 65 | 19.9 |

**Table S2.4. Dpp CRISPR injection table**

| Sl. No. | Concentration | Date | Eggs Injected | Hatchlings | % Hatchlings |
|---|---|---|---|---|---|
| **1.** | 300 ng/µl | 23rd Jan 2020 | 623 | 89 | 14.3 |
| **2.** | 300 ng/µl | 3rd Mar 2020 | 923 | 117 | 12.7 |
| **3.** | 300 ng/µl | 11th Mar 2020 | 427 | 64 | 14.9 |

**Table S2.5. Raw Cq data on the Dorsomorphin and DMSO treated samples**

| **Biological replicates** | **Raw Cq (Dorsomorphin treated)** | | | **Raw Cq (DMSO treated)** | | |
|---|---|---|---|---|---|---|
| | *spalt* | FK506 | UBQL40 | *spalt* | FK506 | UBQL40 |
| 1 (8th July 2019) | 29.09 | 24.16 | 21.68 | 26.56 | 22.76 | 20.91 |
| | 28.96 | 24.18 | 21.65 | 26.34 | 22.76 | 20.91 |
| | - | 24.09 | 21.62 | - | 22.84 | 20.82 |
| 2 (5th July 2019) | 28.36 | 22.50 | 20.89 | 28.10 | 22.69 | 21.05 |
| | 28.36 | 22.61 | 20.96 | 28.05 | 22.55 | 20.94 |
| | 28.20 | 22.54 | 20.76 | 27.76 | 22.67 | 20.94 |
| 3 (14th April 2019) | 31.54 | 22.57 | 20.89 | 30.43 | 22.22 | 20.64 |
| | 31.80 | 22.56 | 20.90 | 30.59 | 22.16 | 20.47 |
| | 31.23 | 22.50 | 20.87 | 30.58 | 22.19 | 20.67 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 4 (21st May 2019) | 32.02 | 22.76 | 21.15 | 30.40 | 22.23 | 20.60 |
| | 31.67 | 22.72 | 21.11 | 30.55 | 22.16 | 20.44 |
| | 31.39 | 22.62 | 21.00 | 30.32 | 22.29 | 20.72 |

Peptides used for antibody development (Highlighted in green)

Spalt

MPRVKPACVRRVSIGESSGSCSEEDVGNAMPDEARDRPEAHMCPRCQ
EQFENLHDFLYHKRLCDEKAMQMGEERMHSDPEDMVVSGDEEMDG
PNKRLEQVRRHRQDAENNNSLEDGEAEIPEADMPPVGLPFPLAGHVTL
EALQNTRVAVAQFAATAMANNANNEAAIQELQVLHNTLYTLQSQQV
FQLQLIRQLQNQLSLTRRKEDDPHSPPPSEPEQNAPSTPARSPSPPRPPR
EPSPVIPSPPTSQSLPSTHTHHTPKTEQISIPKIPTSSPSLMTHPLYSSISSSL
ASSIITNNDPPPSLNEPNTLEMLQKRAQEVLDNASQGLLANNLADELA
FRKSGKMSPYDGKSGGRNEPFFKHRCRYCGKVFGSDSALQIHIRSHTG
ERPFKCNVCGSRFTTKGNLKVHFQRHTSKFPHVKMNPNPVPEHLDKY
HPPLLAQLSPGPIPGMPPHPLQFPPGAPAPFPPNLPLYRPPHHDLLPPRP
LGDKPLSHHPLFAMREEQDAPADLSKPSAPSPPRPASDIFKSEPQDEES
QRDSSFEETDRISPKREIEDNDIGQDAEQDRYPSTSPYDDCSMDSKYSN
EDQIGRDSPHVKPDPDQPENLSSKTSSISGPISIATGLRTFPSFPLFPHSPP
SSVSSGSLTPFHHHPNSTMDSALTRDPLFYNAILPRPGSNDNSWESLIEI
TKTSETTKLQQLVDNINNKVSDPNECIVCHRVLSCKSALQMHYRTHTG
ERPFRCKLCGRAFTTKGNLKTHMGVHRIKPPSQLLHQCPVCHKKFSDP
SMLHQHIRLHTGERNNVFFNQFHDNEINSQSLPGSDVTEYNSFHSIPPPI
FPTPSTPGDRRADSRGTDDESGRDEREPATREFDDEPDIKDRRTSPLSV
CASASEFEVKTITTTASLPSATGSESGRSARGSPPSPSPSPSALSTPPRLP
HHSPLPSPPTPLAALGALGGSPFSPLGLAFPPAVRGNTTCTICYKTFACN
SALEIHYRSHTKERPFKCTVCDRGFSTKSSGGGCQCGRRARAPRPPHA
TALDLWNAFVYPGNMKQHMLTHKIRDMPPGFDKGPGGPSGPPSEEGR
DPSPDRRSSPEKLDLKRSPPVHPPPPMSHPPIDMPPLPKRPTVPSIPSHPP
PSASSKHLCGVCRKNFSSSSALQIHMRTHTGDKPFRCAVCQKAFTTKG
NLKGLLLPATRLISRSTNQATALFGTLGPFIYRLSELYAPPSATSALRLV
ELSDFGSADFR

Armadillo

MSYQIPSSQSRTMSHSNYGGSDVPMAPSKEQQTLMWQQNSYLVDSGI
NSGAATQVPSLTGKEDDEMEGDQLMFDLDQGFAQGFTQEQVDDMNQ
QLSQTRSQRVRAAMFPETLEEGIEIPSTQLDPAQPTAVQRLSEPSQMLK

HAVVNLINYQDDADLATRAIPELIKLLNDEDQVVVSQAAMMVHQLSK
KEASRHAIMNSPQMVAALVRAISNSNDLETTKGAVGTLHNLSHHRQG
LLAIFKSGGIPALVKLLSSPVESVLFYAITTLHNLLLHQDGSKMAVRLA
GGLQKMVALLQRNNVKFLAIVTDCLQILAYGNQESKLIILASQGPIELV
RIMRSFDYEKLLWTTSRVLKVLSVCSSNKPAIVEAGGMQALAMHLGN
PSGRLVQNCLWTLRNLSDAATKVEGLEGLLQSLVQVLASTDVNIVTC
AAGILSNLTCNNQRNKVTVCQAGGVDALVRTVVSAGDREEITEPAVC
ALRHLTSRHVESEMAQNAVRLHYGLPVIVKLLQPPSRWPLVKAVVGL
VRNLALCPANHAPLREHGAVHHLVRLLLRAFNDTQRQRGSVSGGGG
AGGAYADGVRMEEIVEGAVGALHILAREGLNRALIRQQNVIPIFVQLL
FNEIENIQRVAAGVLCELAADKEGAEMIEAEGATAPLTELLHSRNEGV
ATYAAAVLFRMSEDKPHDYKKRLSMELTNSLFRDDHQMWPNDLAM
QPDLQDMLGPEQGYEGLYGTRPSFHQQGYDQIPIDSMQGLEIGSGFGM
DMDIGEADGGGAASADLAFPEPPLDNNNVAAWYDTDL

Rhomboid

MANQQEHNKRYMSGKRTRSYRCAVHQRDREVSSENDFHLLLEDPTLF
ARMVHLVAMEVLPEERDRKYYQERYTCCPPPFFIICVTLLELGVFAWY
AWGAGGVAAAAGPVPVDSPLVYRPDRRRELWRFLTYSVVHAGWLH
LAFNLLVQLAVGLPLEMVHGAVRCGAVYLAGVLGGSLAASVLDPDV
CLAGASGGVYALLAAHLANALLNFHAMRYGAVRLVAALAVASCDV
GFAVHARYTKQEAPPVSYAAHVAGALAGLTIGLLVLKHAQQRLWER
LLWWAALGAYAACTLFAVLYNVFSAPVDELHYMPPDPPPDAGF

Sequence of *engrailed* used for *in-situ* hybridization.

TTGAAGACCGTTGCAGTCCGAACCAGGCCAACAGCCCCGGTCCGGT
CACCGGCAGAGTCCCTGCGCCTCACTCCGAAGTAAGAAACGNGTA
CCAAAGTCAATACACTTGCACGACTATCGATCAAAGGTTTGACAGA
ACGATGACAGTGGTGAAAGTGCAGCCGAATTCACCACCGATGAGT
CCACTGACGTGAAGCCCATAATCCCTGAGTTTGAAGACAAGAGAA
ACCGACAACCACCACCAACCATACCCTTCTCTATCAGCAACATATT
ACACCCAGAATTCGGTTTGACAGCGATTCGAAAAACGAACAAAAT
CGAAGGACCAAAACACGTCGGCCCCAACCACAGCATTTTGTACAA
ACCTTATTTGTCGAACGAGTTATCGAGTTCGAAATTCAATTTCGATT
ATTTAAAATCTAAGGATGATTTCGGTGCATTACCTCCACTTGGCGG
TTTGAGGCAGACCGTGTCGAATATTGGAGAACAGAAGGAGGCACC
AAAGATTATAGAGCAGCAGAAGAGGCCAGATTCAGCCAGCTCTAT
TGTCTCTTCCACATCTAGCGGGGCTTTATCGACGTGTGGCAGCACT
GACGCCAACAGCAGTCAAAGCGGGAACAGCAATCTA

Sequence of *invected* used for *in-situ* hybridization.

GGACCAAAGTGACGAAGAGCACGACCCCTACTCGCCCAACACTAG
AGACACCATCACACCAGACTTCATAGAAGAAGACAAACAAGACAG
GCCTATACACACATCCTCTTTCTCCATACACAATGTCCTTAAGAAG
GAAAGAGACAGTAATAGTCCTGAGAACGTCTTCTCAACTGAAAAG
TTGTTGCAAAGTACACCGAACTTTGAAGATTCTAGGAACTCTGAAA
GCGTTAGTCCGAGACTTGAAGATGATCACAATGAAAGAGCTGATAT
AAGTGTTGATGACAACTCTTGTTGTAGTGATGATACTGTGCTATCTG
TTGGCAATGAAGCCTTACCAACCAATTACCCAAACGACAAAGATCC
GAACCAAGGCTTAACCTCCTTCAAACATATACAAACTCATTTGAAC
GCAATATCACAGTTAAGTCAAAATTTAAACATAAACCAACCAATCC
TCCTACGACCCAACCCAATAACACCAAACCCGTTAATGTTCCTAAA
CCAACCGTTGTTATTCCAAAACCCTTTAATAAACCAAGTGGATTTA
AAATCAGGGTTACCGAGAATCGGCTTGCAGCAAAACAATTTAAATT
TGAACCAAAATTACATGAATTATGCGAGAAAAAATGAACTGAACG
AAAGACGACAGAGTTATTCACCGAAGTTACATGAAAATGAGTCAA
GTAGAGATTTTATTAACCAAGGATGTTTGAAATTTAGCATTGATAA
TATACTGAAAGCTGATTTTGGTAGACGAATTACTGATCCGTTGACA
AAGAGAAAAACGAAGACGAGGCAGTATGAGGCAAAATCTACCCCT
GTCAAAGAGGTTCAGTCTCCCCCTAAAGAGGTAGAGGCTAGAGTG
CCGGA

Sequence of *decapentaplegic* used for *in-situ* hybridization.

GTTCTTCAACGTAAGCGGCGTACCGGCCGACGAGGTGGCGCGCGG
CGCCGACCTCTCGTTCCAACGAGCCGTCGGCACCACCGGCAGACAG
AGACTGTTGTTGTACGACGTGGTGCGCCCTGGCCGCCGCGGCCACT
CCGAGCCGATCCTGCGGCTGCTGGACTCCGTTCCGCTCCGGCCCGG
GGAGGGAATCGTCAACGCCGACGCTCTGGGAGCGGCGCGACGGTG
GCTCAAAGAGCCCAAACATAATCACGGACTATTAGTGCGAGTGTTA
GAAGAAGACGCCGCGAGTGCGAGCAGGGACGCGAAGTTCCCGCAC
GTGCGCGTGCGCAGACGCGTCACGGACGAGGAGGAGGAGTGGCGG
ACGGCGCAGCCGCTGCTCATGCTGTACACGGAGGACGAGCGCGCG
CGCGCGTCGCGGGAGACGAGCGAGCGGCTGACGCGCAGCAAGCGC
GCGGCGCAGCGGCGGGGGCACCGCGCGCACCACCGCCGCAAGGAG
GCGCGCGAGATCTGCCAGCGCCGCCCGCTGTTCGTCGACTTCGCGG
ACGTGGGCTGGAGCGACTGGATCGTGGCCCCGCACGGCTACGACG
CGTACTACTGCCAGGGCGACTGCCCCTTCCCGCTGCCGGACCACCT
CAACGGCACGAACCACGCGATAGTGCAGACTCTGGTCAACTCAGT
GAACCCCGCGACGGTGCCCAAAGCGTGCTGCGTGCCGACGCAACT
CTCATCTATATCTATGTTATATATGGACGAAGTGAACAATGTGGTG
CTTAAAAACTATCAGGACATGATGGTGGTAGGCTGTGG

Sequence of *blistered* used for *in-situ* hybridization.

GCATACGAGCTATCAACGCTGACCGGCACCCAAGTGATGCTGCTGGTCGC
GTCGGAGACCGGCCACGTGTACACGTTCGCGACACGCAAACTGCAGCCGA
TGATCACGTCCGACTCGGGCAAGCGGCTCATACAGACGTGCCTCAACTCG
CCCGACCCGCCCACCACCAGCGAGCAGCGCATGGCCGCCACCGGCTTCGA
GGAGACCGAGCTCACGTATAACGTTGTAGACGACGAGATGAAGGTGAGA
CAACTGGCGTACGCTAACCAGTACCCCATAGAGCACCACCCGGGGTTGGC
GCCGTCGCCACTGCAGCAGTACCACCAGCACCCGCCCTGCCCCTCGCCCC
TCCCCCTCGGCTCGCTGGGCCAGCCGTACTCGCACGCGCATCTATCGCACC
CCCACATGTCTCACCACCCGCAACG

Sequence of *wingless* used for *in-situ* hybridization

CAGCAGCTGGATTTTGTCAGTCCAGCTAGGAAGGGGGGCATAGCAAAGG
CAGGCGAACCAAATAACTTATCACCCTTGTCTCCAAGTGTCCTATACATG
GACCCGGCTGTTCACGCCACCTTGAGGAGGAAACAGAGAAGGCTAGCGA
GGGAGAACCCTGGGGTCCTCGCAGCAATATCCAAGGGAGCCAGCATGGC
TGTGGCCGAATGCCAGCATCAGTTCAAATACAGGAGATGGAACTGTTCTA
CAAGAAATTTTTTGCGAGGGAAGAATCTATTTGGAAAAATTGTTGACAGA
GTTTCGCCGGACAAAGCCCCCCCGGCCGGGGCGGCTATAATTACTAATAT
ACACGTCGACACGCCATTGACGATTGACGCGACATCTTCATTTCATTGTGG
TGTAAACCTCAAGGATCGCATTAACACGGACGATGACAACGGCACAATA

Region of *spalt* targeted by CRISPR-Cas9 (location of guide RNA highlighted in red)

GCATCGACAAGATGCTGAAAATAATAATAGTCTCGAAGACGGCGA
GGCCGAAATACCTGAAGCCGACATGCCCCCCGTGGGTCTGCCGTTC
CCTTTGGCAGGACACGTTACTCTTGAGGCTCTACAAAATACGAGAG
TAGCGGTCGCCCAATTCGCTGCAACAGCGATGGCAAATAATGCGA
ATAACGAAGCTGCTATACAAGAATTACAAGTGTTACACAACACTCT
ATACACTTTACAGTCACAACAAGTATTTCAACTTCAGTTAATACGT
CAGCTTCAGAATCAGTTATCTCTAACTCGACGGAAAGAAGACGATC
CACACAGCCCACCGCCAAGTGAACCAGAACAGAATGCCCCGTCAA
CGCCGGCTCGATCACCGTCGCCGCCGCGTCCGCCACGGGAGCCGTC
GCCTGTTATACCCTCTCCTCCTACTAGCCAAAGTTTGCCGTCGACTC
ACACACATCACACACCCAAAACTGAACAGATATCTATCCCTAAGAT
TCCAACTTCCTCACCATCTTTAATGACCCACCCACTTTATAGTTCAA
TTTCTTCGTCATTAGCATCTTCCATCATAACAAACAATGATCCTCCA
CCGTCCCTAAATGAA

Region of *optix* targeted by CRISPR-Cas9 (location of guide RNAs highlighted in red)

ATGCGCGGCTCCTGGGACGAGTCCACGACGGCGGCGCTGCACGCG
CGCATCCTGGAGGCGCACCGCGGGTCCGCCGCGCCCGACCGCGCC
GAGCCCGCGTGCGAGCCTCCGCCGCTGACGCTGGGCGCGCTGGAG
CTGGCGGCGCCCACGCCGCTGCTGCCGCTGCCCACGCTGAGCTTCA
GCGCCGCGCAGGTGGCCACCGTGTGCGAGACGCTGGAGGAGAGCG
GCGACGTGGAGCGCCTGGCGCGCTTCTTGTGGTCGCTGCCCGTGGC
GCACCCCAACGTGGCCGAGCTGGAGCGCTGCGAAGCCGTGCTGCG
CGCGCGCGCCGTCGTCGCCTTCCACGCCGGCCGCCACCGCGAGCTG
TACGCCATCCTCGAGCGCCACCGCTTCCAGCGCTCCAGCCACGCCA
AGCTGCAAGCGCTGTGGCTGGAGGCGCACTACCAGGAGGCTGAGC
GCCTGCGCGGCCGTCCGCTGGGCCCCGTCGACAAGTACCGCGTGCG
GAAGAAGTTCCCGCTCCCGAGGACGATCTGGGACGGCGAGCAGAA
GACGCACTGTTTCAAGGAGCGGACGCGATCTCTACTCCGAGAATGG
TACCTCCAAGATCCCTACCCGAACCCGACGAAGAAGAGGGAATTG
GCGGCGGCGACGGGTCTGACGCCGACGCAAGTCGGCAACTGGTTC
AAAAACCGACGGCAAAGAGACCGAGCGGCCGCCGCCAAGAACCGC
TCCGCCGTGCTGGGCAGAGGATAA

Region of *dpp* targeted by CRISPR-Cas9 (location of guide RNA highlighted in red)

GTTCTTCAACGTAAGCGGCGTACCGGCCGACGAGGTGGCGCGCGG
CGCCGACCTCTCGTTCCAACGAGCCGTCGGCACCACCGGCAGACAG
AGACTGTTGTTGTACGACGTGGTGCGCCCTGGCCGCCGCGGCCACT
CCGAGCCGATCCTGCGGCTGCTGGACTCCGTTCCGCTCCGGCCCGG
GGAGGGAATCGTCAACGCCGACGCTCTGGGAGCGGCGCGACGGTG
GCTCAAAGAGCCCAAACATAATCACGGACTATTAGTGCGAGTGTTA
GAAGAAGACGCCGCGAGTGCGAGCAGGGACGCGAAGTTCCCGCAC
GTGCGCGTGCGCAGACGCGTCACGGACGAGGAGGAGGAGTGGCGG
ACGGCGCAGCCGCTGCTCATGCTGTACACGGAGGACGAGCGCGCG
CGCGCGTCGCGGGAGACGAGCGAGCGGCTGACGCGCAGCAAGCGC
GCGGCGCAGCGGCGGGGGCACCGCGCGCACCACCGCCGCAAGGAG
GCGCGCGAGATCTGCCAGCGCCGCCCGCTGTTCGTCGACTTCGCGG
ACGTGGGCTGGAGCGACTGGATCGTGGCCCCGCACGGCTACGACG
CGTACTACTGCCAGGGCGACTGCCCCTTCCCGCTGCCGGACCACCT
CAACGGCACGAACCACGCGATAGTGCAGACTCTGGTCAACTCAGT
GAACCCCGCGACGGTGCCCAAAGCGTGCTGCGTGCCGACGCAACT
CTCATCTATATCTATGTTATATATGGACGAAGTGAACAATGTGGTG
CTTAAAAACTATCAGGACATGATGGTGGTAGGCTGTGG

# Chapter 3: Optix: An eye development gene paints the eyespots of *Bicyclus anynana* butterflies via a possible positional information mechanism


## Abstract

*Optix*, a gene essential and sufficient for eye development in *Drosophila melanogaster*, also plays important roles in the development of both the structure and pigmentation of butterfly wing scales. In particular, *optix* regulates wing scale lamina thickness and ommochrome pigment synthesis. Here we explore the involvement of Optix in the wing pattern of the African squinting bush brown butterfly *Bicyclus anynana* by examining its expression on the wings using immunostainings and testing its function via CRISPR-Cas9. Optix in *B. anynana* is expressed in multiple domains, most prominently in the orange ring of the eyespots and in other scattered orange scales, in the white scales at the center of the eyespots, and in patches of silver scales. We showed that *optix* regulates both the pigmentation and the scale structure of the orange scales. We further explored the interaction of Optix with Spalt, a Protein which is involved in the development of black scales during eyespot development. CRISPR knockouts on *optix* and *spalt* followed by immunostaining provide information on how the domain over which Optix is expressed in eyespots is controlled by Spalt in the wings of butterflies, a mechanism similar to the wing anterior-posterior gene regulatory network. We propose a positional information model for eyespot development where Decapentaplegic acts as a central morphogen activating other downstream targets that result in the formation of two eyespot rings. We also propose that

*optix* functions in combination with other genes in color patterning and scale morphology.

**Keywords**

Optix, *decapentaplegic*, Spalt, positional information, eyespots, *Bicyclus anynana*

## Introduction

*Optix,* a member of the *sine oculis homolog* (*six/so)* gene family is involved in eye morphogenesis and wing patterning in the model vinegar fly *Drosophila melanogaster* (Martín et al., 2017; Seimiya and Gehring, 2000). *Optix* has been shown to be expressed in the eye-antenna, wing, and haltere imaginal discs during the third instar larval development of *D. melanogaster* (Martín et al., 2017; Seimiya and Gehring, 2000). Ectopic expression of *optix* is sufficient to induce the development of eyes via a mechanism independent of *eyeless*, a gene previously implicated in eye morphogenesis (Quiring et al., 1994; Seimiya and Gehring, 2000), while reduced level of Optix creates defect in the L2 (R2+3) vein in the wing (Al Khatib et al., 2017; Martín et al., 2017).

Optix is involved in venation patterning of *D. melanogaster* wings via a positional information mechanism (Martín et al., 2017). The morphogen Decapentaplegic (Dpp) expressed along the Anterior-Posterior (A-P) wing margin is involved in the activation of multiple different venation patterning genes in the anterior compartment of the third instar wing imaginal disc which includes the transcription factors *spalt*, *aristaless* (*al*) and *optix* at different concentration threshold (Martín et al., 2017). A high concentration of Dpp activates *spalt* posterior to the vein L2 while lower levels can activate *al* and

*optix* throughout the anterior compartment. However, Spalt is a repressor of *optix* and as a result, *optix* is only expressed in the upper anterior compartment. The complementary expression of Spalt and Optix creates a perfect boundary between these two genes where Al activates the vein specific gene *knirps* (a gene repressed by both Spalt and Optix) that determines the fate of provein cells at L2 (Martín et al., 2017). Thus, the complex interaction of these multiple genes results in the development of a single vein via a positional information mechanism.

In butterflies, Optix has been primarily investigated with respect to color pattern development (Reed et al., 2011; Thayer et al., 2020; Zhang et al., 2017b). Optix is considered one of the hotspot genes along with WntA, Aristaless, and Cortex, where allelic variants result in a wide array of wing pattern variation in *Heliconius* butterflies found in South America (Jiggins et al., 2017; Reed et al., 2011). In *Heliconius*, *Agraulis*, *Junonia*, and *Vanessa* butterflies, Optix has been shown to be involved in the development of ommochrome pigments, as *Optix* crispants result in the loss of red and orange pigments (Zhang et al., 2017b). In *Junonia coenia*, removal of *Optix* results in the development of blue iridescent scales (Zhang et al., 2017b) as a result of decreased ommochrome production and increased lower lamina thickness of the scales (Thayer et al., 2020). Multiple *cis*-elements have been shown to control the domain over which *Optix* is expressed in *Heliconius* wings (Lewis et al., 2019). A recent paper proposes that *optix* is under the control of *cortex* a member of cdc20 family gene that controls the cell cycle (Livraghi et al., 2020). *cortex* knockouts result in the transformation of scales with ommochrome pigments to white and yellow scales (Livraghi et al., 2020).

Furthermore, Optix is expressed along the veins of *Dryas julia* butterflies (Martin et al., 2014), and along with *doublesex* might be involved in development black and gold scales that are used to protect the pheromone diffusing structures along the veins in male butterflies (Martin et al., 2014; Rauser and Rutowski, 2003).

Eyespots in *Bicyclus anynana* butterflies contain a circular ring which is orange in color and likely involves Optix for scale differentiation and pigmentation but the function of this gene in eyespot development has not yet been explored. Over a dozen genes have been visualized as expressed in the center of eyespots, using *in-situ* and antibody stainings, and over 180 transcripts have been proposed to be differentially expressed in the eyespot centers relative to flanking tissue using RNAseq (Monteiro, 2015; Özsu and Monteiro, 2017). Some of the genes that have been studied include *distal-less* (*dll*)(Connahs et al., 2019; Monteiro et al., 2013), *decapentaplegic* (*dpp*) (Connahs et al., 2019), *spalt* (Monteiro et al., 2006; Oliver et al., 2012), *wingless* (*wg*) (Özsu et al., 2017), and *engrailed* (*en*) (Banerjee et al., 2020; Brunetti et al., 2001). These genes have also been shown to be differentially expressed in the different domains of the eyespots (Banerjee et al., 2020; Brunetti et al., 2001; Connahs et al., 2019). A reaction-diffusion model involving the morphogens *dpp* and *wg* and the transcription factor Dll has been proposed in setting up the center of the eyespot in the larval stage (Connahs et al., 2019). A positional information model has been proposed to be involved in setting up the rings of the eyespot where the signal from a morphogen secreted from the center of the eyespot is received by cells surrounding it and based on different concentrations perceived by these cells,

downstream target genes are activated via a threshold-like mechanism in the rings (Nijhout, 1978; Nijhout, 1980).

The role of *Optix* and its involvement with other genes in *B. anynana* has remained unexplored. In the present research we studied *Optix* both at the expression and functional levels in the overall patterning of the wing, focusing mostly on the eyespots. We also explored the possible interaction of *Optix* with other genes involved in eyespot development such as Dpp and Spalt. Based on these results, we propose a positional information model for setting up the rings of the eyespots where Dpp likely acts as a central morphogen activating Spalt at a higher concentration in the black scale region and Optix at a lower concentration in the outer ring surrounding Sal expression. Spalt represses the activation of Optix in the black scales and as a result, Optix is present only in the orange scale region.

**Methods**

Rearing *Bicyclus anynana*

*B. anynana* butterflies were raised in lab condition at 27°C and 12-12 hrs day-night cycle. Larvae were fed young corn leaves and adults were fed mashed bananas.

CRISPR-Cas9

Optix and Spalt CRISPR experiments were carried out based on a protocol previously described (Banerjee and Monteiro, 2018). To summarize guides were designed targeting the coding sequence of the genes (see supplementary file for sequence and the regions targeted). Solution containing Cas9 protein (IDT, Cat No. 1081058) and guide RNA at the concentration of 300ng/ul each

along with Cas9 buffer and a small amount of food dye was injected in 1509 embryos for *optix* and 863 embryos for *spalt* (**Table S3.1 and S3.2**). A few of these injected individuals were dissected at 24 hrs after pupation to perform immunostaining. The rest were allowed to grow till adulthood. After eclosing, the adults were frozen at -20 C and imaged under a Leica DMS1000 microscope.

*In-situ* hybridization

For the visualization of *dpp* expression in the pupal wings of *B. anynana*, 20-24 hrs old pupal wings were timed using an Olympus tough tg-6 camera. The wings were dissected in 1x PBS under a Zeiss stemi 305 microscope and transferred to 1xPBST with 4% formaldehyde. After fixation for 30 mins the wings were washed three times in 1x PBST for 5 mins each. The wings were then treated with proteinase K and glycine and washed again three times using 1x PBST. The wings were then gradually transferred into prehyb buffer (see **table A1** for composition) and heated at 65ºC for 1 hr. Hybridization buffer with a *dpp* probe (see **table A1** for composition) was added to the wings and incubated for 16 hrs at 65ºC. Wings were then washed five times for 30 mins each in prehyb buffer. After washing, wings were moved to room temperature and gradually transferred to 1x PBST and washed in 1x PBST. After, the wings were incubated in block buffer (see **table A1** for composition) for 1 hr, after which anti-digoxygenin (Sigma-Aldrich, Cat No. 11093274910) diluted 1/3000 times in block buffer was added. After 1 hr of incubation the wings were washed five time, five mins each in block buffer. Finally, wings were transferred to alkaline-phosphatase buffer (see supp table S4 for composition) supplemented with NBT-BCIP (Promega, Cat No. S3771) and incubated in the

dark till the development of color. The wings were then imaged under a Leica DMS1000 microscope.

Immunostaining

For immunostaining, pupal wings were dissected at different timepoints under a Zeiss Stemi 305 microscope in 1x PBS at room temperature. Wings were fixed using 4% formaldehyde in fix buffer (see **Table A2** for details), followed by four washes in 1x PBS, five mins each. After the washes, the wings were incubated in block buffer at 4ºC for overnight. The next day primary antibodies against Optix (1:3000, rat, a gift from Robert D. Reed) and Sal (1:20000, guinea pig GP66.1), were added in wash buffer and incubated at 4ºC for 24 hrs. The next day anti-rat AF488 (Invitrogen, #A-11006) and ant-guniea pig AF555 (Invitrogen, # A-21435) secondary antibodies at the concentration of 1:500 in wash buffer were added followed by four washes in wash buffer, 20 mins each. Wings were then mounted on an inhouse mounting media and imaged under an Olympus fv3000 confocal microscope.

Scanning Electron microscope

A fine needle was used to pick scales from adult WT and optix crispant wings. The scales were then mounted on a carbon tape fixed to an SEM stub. For the measurement of the scale lamina thickness, wing tissue containing scales of interest were frozen in liquid nitrogen for 5 mins. The scales were then broken using a fine blade (Swann-Morton, Blade no. 4) and mounted on carbon tape and imaged under an SEM.

Scale transmission measurement

Affected regions from wings from four adult optix crispants were imaged and compared with homologous regions in four WT individuals under an Ocean-optics spectrophotometer. The absorbance spectra readings were recorded using spectra suite software and saved in txt format. The files were analysed in R using 'pavo' package (R codes are mentioned in the supplementary file).

**Results**

Expression and function of Optix in eyespots

To examine whether Optix is expressed in the pupal wings of *B. anynana* we used immunostains. Optix is expressed prominently in the orange ring of the eyespots from 24-72 hrs **(Fig. 3.1C-F**) of pupal development. To test the function of *optix* we used injections of Crispr/Cas9 into ealy embryos. Disruptions to Optix resulted in the transformation of the orange colored scales to brown (**Fig. 3.1G-J**). To confirm that these phenotypes were due to the loss of Optix expression, we performed Optix immunostains in some of the injected animals at the pupal stage of development. This revealed that Optix expression was missing from cells of the orange ring of the eyespots (**Fig. 3.1K and L**). Finally, we confrmed that knocking out Optix leads to deletion of nucleotides mostly near the targeted sites (**Fig. 3.1M and N**).

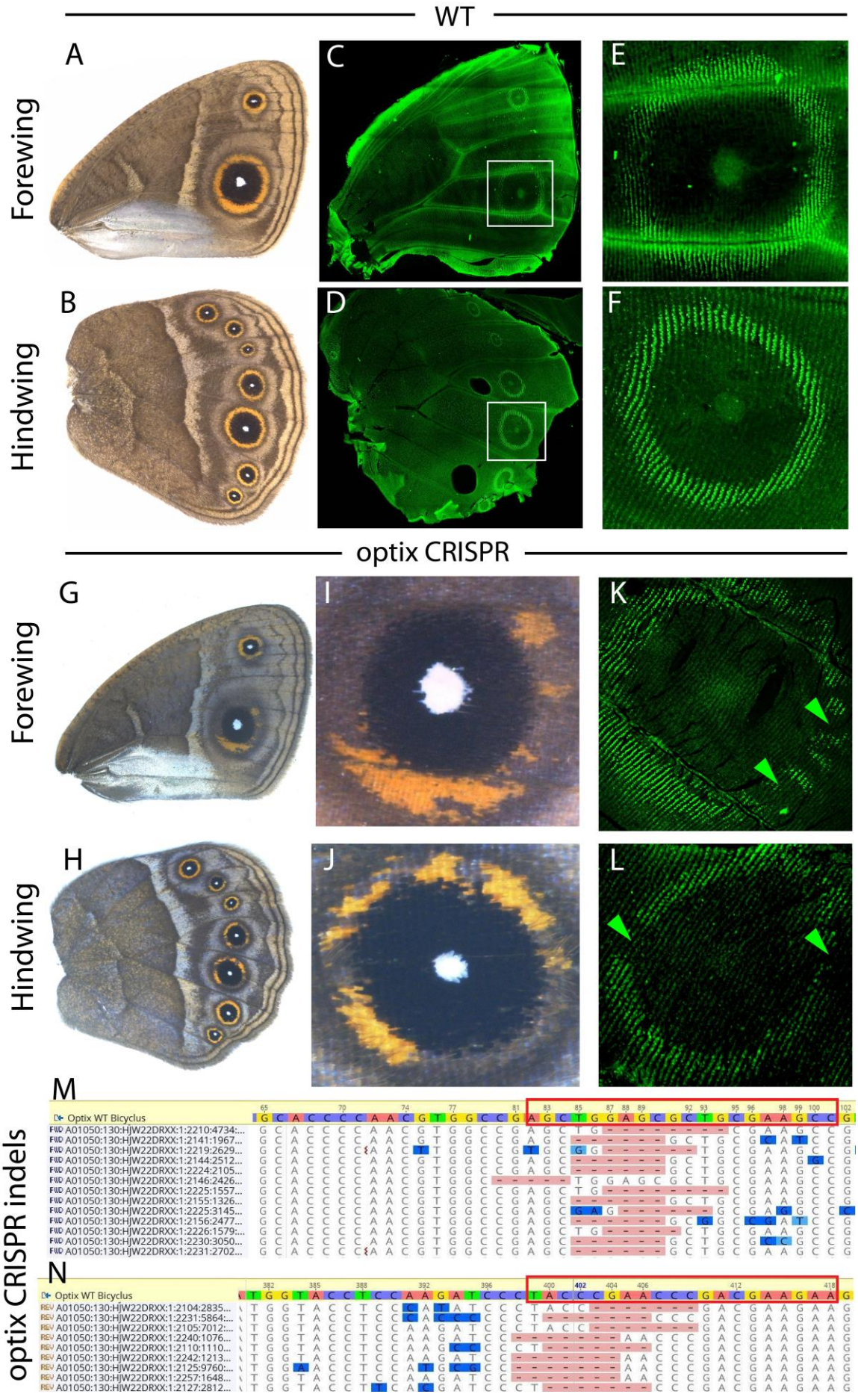

WT
A
B
C
D
E
F
Forewing
Hindwing
optix CRISPR
G
H
I
J
K
L
Forewing
Hindwing
M
N
optix CRISPR indels
Optix WT Bicyclus

**Figure 3.1. Expression and function of Optix in the wings of *B. anynana* two days after pupation. (A)** WT forewing, **(B)** WT hindwing. **(C)** The expression of Optix in the forewing and **(D)** in the hindwing. **(E and F)** Expression of Optix in the eyespot orange ring of Cu1 eyespots. (**G-J**) CRISPR-Cas9 knockout of *optix* in the forewing and hindwing produces defects in the development of the eyespot's orange ring. (**K and L**) optix CRISPR results in the loss of Optix in cells of the future orange ring area of the eyespots. **(M and N)** CRISPR deletions at the optix target sites.

Transformation of scale structure in the eyespots due to *optix* knock-out

Knocking-out *optix* resulted in the loss of a partial upper lamina that connects the cross-ribs in the orange scales (**Fig. 3.2C and D**) making these scales resemble WT brown scales (**Fig. 3.2E**). The structure of black and white scales was not affected (**Fig. S3.2C, D, G, and H**).

Changes in the pigmentation due to *optix* knock-out

Knocking out *optix* resulted in the change of absorbance spectra of the wing from orange to brown at the orange ring region (**Fig. 3.2F**) resembling WT brown scales (**Fig. 3.2E**).

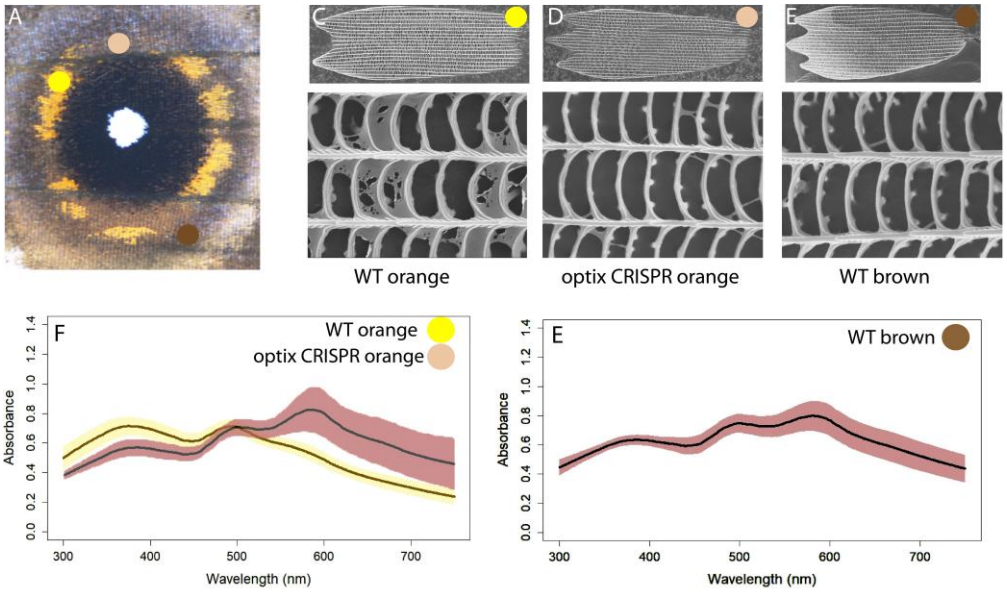


**Figure 3.2. Optix promotes the development of the upper lamina and prevents the development of ommochromes in the orange scales.** (**A and B**) Optix knockout results in the conversion of orange scales into brown scales. SEM images of (**C**) orange scale from WT scale region, (**D**) orange scale from optix CRISPR scale region, and (**E**) brown scale from WT region. Knocking out *optix* results in the reduction of an upper lamina found in orange scales making these scales resemble brown scales. (**F and E**) Absorbance spectra of WT and optix CRISPR scale patches. Knocking-out *optix* results in the conversion of the absorbance spectra of orange scales to that of brown like scales, indicating a change in pigments. Colored dots represent the areas used for spectrophotometric analysis.

Expression of *wingless* (*wg*), *decapentaplegic* (*dpp*), Spalt, Armadillo (Arm) and Engrailed (En) in the eyespots

*In-situ* hybridization on the transcript of *wg* and *dpp* showed expression in the eyespot center in one day old pupal wings (**Fig. 3.3A-D**). Immunostaining against Spalt showed expression in the eyespot center and in the tissues where the future black scales are going to form (**Fig. 3.3C and D**). Immunostaining against Arm (the signal transducer of Wnt (Wg) signaling) showed expression in the eyespot center (**Fig. 3.3K**) which overlaps with the expression of En at the center of the eyespot (**Fig. 3.3M**). En is also expressed in the cells that will form the orange ring (**Fig. 3.3L**) which will be explored in chapter 4.

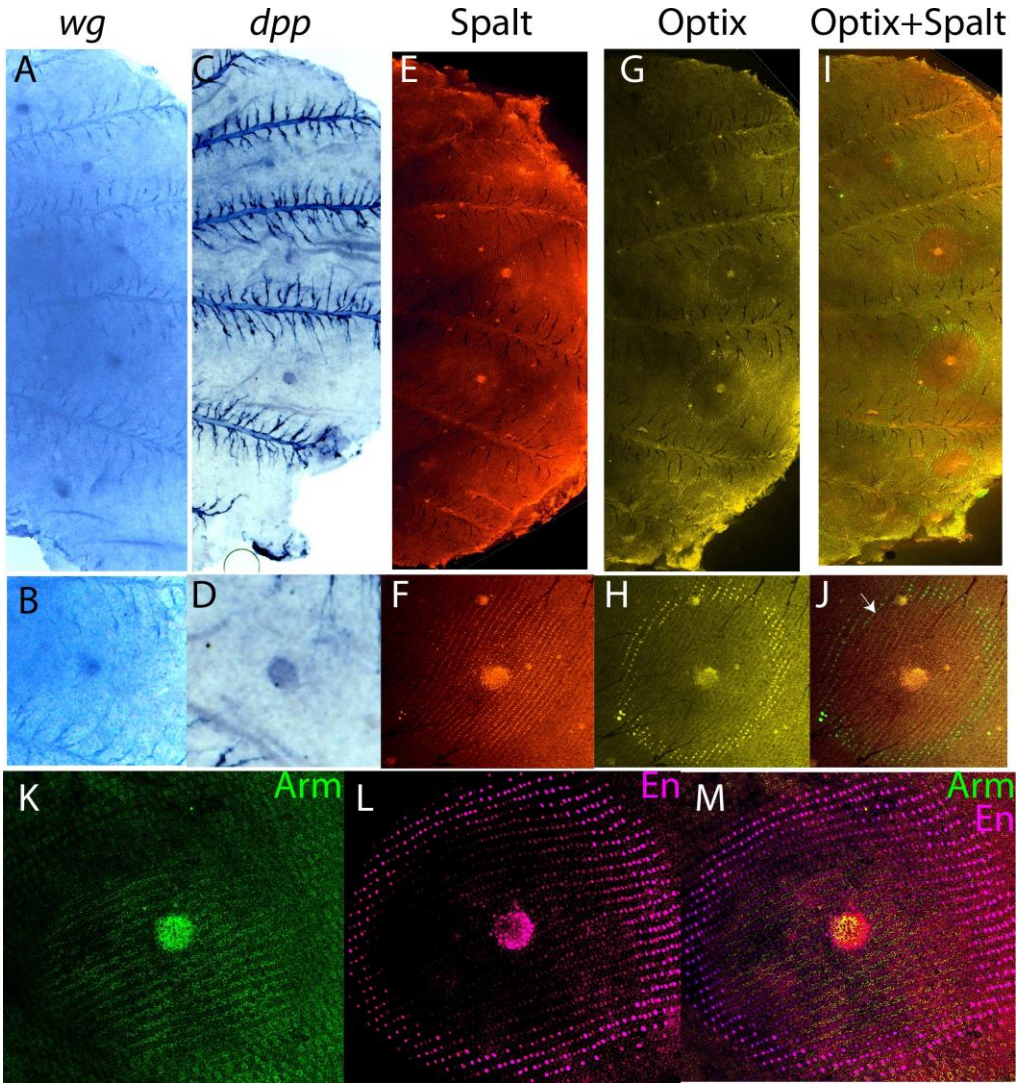



**Figure 3.3. Expression of *wingless* (*wg*), *decapentaplegic (dpp)*, Spalt, Optix, Armadillo (Arm) and Engrailed (En) in 18-24 hrs pupal wing of *B. anynana*. (A and B)** Expression of *wg* in the eyespot center. **(C and D)** Expression of *dpp* in the eyespot center **(E and F)** Expression of Spalt in the eyespot center and in the future black scales. **(G and H)** Expression of Optix in the eyespot center and the future orange ring. (**I and J**) Expression of Optix and Spalt. The boundary where Sal expression stops and Optix expression starts is marked by white arrow. **(K)** Expression of Arm in the eyespot center. **(L)** Expression of En in the eyespot center and in the orange ring. **(M)** Co-expression of Arm and En.

Interaction of Optix with Spalt in the eyespot

Deletion of *optix* resulted in a loss of orange pigments in the orange scales **(Fig. 3.4E)** due to the loss of Optix expression in the orange ring **(Fig. 3.4F)**.

However, deletion of *optix* doesn't affect the expression of Spalt **(Fig. 3.4G)**.

Deletion of *spalt* resulted in the development of orange scales in the black scale regions of the eyespot **(Fig. 3.4I)** due to the loss of Spalt **(Fig. 3.4K)**, which resulted in the ectopic expression of Optix in the black scale region **(Fig. 3.4J)**.

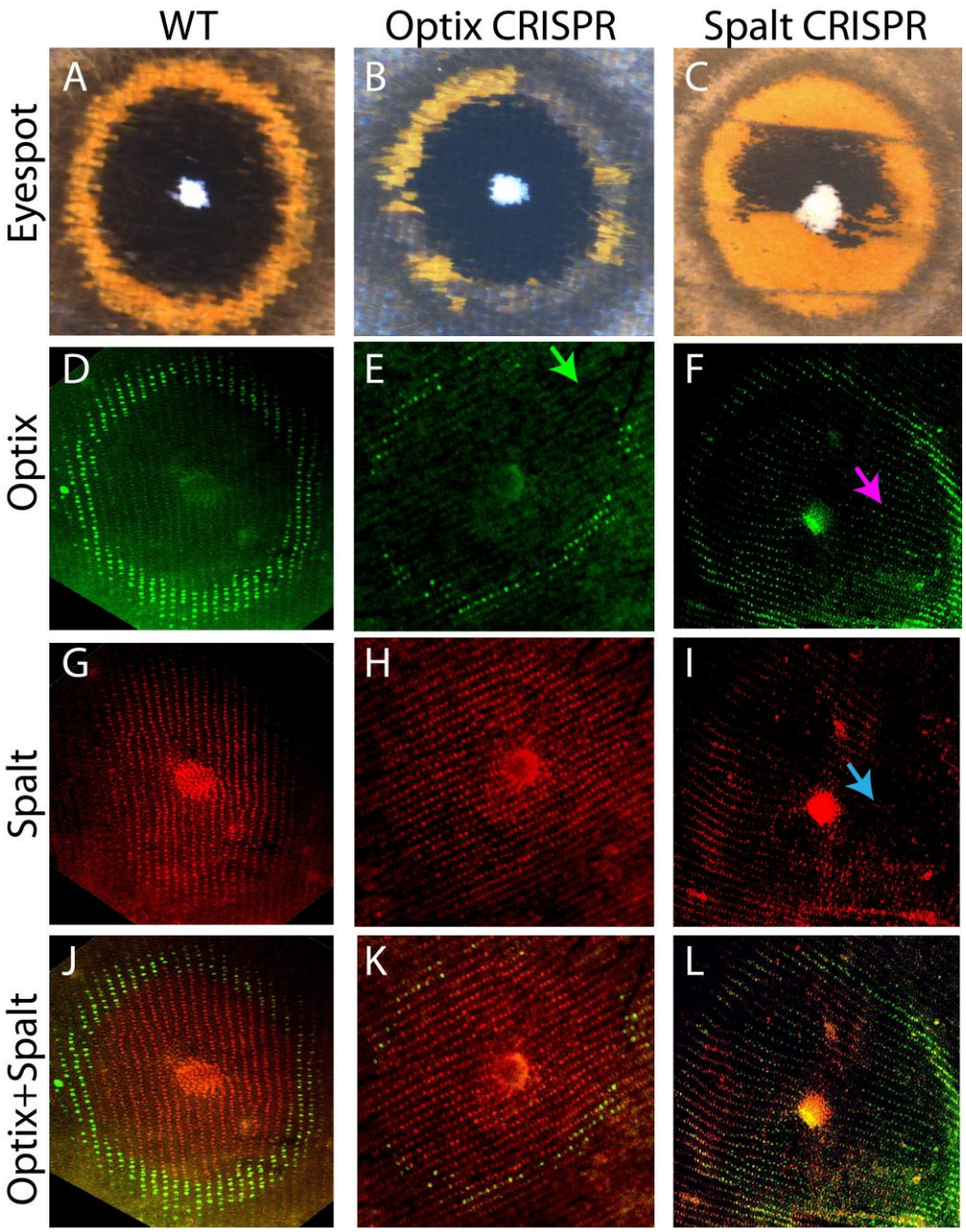



**Figure 3.4. Spalt prevents *optix* from being expressed in the black disc region of the eyespot (18-24 hrs pupal wings)**. **(A)** WT eyespot, **(B)** *optix* CRISPR eyespot, and **(C)** *spalt* CRISPR eyespot. Eyespots stained with the antibody against Optix in **(D)** WT, **(E)** *optix* CRISPR, and **(F)** *spalt* CRISPR individuals. Eyespot stained with the antibody against Spalt in **(G)** WT, **(H)**

*optix* CRISPR and **(I)** *spalt* CRISPR individuals. **(J-L)** co-staining of Spalt and Optix. Green arrow marks the region of missing Optix protein in the *optix* CRISPR individual. The expression of Spalt is not affect due to deletion of *optix*. Pink arrow marks the ectopic expression of Optix, where Spalt expression is missing (blue arrow) in a *spalt* CRISPR individual.

Interaction of Optix with Spalt

Deletion of *optix* doesn't affect the expression pattern of Spalt (**Fig. 3.4B, E, H, and K**). Optix is expressed in a clear halo like pattern in the orange precursor scale cells, while Spalt is expressed in the future black scale cells. Deletion of *optix* via CRISPR-Cas9 results in loss of Optix in the orange ring of scale cells and loss of orange color in those scales. However, no changes in the expression of Spalt and in the black scale phenotype is observed, indicating that Optix is not involved in regulating the outer perimeter of *spalt* expression in the eyespots.

Spalt represses *optix* expression (**Fig. 3.4C, F, I, and L**). Knocking out spalt using CRISPR-Cas9 resulted in the loss of black pigments and formation of orange scales in the black scale region (**Fig. 3.4C**). Immunostaining *spalt* crispants in the eyespot region revealed that loss of Spalt in the black scale cells led to the presence of Optix in those cells (**Fig. 3.4F, I, and L**). These data indicate that Spalt is a repressor of *optix*, where loss of *spalt* leads to activation of *optix*.

Both Optix and Spalt are co-expressed in the center of the eyespots, but Optix neither controls *spalt* expression nor plays any role in the differentiation of the eyespot white center cells (**Fig. 3.4E, H and K**). Immunostaining of Optix and Spalt shows that both the genes are co-expressed in the white centers of the eyespots (**Fig. 3.3I and J; Fig. 3.4D, G and J**). Deletion of *optix*, however,

doesn't produce any defects in the pigmentation and scale ultrastructure of the white scales (**Fig. 3.4B; Fig. S3.2G and H**). The expression of both Optix and Spalt in the eyespot center remained unchanged in the *optix* CRISPR wings when compared to the WT expression (**Fig. 3.4E, H, and K**). These results indicate that Optix is not likely involved in the differentiation of the white center of the eyespot.

## Discussion

### Optix is expressed in novel domains in eyespots where it controls both pigmentation and scale ultrastructure development

Optix in *B. anynana* is expressed in the cells responsible for the formation of the orange ring where it controls both pigmentation and cuticularization of upper lamina in between the cross ribs (**Fig. 3.1 and 3.2**). Optix has been shown to be involved in the formation of red and orange ommochrome pigments in butterflies as *optix* knock outs result in loss of these pigments in *Heliconius erato, Agraulis vanilla, Vanessa cardui* and *Junonia coenia* (Reed et al., 2011; Zhang et al., 2017b). Furthermore, *optix* has been shown to be involved in the formation of the lower lamina of the scales in *Junonia coenia*, where deletion of *optix* results in increased thickness of the lower lamina along with loss of ommochrome and transformation of brown (dark orange brown) scales to thicker blue iridescent scales (Thayer et al., 2020). Thayer et al., (2020) proposed that Optix might be involved in regulation of chitin synthase. Immunostaining of Optix in *B. anynana* shows expression in the cells involved in the formation of the orange ring of the eyespot, and *optix* deletions result in the loss of both the orange pigment (likely a ommochrome)

and the upper lamina in the transformed orange scales (**Fig. 3.2**). Though the function of Optix in the development of ommochrome pigments was known (Zhang et al., 2017b), the loss of the upper lamina in *optix* negative scale cells is novel (**Fig. 3.2B and C**). We also observe similar reductions in the extent of the upper lamina in the silver scales where Optix is also expressed (**Fig. S3.5**). The reduction in the formation of the upper lamina could be due to loss of chitin synthase i.e., the gene involved in the formation of chitin (Merzendorfer and Zimoch, 2003) in between the cross-ribs. The loss of chitin in the upper lamina is, however, opposite to the gain of chitin in the lower lamina in *J. coenia* where loss of *optix* results in increased thickness of the chitin creating a blue iridescent color (Thayer et al., 2020).

Spalt controls the domain over which *optix* is expressed but not the other way around

Deletion of *optix* results in the loss of Optix in the orange ring area, but it doesn't affect the expression domain of Spalt (**Fig. 3.4**). These results are consistent with the study on *D. melanogaster* wing vein patterning, where inhibition of *optix* doesn't result in the movement of Spalt into the upper anterior domain of the wing where Optix is expressed (Martín et al., 2017). Overexpression of *optix*, however, shrunk the Spalt domain in the wing disc (Martín et al., 2017).

Loss of *spalt* results in ectopic expression of Optix in the black scale area, indicating that Spalt is repressing the activation of *optix* in the black scale area

(**Fig. 3.4L**). These data are consistent with the interaction of *optix* and *spalt* in *D. melanogaster* wing disc, where inhibition of *spalt* expands the domain over which *optix* is expressed in the anterior compartment (Martín et al., 2017).

The expression of Optix in the white scale cells must be in response to some upstream gene involved in the formation of the eyespot center which leads to its expression but no function in the white scale. Over 180 genes have been proposed to be associated with the center of the eyespots (Özsu and Monteiro, 2017) and perhaps one of these genes is involved in the activation of *optix* in the center. None of these genes, however, includes *cortex*, which was proposed to activate *optix* in *Heliconius* melanic and red patches (Livraghi et al., 2020). It would be interesting to explore the expression of *cortex* in *B. anynana* and see if it is involved in the formation of the eyespot center and the orange ring.

A positional information system is likely involved in setting up the rings of the eyespot

Researchers have hypothesised that a positional information mechanism involving a central morphogen is involved in setting up the rings of eyespots (Monteiro et al., 2006; Nijhout, 1978). Wingless (Wg) and Dpp are two such morphogens proposed to be involved in eyespot ring differentiation (Monteiro et al., 2006) which are proposed to activate *spalt* and *distal-less* in the black scale region; and *engrailed* in the orange scales (Beldade and Peralta, 2017; Brunetti et al., 2001; Monteiro et al., 2006). A recent study used RNAi to show that Wg could be the morphogen, where the inhibition of *wg* resulted in the reduction of the overall size of the eyespots (Özsu et al., 2017). The other

morphogen *dpp* is expressed during the early pupal wing development (within day one after pupation) (**Fig. 3.3C and D**). Furthermore, the signal transducer of Dpp, pSMAD is also expressed in the center of the eyespot during early pupal wing development, where it is likely involved in activating other downstream target genes (Monteiro, 2006).

The rings of the eyespots are colored by a probable positional information system involving Dpp, which activates *spalt* and *optix* at different concentration thresholds (**Fig. 3.5C**). Dpp is a long-range morphogen that has been shown to activate *spalt* at higher concentration threshold and *optix* at lower concentration threshold via a positional information mechanism in the larval wing discs of *D. melanogaster* (**Fig. 3.5A**; (Martín et al., 2017)). Spalt is a repressor of *optix* and as a result *optix* expression is only observed in the upper anterior compartment of the fly wing disc (**Fig. 3.5A**, (Martín et al., 2017)). A similar mechanism has been proposed for butterfly larval wing discs as well, as discussed in the chapter 2. In the eyespots we observed the presence of *dpp* in the center (**Fig. 3.3C and D**) where it likely acts as a morphogen activating *spalt* at high concentration in a disc around the center (**Fig. 3.3E and F**) and *optix* at low concentration in the periphery of the black disc region (**Fig. 3.3G and H**). Spalt is then involved in the repression of *optix* in the black scale region resulting in the activation of *optix* in a halo like pattern in the orange scale region (**Fig. 3.4L; Fig. 3.5C**). Both Spalt and Optix are co-expressed in the eyespot center, so some other gene is probably preventing the negative cross-regulation of *optix* by Spalt in this region.

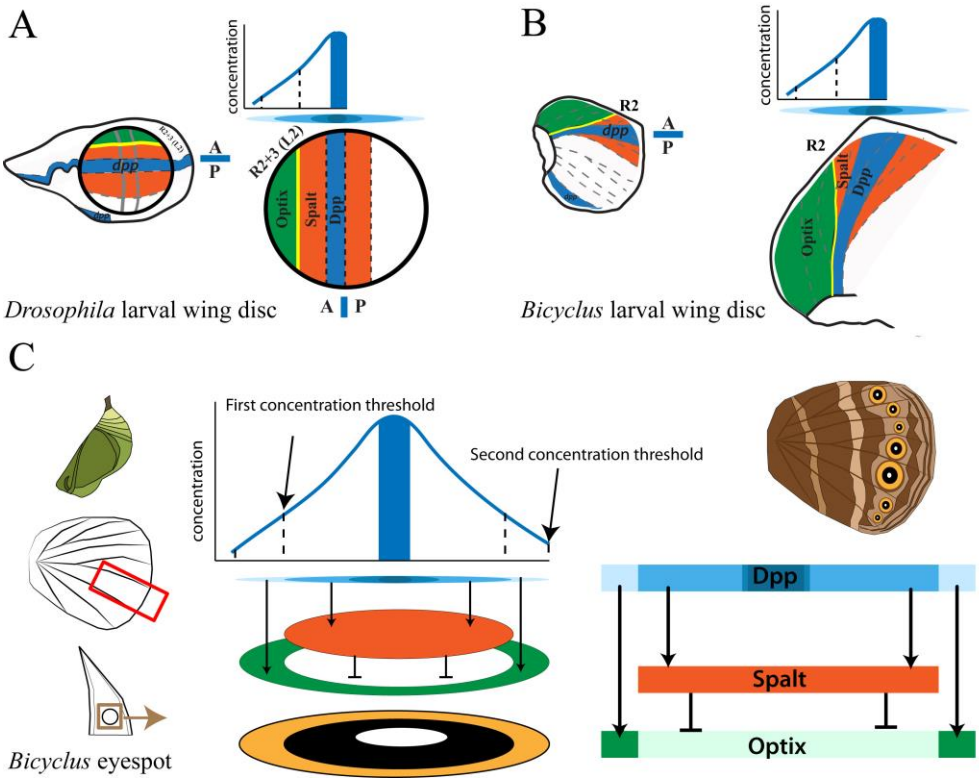



**Figure 3.5. A positional information model for the eyespot pattern formation via a mechanism similar to the anterior-posterior wing network. (A)** Positioning of Drosophila R2+3 (L2) vein via Dpp (blue). Dpp diffuses from the AP boundary and activates Sal (orange) and Optix (green) at different concentration threshold. Sal represses the activation Optix at the cell's posterior to the vein L2. The cells in between Sal and Optix expresses knirps (co-repressed by Sal and Optix) which leads to the development of L2 provein cells. (ref). **(B)** Positioning of Bicyclus R2 vein via Dpp (blue) via the same mechanism as discussed above for *Drosophila*. **(C)** A mechanism similar to that in the AP wing network in involved in the eyespots. In the eyespots Dpp (blue) is expressed in the center which diffuses in the surrounding tissues activating Sal and Optix at different concentration thresholds. Sal represses the activation of *optix* in the black scale areas and as a result Optix is expressed only in the orange ring region.

A reaction-diffusion and positional information systems likely work together to form the eyespots in butterflies

The two big theories in biological patterning are reaction-diffusion and positional information. Reaction-diffusion involves interaction between at least two diffusible molecules (morphogens) while positional information can achieve patterning by diffusion of a single morphogen. It has been proposed

that the two mechanism could work together in one of three ways; (1) reaction-diffusion is upstream of positional information, where reaction-diffusion creates high and low levels of one of the interacting morphogen which then based on different concentration thresholds activate other downstream genes; (2) reaction-diffusion and positional information work in parallel where the information generated from both the mechanisms are used to create pattern; and finally (3) positional information is upstream of reaction-diffusion where the initial positions of the pattern are generated by a positional information mechanism upon which a reaction-diffusion mechanism acts (Green and Sharpe, 2015).

A reaction-diffusion mechanism is likely involved in setting up the center of the eyespots followed by a positional information mechanism patterning the rings of the eyespots. A previous study has proposed that a reaction-diffusion mechanism is involved in setting up the center of the eyespots (Connahs et al., 2019). By studying the gene expression patterns of Wnt (Armadillo: The signal transducer of Wnt signaling) and BMP (Dpp: the morphogen of BMP signaling) along with CRISPR-Cas9 data on *dll*, researchers have proposed that a three molecule substrate depletion reaction-diffusion system is involved in setting up the center of the eyespots during the fifth instar larval wing development. The study showed that the BMP (Dpp) module (absent in the center of the eyespot) acts as a substrate for the Wnt module which is expressed in the center of the eyespot together with Dll. In this study the simulation experiments by gray-scott model accurately replicated the observed *dll* CRISPR mutant phenotypes. The reaction-diffusion system created high

levels of Wnt signaling in the eyespot centers and low levels in the surrounding area.

The high level of Wnt signaling is likely involved directly or indirectly in the activation of *dpp* (absent in the eyespot center during the larval stage) during later developmental stages. The activation of *dpp* could be due to the presence of Cubitus interruptus (Ci) in the eyespot center (Keys et al., 1999) set up by Wnt signaling. In such a scenario a gene circuit similar to the *D. melanogaster* wing venation circuit (Blair, 2007) could be involved in patterning the eyespot rings. In such a mechanism, En activates a short-range morphogen *hedgehog* (*hh*) and Hh activates Ci which then activates *dpp* (Brook et al., 1996). This mechanism has been proposed to be conserved across wing, leg and antennal imaginal discs in *D. melanogaster* (Raftery et al., 1991) and might have been co-opted to the eyespots. The wing venation circuit was previously proposed to be involved in the development of eyespot in *J. coenia* once three of these genes, Hh, En, and Ci, had been visualized in and around the eyespot centers (Keys et al., 1999).

In *B. anynana,* Wg (Wnt1) likely activates *dpp* in the eyespot center. Wg (Wnt) signaling has been shown to be involved in the activation of *en* in *D. melanogaster* embryo (Sanson et al., 1999), where Wg (Wnt1) secreted from the cells binds to the Frizzled receptor in the nearby cells (Müller et al., 1999; Sanson et al., 1999). The binding of Wg (Wnt1) results in the activation of *en* by the Armadillo (Arm) signal transducer (Van de Wetering et al., 1997). In *B. anynana* pupal wings we observe the expression of both *wg* and its signal transducer Arm in the center of the eyespot (**Fig. 3.3A, B and K**). The expression of Arm overlaps with the cells that express En in the center of the

eyespot (**Fig. 3.3M**) indicating that a similar mechanism as explained earlier might be in play, where Arm activates *en*. En is likely involved with the regulation of the active form of Cubitus interruptus (Ci) either via a Hedgehog (Hh) dependent or an Hh independent mechanism. In *B. anynana* no Hh has been observed yet in the center of the eyespots but Ci is expressed at high levels in these central cells (Keys et al., 1999), and Ci is required for the activation of *dpp* (Hepker et al., 1999). The high level of Ci can activate *dpp* in the eyespot center (**Fig. 3.3C and D**) and the diffusion of Dpp from the eyespot center then likely patterns the rings of the eyespot via a positional information mechanism activating *spalt* and *optix* at different concentrations as mentioned in the discussion section above. The reduction of eyespot size due to *wg* (*wnt1*) RNAi (Özsu et al., 2017) also supports such a mechanism where reduction in the level of Wg (Wnt1) will reduce the level of Dpp secreted from the center of the eyespot, which will result in a smaller domain over which Dpp can activate *spalt* and *optix*.

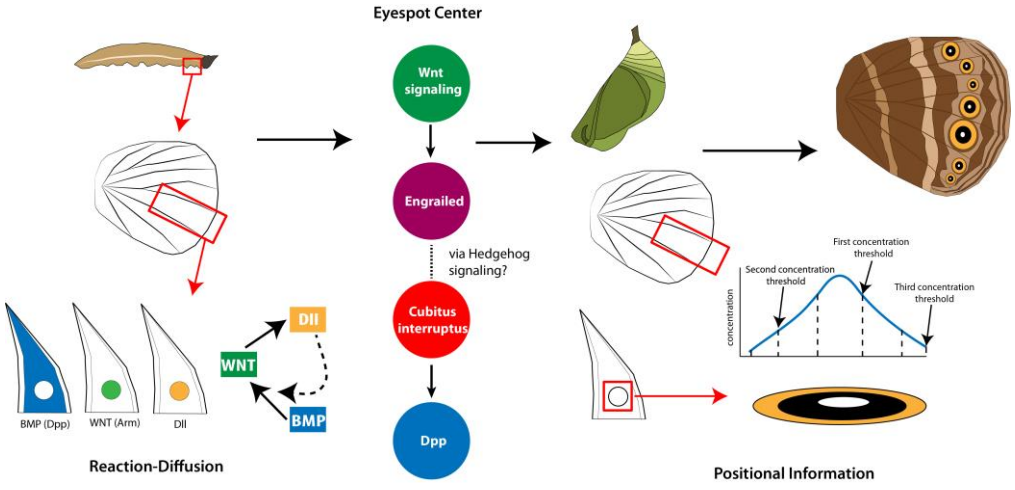



**Figure 3.6. A model of Reaction-diffusion followed by positional information involved in the development of eyespots in *B. anynana* butterflies.** During the development of the larval wing BMP (Dpp) signaling

is observed everywhere in the wing compartment except for the eyespot center and the veins, WNT (Arm) signaling is observed in the center of the eyespot overlapping with the expression of Distal-less (Dll). A model has been proposed where these three signaling modules interact via the process of reaction-diffusion that lead to the positioning of the eyespot center. Wnt signaling likely activates Dpp in the eyespot center. After the center is formed, during the late larval/pupal stage of wing development a positional-information comes into play, where a central morphogen, likely Dpp diffuses to the surrounding tissues and activates genes responsible for the formation of the rings.

## Conclusion

We found that *optix* is involved in both pigmentation and scale structure development in the orange ring region of the eyespot in *B. anynana* butterflies. Knocking-out *optix* results in the transformation of scale color from orange to brown as a result of loss of pigmentation and changes in scale ultrastructure. Spalt is involved in the repression of *optix* in the cells involved in the formation of the black scales and as a result Optix is present only in the orange ring of the eyespot, where Spalt is absent. We propose that Dpp a morphogen present in the center of the eyespot is likely involved in the activation of both *optix* and *spalt* via a possible positional-information mechanism, where Dpp acts as a central morphogen activating *spalt* at a higher concentration threshold and *optix* at a lower threshold. Previous work has shown that a reaction-diffusion mechanism is likely involved in setting up the center of the eyespot in *B. anynana* larval wings. After the eyespot centers are determined, a positional information mechanism most probably comes into play coloring the rings of the eyespot, indicating that the two mechanisms of biological patterning are possibly working in sequence to pattern these magnificent marvels of nature.

**Acknowledgement**

This work was supported by the National Research Foundation, Singapore, award NRF-NRFI05-2019-0006. TDB was supported by a Yale-NUS scholarship. We would like to thank Robert Reed for the anti-Optix antibody; and Lee Ka Yau (SEM facility, Department of Chemistry, NUS) and Anupama Prakash for carrying out the SEM imaging. We would like to thank DBS-CBIS confocal facility and Tong Yan for access and help with the confocal microscopes. The authors declare no conflict of interest.

**Authors contribution**

TDB designed the experiments, performed the experiments, analysed the results, and wrote the manuscript, with input from AM. Both authors read and agreed to the final version of the manuscript.

**Supplementary materials**

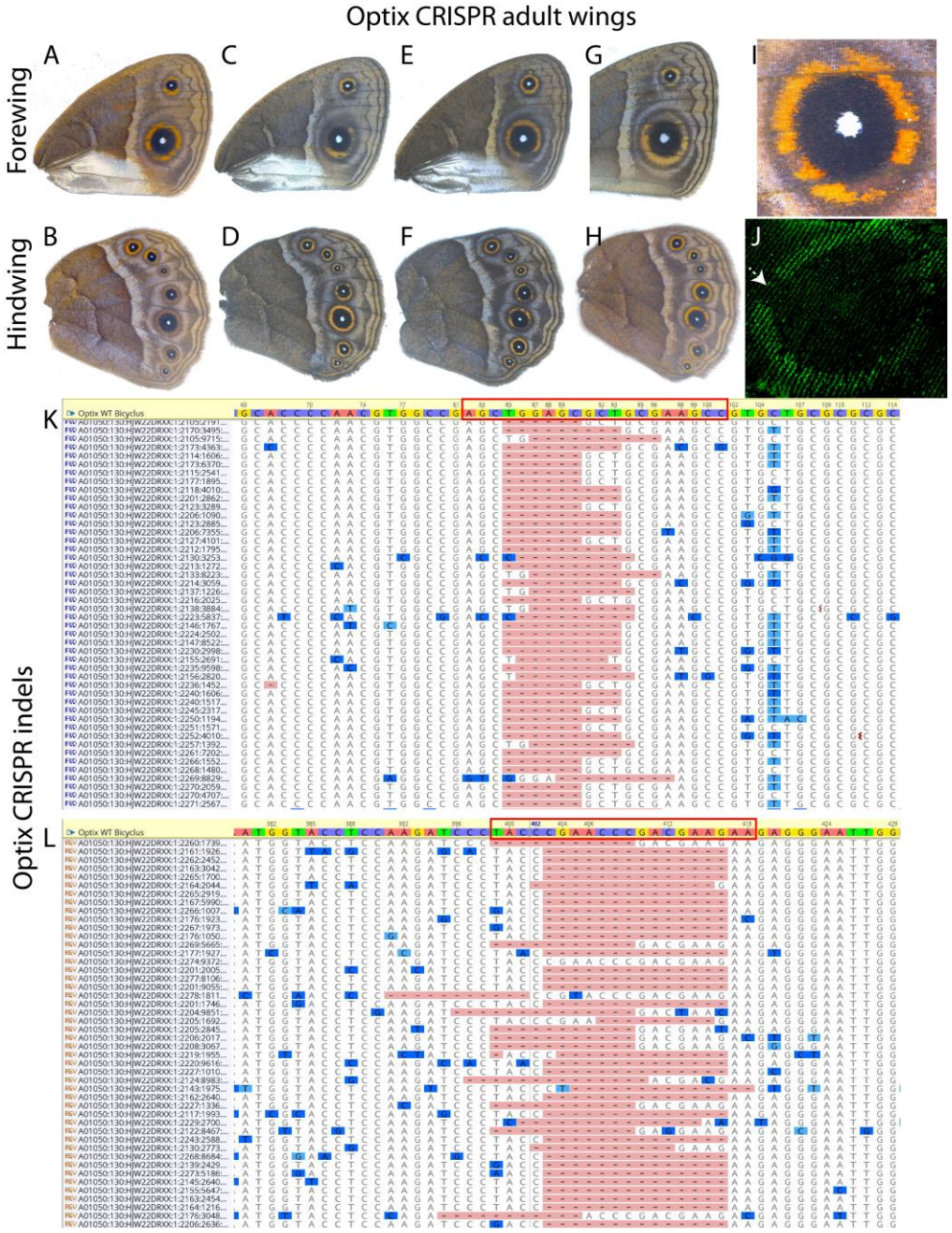



**Figure S3.1. Optix CRISPR in *Bicyclus anynana* butterflies. (A-H)** *optix* CRISPR adult wings. **(I)** optix CRISPR eyespot with conversion of orange scales into brown scales. **(J)** Antibody staining of Optix at the eyespot of *optix* CRISPR individual. **(K and L)** Deletions at the two sites targeted for *optix* CRISPR.

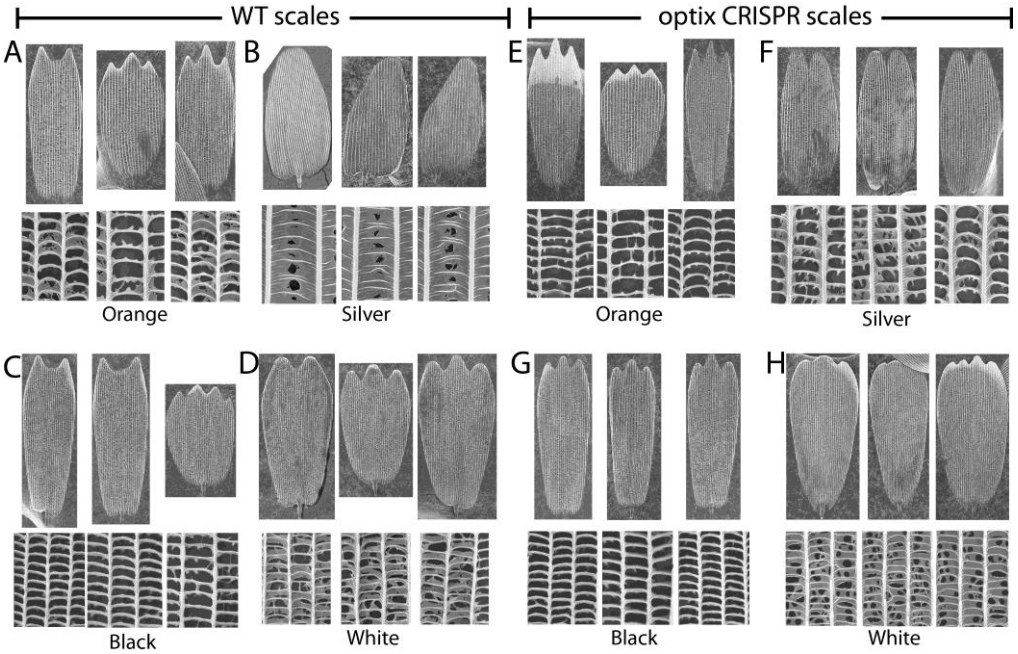



**Figure S3.2. WT and Optix CRISPR scale ultrastructure.** WT orange and silver scales (**A and B**) have higher amount of upper lamina in between the cross-ribs compared to optix CRISR orange and silver scales (**E and F**). There are no changes in the scale structure of black and white scale structures in between the WT and *optix* CRISPR (**C, D, G and H**).

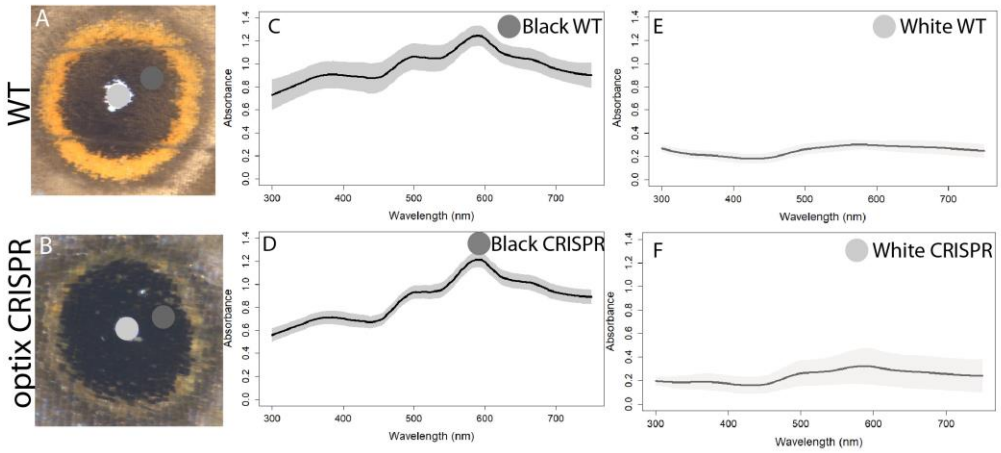



**Figure S3.3: Absorbance spectra of WT and optix CRISPR scales**. (**A**) WT eyespot. (**B**) optix CRISPR eyespot. Absorbance spectra from (**C)** black and (**E**) white scale regions of **A**. Absorbance spectra from (**D**) black and (**F**) white scale regions of **B**.

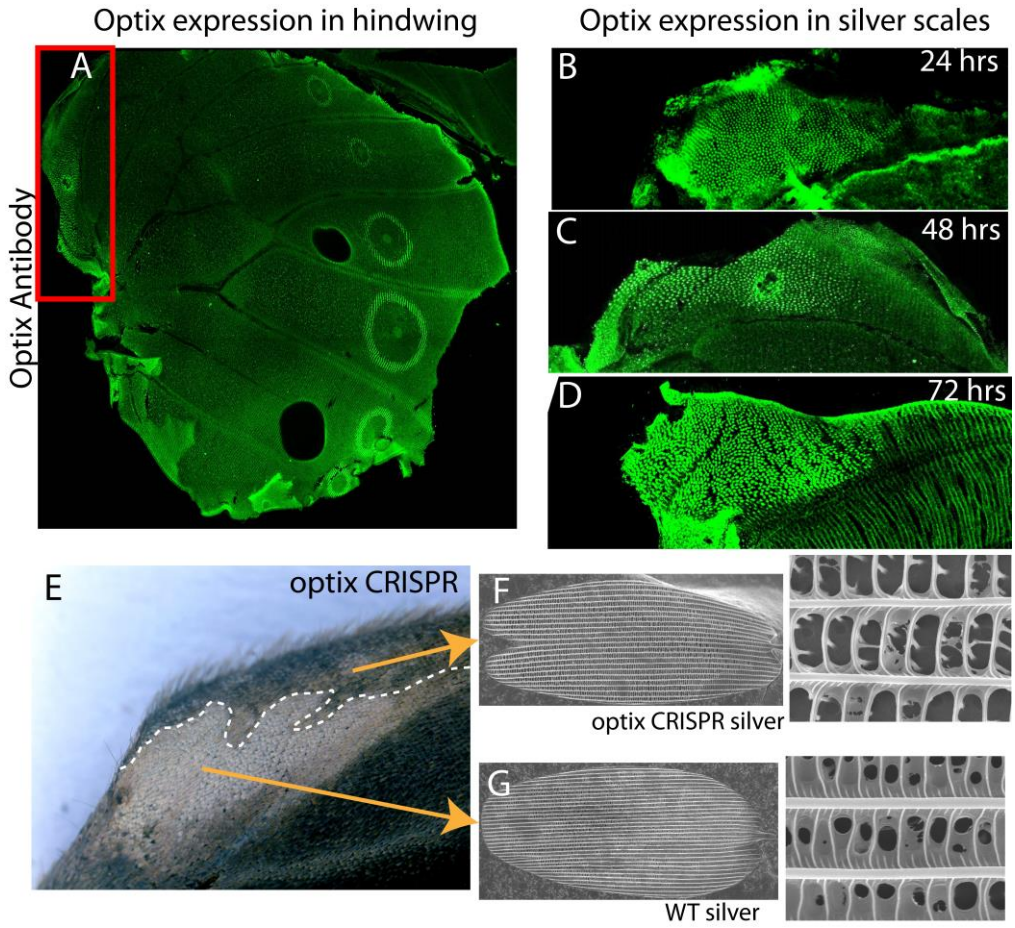



**Figure S3.4. Optix is involved in silver scale development of *B. anynana* wings with a partial upper lamina. (A)** Hindwing showing expression of Optix in the future silver scale region (red rectangle). Expression of Optix at **(B)** 24 hrs, **(C)** 48 hrs and **(D)** 72 hrs pupal wing. **(E)** Optix crispant individual with both WT silver scales and modified scales. **(F)** Ultrastructure of optix CRISPR silver scale. **(G)** Ultrastructure of WT silver scales. Knocking out *optix* results in the reduction of the upper lamina in between cross-ribs in orange scales.

**Table S3.1. Primer table**

| Sl. No. | Primer Name | Sequence | Description |
|---|---|---|---|
| **1.** | Sal_CRISPR_Guide | GAAATTAATACGACTCACTATAGGTGATCGAGCCGGCGTTGAGTTTTAGAGCTAGAAATAGC | Forward primer for guide synthesis to knockout *sal* |
| **2.** | Optix_CRISPR_Guide_1 | GAAATTAATACGACTCACTATAGGGGCTTCGCAGCGCTCCAGCTGTTTTAG | Forward primer 1 for guide synthesis to knockout *optix* |

| | | | |
|---|---|---|---|
| | | AGCTAGAAATAGC | |
| **3.** | Optix_CRISPR_Guide_2 | GAAATTAATACGACTCACTATAGGTTCTTCGTCGGGTTCGGGTAGTTTTAGAGCTAGAAATAGC | Forward primer 2 for guide synthesis to knockout *optix* |
| **4.** | CRISPR_Guide_R | AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAAC | Reverse primer for guide synthesis CRISPR guides |

**Table S3.2. Optix CRISPR-Cas9 injection table**

| Sl. No. | Concentration | Date | Eggs Injected | Hatchlings | % Hatchlings |
|---|---|---|---|---|---|
| **1.** | 300 ng/µl | 11th March 2020 | 785 | 85 | 10.8 |
| **2.** | 300 ng/µl | 12th March 2020 | 398 | 47 | 11.9 |
| **3.** | 300 ng/µl | 6th June 2020 | 326 | 65 | 19.9 |

**Table S3.3. Spalt CRISPR-Cas9 injection table**

| Sl. No. | Concentration | Date | Eggs Injected | Hatchlings | % Hatchlings |
|---|---|---|---|---|---|
| **1.** | 300 ng/µl | 28th Sept 2018 | 302 | 48 | 15.9 |
| **2.** | 300 ng/µl | 10th Oct 2018 | 306 | 25 | 8.2 |
| **3.** | 300 ng/µl | 11th Nov 2018 | 120 | 18 | 15.0 |

| | | | | | |
|---|---|---|---|---|---|
| **4.** | 300 ng/µl | 9th Feb 2019 | 135 | 8 | 5.9 |

Sequence of *decapentaplegic* used for *in-situ* hybridization.

GTTCTTCAACGTAAGCGGCGTACCGGCCGACGAGGTGGCGCGCGG
CGCCGACCTCTCGTTCCAACGAGCCGTCGGCACCACCGGCAGACAG
AGACTGTTGTTGTACGACGTGGTGCGCCCTGGCCGCCGCGGCCACT
CCGAGCCGATCCTGCGGCTGCTGGACTCCGTTCCGCTCCGGCCCGG
GGAGGGAATCGTCAACGCCGACGCTCTGGGAGCGGCGCGACGGTG
GCTCAAAGAGCCCAAACATAATCACGGACTATTAGTGCGAGTGTTA
GAAGAAGACGCCGCGAGTGCGAGCAGGGACGCGAAGTTCCCGCAC
GTGCGCGTGCGCAGACGCGTCACGGACGAGGAGGAGGAGTGGCGG
ACGGCGCAGCCGCTGCTCATGCTGTACACGGAGGACGAGCGCGCG
CGCGCGTCGCGGGAGACGAGCGAGCGGCTGACGCGCAGCAAGCGC
GCGGCGCAGCGGCGGGGGCACCGCGCGCACCACCGCCGCAAGGAG
GCGCGCGAGATCTGCCAGCGCCGCCCGCTGTTCGTCGACTTCGCGG
ACGTGGGCTGGAGCGACTGGATCGTGGCCCCGCACGGCTACGACG
CGTACTACTGCCAGGGCGACTGCCCCTTCCCGCTGCCGGACCACCT
CAACGGCACGAACCACGCGATAGTGCAGACTCTGGTCAACTCAGT
GAACCCCGCGACGGTGCCCAAAGCGTGCTGCGTGCCGACGCAACT
CTCATCTATATCTATGTTATATATGGACGAAGTGAACAATGTGGTG
CTTAAAAACTATCAGGACATGATGGTGGTAGGCTGTGG

Region of *optix* used for CRISPR-Cas9 (Highlighted in red)

ATGCGCGGCTCCTGGGACGAGTCCACGACGGCGGCGCTGCACGCG
CGCATCCTGGAGGCGCACCGCGGGTCCGCCGCGCCCGACCGCGCC
GAGCCCGCGTGCGAGCCTCCGCCGCTGACGCTGGGCGCGCTGGAG
CTGGCGGCGCCCACGCCGCTGCTGCCGCTGCCCACGCTGAGCTTCA
GCGCCGCGCAGGTGGCCACCGTGTGCGAGACGCTGGAGGAGAGCG
GCGACGTGGAGCGCCTGGCGCGCTTCTTGTGGTCGCTGCCCGTGGC
GCACCCCAACGTGGCCGAGCTGGAGCGCTGCGAAGCCGTGCTGCG
CGCGCGCGCCGTCGTCGCCTTCCACGCCGGCCGCCACCGCGAGCTG
TACGCCATCCTCGAGCGCCACCGCTTCCAGCGCTCCAGCCACGCCA
AGCTGCAAGCGCTGTGGCTGGAGGCGCACTACCAGGAGGCTGAGC
GCCTGCGCGGCCGTCCGCTGGGCCCCGTCGACAAGTACCGCGTGCG
GAAGAAGTTCCCGCTCCCGAGGACGATCTGGGACGGCGAGCAGAA
GACGCACTGTTTCAAGGAGCGGACGCGATCTCTACTCCGAGAATGG
TACCTCCAAGATCCCTACCCGAACCCGACGAAGAAGAGGGAATTG
GCGGCGGCGACGGGTCTGACGCCGACGCAAGTCGGCAACTGGTTC
AAAAACCGACGGCAAAGAGACCGAGCGGCCGCCGCCAAGAACCGC
TCCGCCGTGCTGGGCAGAGGATAA

Region of *spalt* targeted by CRISPR-Cas9 (location of guide RNA highlighted in red)

GCATCGACAAGATGCTGAAAATAATAATAGTCTCGAAGACGGCGA
GGCCGAAATACCTGAAGCCGACATGCCCCCCGTGGGTCTGCCGTTC
CCTTTGGCAGGACACGTTACTCTTGAGGCTCTACAAAATACGAGAG
TAGCGGTCGCCCAATTCGCTGCAACAGCGATGGCAAATAATGCGA
ATAACGAAGCTGCTATACAAGAATTACAAGTGTTACACAACACTCT
ATACACTTTACAGTCACAACAAGTATTTCAACTTCAGTTAATACGT
CAGCTTCAGAATCAGTTATCTCTAACTCGACGGAAAGAAGACGATC
CACACAGCCCACCGCCAAGTGAACCAGAACAGAATGCCCCGTCAA
CGCCGGCTCGATCACCGTCGCCGCCGCGTCCGCCACGGGAGCCGTC
GCCTGTTATACCCTCTCCTCCTACTAGCCAAAGTTTGCCGTCGACTC
ACACACATCACACACCCAAAACTGAACAGATATCTATCCCTAAGAT
TCCAACTTCCTCACCATCTTTAATGACCCACCCACTTTATAGTTCAA
TTTCTTCGTCATTAGCATCTTCCATCATAACAAACAATGATCCTCCA
CCGTCCCTAAATGAA

R-code for the absorbance spectra analysis

```
library(pavo)

specs <- getspec( where = getwd(), ext = "txt", lim =
c(300, 750))

plot(specs, ylab = "Absorbance", col = rainbow(3))

specs <- procspec (specs, opt='smooth')

specs

spp <- substr(names(specs), 1, 4)

aggplot (specs, by=spp, ylab = "Absorbance")


#Plot for orange spectra

specs.orange <- subset(specs, "orange")

spp.orange <- substr(names(specs.orange),1,4)

aggplot(specs.orange, by=spp.orange, ylab = "Absorbance",
lcol = "darkgoldenrod4", shadecol = "gold", ylim=c(0.0,
1.4), cex.axis = 1.3, cex.lab = 1.5, lwd = 4)


#Plot for orangecrispr spectra

specs.orangecrispr <- subset(specs, "orangecrispr ")

spp.orangecrispr <- substr(names(specs.orangecrispr),1,4)

aggplot(specs.orangecrispr, by=spp.orangecrispr, ylab =
"Absorbance", lcol = "Black", shadecol = "Brown",
ylim=c(0.0, 1.4), cex.axis = 1.3, cex.lab = 1.5, lwd = 4)
```

```
#Plot for black spectra

specs.black <- subset(specs, "black")

spp.black <- substr(names(specs.black),1,4)

aggplot(specs.black, by=spp.black, ylab = "Absorbance",
lcol = "Grey0", shadecol = "Grey3", ylim=c(0.0, 1.4),
cex.axis = 1.3, cex.lab = 1.5, lwd = 4)


#Plot for brown spectra

specs.brown <- subset(specs, "brown")

spp.brown <- substr(names(specs.brown),1,4)

aggplot(specs.brown, by=spp.brown, ylab = "Absorbance",
lcol = "Black", shadecol = "Brown", ylim=c(0.0, 1.4),
cex.axis = 1.3, cex.lab = 1.5, lwd = 4)


#Plot for white spectra

specs.white <- subset(specs, "white")

spp.white <- substr(names(specs.white),1,4)

aggplot(specs.white, by=spp.white, ylab = "Absorbance",
lcol = "Grey40", shadecol = "seashell3", ylim=c(0.0,
1.4), cex.axis = 1.3, cex.lab = 1.5, lwd = 4)
```

## Chapter 4: Expression of multiple *engrailed* family genes in eyespots of *Bicyclus anynana* butterflies does not implicate the duplication events in the evolution of this morphological novelty

Banerjee, T. D., Ramos, D. and Monteiro, A. (2020). Expression of multiple *engrailed* family genes in eyespots of *Bicyclus anynana* butterflies does not implicate the duplication events in the evolution of this morphological novelty. *Front. Ecol. Evol*. 8, 1–12.

### Abstract

Gene duplication events often create genetic redundancy that can either lead to the appearance of pseudogenes or, instead, create opportunities for the evolution of novel proteins that can take on new functions. One of the genes which has been widely studied with respect to gene duplication is *engrailed* (*en*). En-family proteins are expressed in a morphological novelty, eyespots (in the center and in the outer gold ring), in the African squinting bush brown butterfly *Bicyclus anynana*, as well as in a more conserved pattern, the posterior compartment of a wing. In the present study, we used whole-genome sequencing and transcriptome data to show the presence of three *en*- family genes and their differential expression on the pupal wings of *B. anynana* using *in-situ* hybridization. The results suggest two duplication events of en-family genes, the first evidence of a two-fold duplication in the Lepidoptera. We propose that all copies initially had posterior wing compartment expression and all copies subsequently gained a novel expression domain associated with eyespot centers. Two copies secondarily lost the posterior compartment expression, and one copy alone gained the outer ring expression domain. By dating the origin of both duplication events, however, we conclude that they

predate the origin of eyespots by at least 60 mya, and hence our data does not support the retention of the multiple *en* gene duplicates in the genome via their involvement with the novel eyespot evolutionary innovation.

**Introduction**

Gene duplication plays an essential role in the evolution of novel proteins, and this process has also been proposed to lead to the evolution of major morphological innovations such as flowers; the chambers of the heart; and brains, bones, and cartilage in vertebrates (Conant and Wolfe, 2008; Olson, 2006; Theißen, 2001; Wagner and Lynch, 2010; Zhang et al., 2015). When genes duplicate, there is initially genetic redundancy which allows one of the copies to quickly accumulate mutations and become converted into a pseudogene (Lynch et al., 2001; Zhang, 2003). However, genetic redundancy can also lead to opportunity. When gene duplicates remain intact during the evolutionary processes this usually means one of three things: 1) that either the double dose of RNA and/or protein increased fitness (Zhang, 2003); 2) that each copy divided the original functions and/or expression domains of the protein among themselves, a process known as sub-functionalization that reduces the total number of functions taken over by each copy and lessens pleiotropy (Force et al., 1999); 3) or that one of the copies gained a novel function, known as neo-functionalization (Force et al., 1999). Examples include sub-functionalization of the gene *engrailed (en)* in zebrafish, which resulted in the differential expression of the copies in the pectoral appendage, and in neurons (Force et al., 1999); and neo-functionalization of eosinophil-derived neurotoxin (EDN) and eosinophil cationic protein (ECP) in humans and old world monkeys from EDN, where after the duplication ECP developed

a novel anti-pathogenic function (Zhang et al., 1998). Neo-functionalization has been of particular interest since the origin of novel gene duplicates has been loosely associated with the origin of evolutionary innovations (Conant and Wolfe, 2008; Wagner and Lynch, 2010), such as with beetle horns (Zattara et al., 2016), human vision (Yokoyama and Yokoyama, 1989), and betalain pigments in plants (Brockington et al., 2015).

The gene *engrailed* (*en*) has been heavily studied with respect to gene duplications(Force et al., 1999). Duplication of *en* in several independent metazoan lineages has generated a family of related genes, varying from one to four copies, where the paralogs display varying degrees of divergence in expression and function (Abzhanov and Kaufman, 2000; Damen, 2002; Gibert et al., 2000; Marie and Blagburn, 2003; Peel et al., 2006; Peterson et al., 1998).

In addition to its conserved function in segmentation during invertebrate embryogenesis(Brown et al., 1994; Fujioka et al., 2002; Manzanares et al., 1993), *en* is also associated with morphological novelties in several species. For instance, *en* is involved in neurogenesis (Patel et al., 1989), in determining the fate of glial and neuronal cells in grasshopper median neuroblast(Condron et al., 1994), in axonal targeting in centepedes (Whitington et al., 1991), and in the patterning of insect wing veins (Banerjee and Monteiro, 2020b; Guillén et al., 1995). In addition, while no function is yet known, *en* is expressed in precursor cells that build a mollusc's shell (Nederbragt et al., 2002), in the tentacles of the cephalopod *Sepia officianalis* (Baratte and Bonnaud, 2009), in bacteriocytes of the aphid, *Acyrthosiphon pisum* (Braendle et al., 2003), and in

the eyespots of saturniid moths (Monteiro et al., 2006) and nymphalid butterflies (Brunetti et al., 2001; Keys et al., 1999).

Eyespots in butterfly wings are novel traits that appear to be specified by old transcription factors and signaling pathways, deployed in novel ways (Brunetti et al., 2001; Keys et al., 1999; Monteiro, 2015; Oliver et al., 2012). Using an antibody that detects a variety of Engrailed proteins (Patel et al., 1989), Keys et al. (1999) found that En-family proteins are expressed in both conserved and novel domains in butterfly wings (Keys et al., 1999). En-family proteins maintained their conserved expression domain, and likely their role, in establishing the posterior compartment of each embryonic body segment including that of the wings during the larval stage (Banerjee and Monteiro, 2020b; Carroll et al., 1994; Keys et al., 1999). However, one or more of the genes encoding En or its paralog Invected (Inv) were also expressed in the eyespot centers, in late larval and pupal wings, and in a pattern mapping to the ring of gold scales during the pupal stage of wing development in *Bicyclus anynana* butterflies (Brakefield et al., 1996; Brunetti et al., 2001). A detailed investigation of the number of *en* family gene copies present in the genome of *B. anynana*, however, and of the expression domains of each paralog in eyespot patterns is still missing.

Here we provide additional insights on the association between the retention of duplicated genes and their involvement in the origin of novel traits by detailing the expression domain of multiple *en* copies in butterfly wings, asking whether one or more copies are exclusively associated with the eyespot patterns, and testing whether the gene duplication events are closely associated with the origin of eyespots. We report the presence of three *en*-family genes in

*B. anynana*, two of which contain all the five major domains of En-family proteins (Peel et al., 2006) and one contains four of these domains. We used sequencing data, to analyze the phylogenetic relationship between the copies, and whole-mount *in-situ* hybridizations to examine the expression of all three copies in the pupal wings. Two of the copies are closely related to each other and have a single expression domain on the pupal wing, in the eyespot centers, while the other copy displays three wing expression domains, in the posterior compartment, in the eyespot centers, and in the orange ring. The timing between the last *en*-family gene duplication and the origin of eyespots, however, spans at least 60 mya. We discuss the implications of our data on the mechanisms of evolution of these morphological novelties.

## Methods

### Rearing *Bicyclus anynana*

*B. anynana* butterflies were reared in the lab at 27°C, 60% relative humidity, and 12-12 hrs day-night cycle. Larvae were fed young corn leaves and adults were fed mashed bananas.

### Mining for *en* paralogs and splice variants

The sequences of *en*-family genes and splice variants of *B. anynana* were obtained via whole genome sequencing data (*Bicyclus anynana* v1.2 (Nowell et al., 2017)), transcriptome (*B. anynana*; eyespot (Özsu and Monteiro, 2017) and pupal brain (Macias-Munoz et al., 2016)), and 5’ and 3’ RACE (Rapid Amplification of cDNA Ends). Protein sequences of En and Inv of *D. melanogaster* were obtained from FlyBase. En-family protein sequences of

other Lepidopteran species were obtained from Lepbase (Challis et al., 2016), NCBI, EMBL and transcriptome data (Connahs et al., 2016; Daniels et al., 2014; Ferguson et al., 2010) and analysed for conserved common and paralog-specific domains.

Sequence alignment and phylogenetic tree construction

Nucleotide sequence alignment was carried out using ClustalW (Thompson et al., 1994) with the default parameters in 'SLOW/ACCURATE' pairwise alignment option in GenomeNet and visualized in Geneious v10.1.3 (Kearse et al., 2012). Nucleotides were converted into protein sequence via NCBI ORF finder. The protein sequences were aligned via ClustalW (Thompson et al., 1994) with gap open cost: 10, and gap extend cost 0.1; and afterwards, manually edited and visualized in Geneious v10.1.3(Kearse et al., 2012).

A maximum likelihood tree was inferred using RAxML v8.1.20 ran with model PROTGAMMAJTT and default parameters with 100 bootstraps(Stamatakis, 2014) using ETE v3.1.1(Huerta-Cepas et al., 2016) implemented on GenomeNet. The same tree topology was also obtained using the following other methods and software:

a) PhyML v20160115 ran with model JTT and parameters: -f m --nclasses 4 --alpha e --pinv e -o tlr --bootstrap 100(Guindon et al., 2010), using ETE3 v3.1.1(Huerta-Cepas et al., 2016);
b) fasttree with slow NNI and MLACC=3 (to make the maximum-likelihood NNIs more exhaustive) (Price et al., 2009) using ETE v3.1.1(Huerta-Cepas et al., 2016); and

c) geneious tree builder using Jukes-Cantor as the genetic distance model, UPGMA as tree build method, and 1000 bootstraps(Kearse et al., 2012).

Probe design of the *en*-family genes

Sequence used for en probe: A 628 bp region in the exon 1 of en CDS was chosen for designing a probe. This region is absent in *inv* and *inv-like* and does not align well with these genes (~35% bp similarity) (**Fig S4.5**). Sequences used for *inv* and *inv-like* probes: The regions used

for probe preparation are Exon 1 of *inv*, which is not present in *inv-like*, and the 5? UTR of *inv-like*, which is not present in inv. The sequence used for the inv probe has 24.9% identical bps to that of the en gene sequence, and 36.03% identical bps to that of the *inv-like* gene sequence in their overlapping regions. The sequence used for the *inv-like* probe has 30.2% identical bps to that of the en gene sequence, and 36.8% identical bps to that of the *inv* gene sequence in the overlapping region (**Fig S4.5**). This level of sequence similarity is lower than that observed between highly unrelated genes such as en and optix, which share 41.2% identical sequences. The presence of a very low sequence identity and a small overlap as observed in Supplementary Figure S5 makes it unlikely for the two probes to produce results due to cross-reaction. Control stainings: *inv-like* and *inv* are mostly identical in the

CDS region. Due to this limitation, we used a 182 bps long region in the 5' UTR of inv-like (**Fig 4.1a**) which is small in size and can produce non-specific stainings. To test for presence of non-specific stainings we performed *inv-like* sense stainings using a probe prepared with the same sequence (**Fig S4.1p, q**).

Sense stainings were also performed for *en* and *inv* (**Fig S4.1d–f, j–l**). Prior to probe preparation, amplified fragments of each paralog were sequenced and aligned to ensure that the correct amplicons were used for *in situ* hybridization.

*In-situ* hybridization

One to two days old pupal wings, timed for moment of pupation using a time-lapse camera, were dissected in 1X PBS at room temperature based on a previously published protocol(Banerjee and Monteiro, 2020a) and transferred into 1X PBST with 4% formaldehyde for 30-35 mins. Afterwards, the wings were washed three times in 1X PBST and treated with 20 mg/ml proteinase K (NEB, P8107S) in 1 ml 1X PBST and then with 2 mg/ml glycine in 1X PBST. Wings were then washed three times in 1X PBST and gradually transferred into a prehybridization buffer (composition in Table S4) and incubated at 60°C for one hour prior to transfer into hybridization buffer (composition in **Table A1**) supplemented with 100ng/µl of probe. The wings were incubated in hybridization buffer at 60°C for 16-20 hrs and washed 5 times with prehybridization buffer at 60°C to remove any non-specific staining. To prevent cross-hybridization of the probes, experiments were performed on separate days. Afterwards, wings were brought to room temperature and gradually transferred to 1X PBST and then incubated in block buffer (composition in Table S4) for 1 hr. 1:3000 dilution of anti-Digoxygenin was then added and incubated for 1 hr followed by 5 washes with block buffer. The wings were then transferred to an alkaline-phosphatase buffer (composition in **Table A1**) supplemented with NBT/BCIP. Imaging was done under a Leica DMS1000 microscope.

Antibody staining

One to two days old pupal wings, timed using a time-lapse camera, were dissected in 1X PBS and immediately transferred into a fixation buffer supplemented with 4% formaldehyde at room temperature (composition in **Table A2**) for 30-35 mins. Wings were then moved to ice and washed four times with 1X PBS and transferred to block buffer for one day (composition in Table S5). The next day wings were incubated in primary antibodies against En/Inv (1:15, mouse 4F11, a gift from Nipam Patel(Patel et al., 1989)), followed by anti-mouse AF488 (Invitrogen, #A28175) secondary antibody. Wings were then washed and mounted on an inhouse mounting media (composition in **Table A2**) and imaged under an Olympus fv3000 confocal microscope.

**Results**

Three *en*-family copies are present in the genome of *Bicyclus anynana*

Whole-genome sequencing(Nowell et al., 2017) and transcriptomic data(Macias-Munoz et al., 2016),(Özsu and Monteiro, 2017) revealed three *en*-family genes present in the genome of *B. anynana* within a single scaffold. We named the first copy *engrailed* (*en*) based on the presence of specific sequence domains described below. It has three exons and three splice variants. The 5' UTR of the splice variants are the same, but the 3' UTR region has different lengths across the three variants. The sequences are present between 136,804 and 158,542 bps of Scaffold 448: *Bicyclus anynana* v1.2 (**Fig 4.1a**). The transcript is translated into a 352 aa sequence containing all the five conserved domains of En-family proteins and the two En specific domains(Hui et al., 1992) (**Fig 4.2a**). The second copy we named *invected* (*inv*). It has two exons

present between 249,290 and 235,356 bps of the same scaffold. The transcript is translated into a 479 aa sequence containing all the five conserved domains of En-family proteins and the Inv specific LVSG and RS domains(Hui et al., 1992; Peel et al., 2006). The third copy we named *invected-like* (*inv-like*). It has two exons with a 5' UTR as identified using a pupal brain (Macias-Munoz et al., 2016) and an eyespot-specific transcriptome (Özsu and Monteiro, 2017) (See suppl file for sequence). The sequence is present between 189,552 and 191,029 bps of the same scaffold (**Fig 4.1a**). The transcript is translated into a 125 aa sequence containing four domains of En-family proteins with high similarity to Inv (**Fig 4.2a** and results below for details).

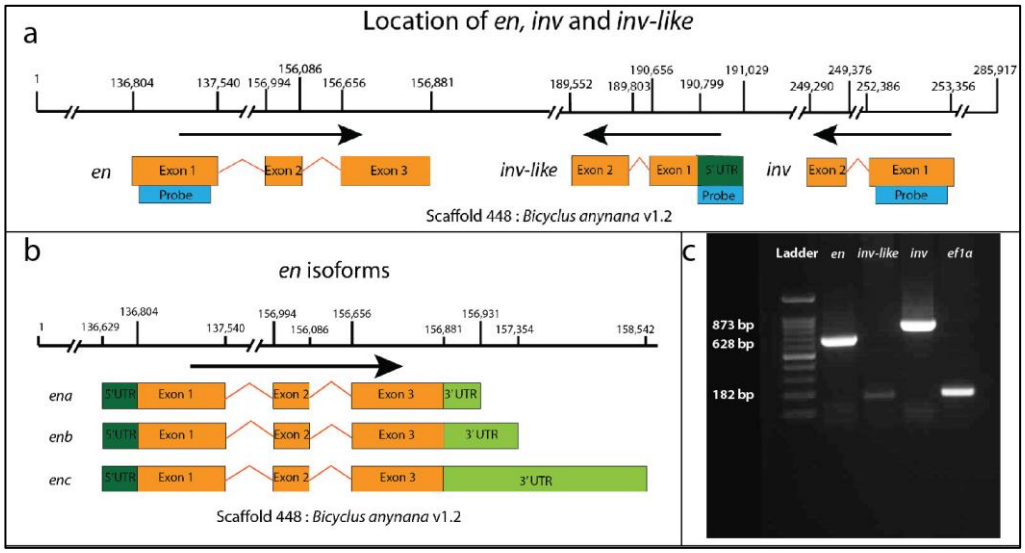



**Figure 4.1. Mapping of *engrailed* (*en*) family genes in the genome of *B. anynana*, *en* isoforms, and RT-PCR based verification of the three copies.** (a) Three *en*-like paralogs are present in the genome of *B. anynana*. All three copies are present close to each other in scaffold 448 of *Bicyclus anynana* v1.2 genome assembly (Nowell et al., 2017). Blue bars indicate sequences specific to each of the copies that were used for designing probes for *in-situ* hybridization. (b) Isoforms of *en*. Three isoforms of *en* are present in the scaffold 448. (c) RT-PCR based verification of *en*, *inv-like*, and *inv* in the pupal wings of *B. anynana*. Ef1α was used as a positive control. The probe for

exon 1 of *inv* is absent in *inv-like* and hence the chance of cross-reactivity is low.

Conserved domains of En-family proteins in *B. anynana*

En and Inv proteins contain all five domains of high conservation previously identified for En-family proteins, while Inv-like contains three full domains and one partial domain (**Fig 4.2a**) (Logan et al., 1992). The function of most of these regions has been determined: EH1 and EH2 are binding sites for the transcriptional co-repressor Groucho (Tolkunova et al., 1998) and the co-activator Exd (Peltenburg and Murre, 1996) respectively, EH4 is the homeodomain (Fjose et al., 1985) and EH5 represses En targets (Han and Manley, 1993). The function of EH3 is not known. The EH1 domain of En and Inv are 13 aa each and share 61.5 % identity. The EH2 domain of En and Inv are 19 and 21 aa respectively and share 90.5 % identity, while Inv-like has only three aa from this domain. The EH3 domain is highly divergent among species and paralogs. En has 11 aa, while Inv and Inv-like have 15 aa each for this domain, respectively. The EH4 domain is a 60 aa homeodomain region which is highly conserved in the three paralogs. Inv and Inv-like are 100% identical in this domain while En has 95% identity with the other two. The EH5 domain is an 18 aa region which is also highly conserved. Inv and Inv-like are 100% identical in this region while En shares 94.4% identity with the other two (**Fig 4.2b**).

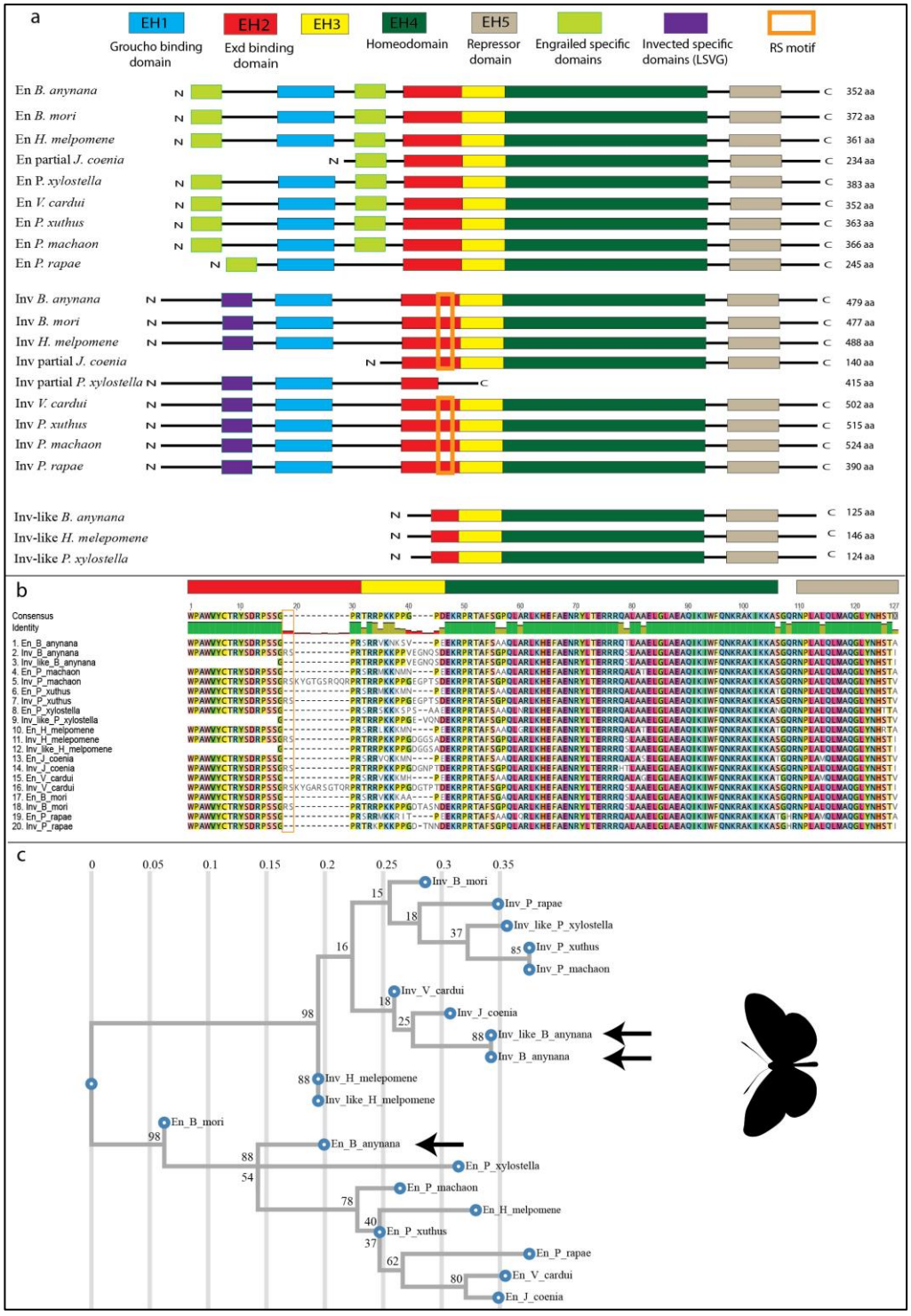



**Figure 4.2: Alignment and Phylogenetic tree of En-family proteins in Lepidoptera.** (**a**) En-family proteins have five conserved domains named EH1 to EH5. En and Inv contain all the five En-family domains along with gene-specific domains, while Inv-like only contains three complete domains (EH3 to EH5) and one partial domain (EH2). (**b**) Alignment of En-family proteins around the conserved EH4-EH5. (**c**) Phylogenetic tree of En-family protein created using RaxML (Stamatakis, 2014). En proteins cluster together; and Inv, and Inv-like cluster together. Arrows point to the *B. anynana* paralogs. The values on the top axis represent mean number of amino acid substitutions

per site and the node values represent branch support calculated based on 100 bootstraps.

Phylogenetic analysis of En-family proteins in Lepidoptera

Phylogenetic analysis of the highly conserved sequences from the EH2 to EH5 domains of En-family proteins from nine lepidopteran species using RaxML (Randomly Accelerated Maximum likelihood) (Stamatakis, 2014) resulted in the clustering of all En sequences as a sister lineage to Inv and Inv-like sequences in Lepidoptera (**Fig 4.2c**). The same results were obtained using PhyMl (Guindon et al., 2010) (**Fig S4.2**), fasttree (Price et al., 2009) (**Fig S4.3**), and Geneious Tree builder (Kearse et al., 2012) (**Fig S4.3**). The protein sequences of Inv and Inv-like at the conserved EH2 to EH5 have 98.9 % similarity; while En has 90.1% similarity to Inv, and 91.2% similarity to Inv-like. Furthermore, the nucleotide sequences of *inv* and *inv-like* are 98.25% similar; while *en* is 80.07% similar to *inv*, 81.81% similar to *inv-like* at the conserved EH4 and EH5 regions (**Fig S4.5; Table S4.1**). The presence of highly conserved sequences between Inv and Inv-like compared to En indicate that *inv* and *inv-like* underwent a more recent duplication.

Expression of *en* paralogs in the pupal wings of *B. anynana*

*In-situ* hybridization results indicate expression differences between the three *en* paralogs in the pupal wings of *B. anynana*. *en* is expressed in the posterior compartment of the wing, in the eyespot centers, and in the orange ring area surrounding each center (**Fig 4.3, d-f; Fig S4.1, a-c**). *inv* and *inv-like*, however, are only present in the eyespot centers (**Fig 4.3, i-n; Fig S4.1, g-i and m-o**).

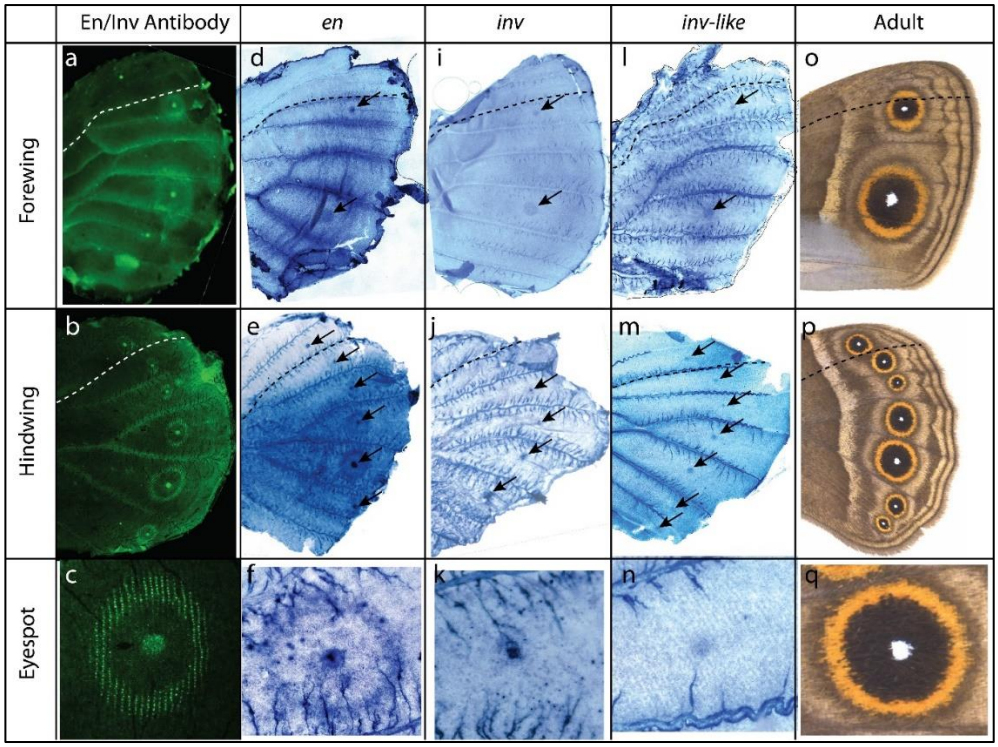



**Figure 4.3: Expression of *en*, *inv*, and *inv-like* in *B. anynana* butterflies' pupal wings.** (**a-d**) Antibody staining of En-like proteins using 4F11 antibody, En-family proteins are present in the posterior compartment, eyespot centers, and orange rings; (**a**) forewing, (**b**) hindwing, (**c**) eyespots. (**d-f**) *In-situ* hybridization of *en* using a specific probe. *en* is expressed in the posterior compartment of each wing, eyespot centers, and orange rings; (**d**) forewing, (**e**) hindwing, (**f**) eyespots. (**i-k**) *In-situ* hybridization of *inv* using a specific probe. *inv* is expressed in eyespot centers; (**i**) forewing, (**j**) hindwing, (**k**) eyespots. (**l-n**) *In-situ* hybridization of *inv-like* using a specific probe. *inv-like* is expressed in the eyespot centers. (**m**) forewing, (**n**) hindwing, (**o**) eyespots. (**o-q)** Adult phenotype. Anterior and posterior compartment are separated by dotted lines. Black arrows point to the eyespot centers. The staining in the veins and side trachea (as in panel **n**) are non-specific stains also observed with control probes (sense probes; **Fig S4.1 d-f, j-l, p and q**).

## Discussion

Duplication of *en* paralogs in holometabolous insects

Recent research proposed that a common ancestor to all Lepidoptera has undergone a large-scale genome duplication that is consistent with a whole-genome duplication, while hexapod ancestors to both Diptera and Lepidoptera have undergone two other large scale bursts of gene duplication(Li et al.,

2018). These duplications are believed to play important roles in the evolution of hexapods by increasing genetic redundancy and freeing up redundant copies to evolve new functions(Doyle and Coate, 2019; Li et al., 2018). The duplication of *en* into *en* and *inv* happened to a common ancestor to all hexapods including the holometabolous orders Diptera (containing *Drosophila*) and Lepidoptera (butterflies and moths) (Peel et al., 2006). Two of the families within the Lepidoptera, including Pieridae (containing *Pieris rapae*), and Nymphalidae (*Junonia coenia*, *Vanessa cardui*, and *B. anynana*) have butterflies whose genomes and transcriptomes have at least two copies of *en*-family genes (Challis et al., 2016; Connahs et al., 2016; Daniels et al., 2014) with characteristic En-family conserved domains. Inv has the conserved RS and LSVG motifs present across holometabolous insects (Peel et al., 2006), and En proteins one or both En-specific motifs (**Fig 4.2a**, see supplementary file for sequences and motifs).

The gene phylogeny suggested that the duplication of a common ancestor into *inv* and *inv-like* might have occurred independently in at least three different lineages: *B. anynana*, *Heliconius melpomene*, and *Plutella xylostella* (**Fig 4.2c**). This phylogenetic pattern, of Inv-like clustering with its paralog Inv in two of these lineages (*H. melpomene* and *B. anynana*) before clustering with its other orthologs (Inv-like), can also derive from the phenomenon of gene conversion (discussed below). Independent evidence using genomic copy position and orientation suggests that the duplication of these genes likely happened to a lineage ancestral to the divergence of all three species. The three *en*-like paralogs from *P. xylostella* map to contig unitig_1988 of *P. xylostella* pacbiov1 genome (Challis et al., 2016); and the same three paralogs

from *H. melpomene melpomene* map to scaffold Hmel207002 of *H. melpomene melpomene* Hmel2 genome (Challis et al., 2016; Davey et al., 2015). All three paralogs map next to each other and in the same genomic orientation as in *B. anynana* (**Fig S4.6 and S4.7**; **Table S4.1; Fig 4.1a**; sequence in suppl file). The conserved genomic architecture suggests that this gene duplication event happened only once in a common ancestor to the three lineages, rather than in the three different lineages independently, and this must have happened at least 130 mya, which is an estimated date for the split of *P. xylostella* moths from the two butterflies obtained from a recent phylogenomic study that included fossil calibrations (Kawahara et al., 2019).

Concerted evolution of En-family proteins in *B. anynana*

En-family proteins seem to have evolved under concerted evolution in *B. anynana* (as well as in *H. melpomene*). Gene duplicates that undergo concerted evolution are often found in close physical proximity and maintain high sequence similarity due to unequal crossing over and frequent gene conversion (Zhang, 2003). Due to concerted evolution high sequence similarity is maintained between En-family members which has often led to misleading phylogenetic trees (Abzhanov and Kaufman, 2000; Marie and Bacon, 2000; Peel et al., 2006; Peterson et al., 1998). In particular, higher sequence similarity is maintained between paralogs within a species when compared to similarity between orthologs across species, giving a false phylogenetic signal of multiple duplication events having happened independently in each lineage (Liao, 1999). Specific conserved LSVG and RS motifs in Inv paralogs, however, appear to be immune to gene conversion allowing researchers to conclude that the divergence of *en* and *inv* happened before the divergence of

hexapods (Peel et al., 2006). We have identified both LSVG and RS motifs in *B. anynana* and other Lepidopteran species (Fig 2a, b) but also observed clear signs of concerted evolution in that *B. anynana* Inv has more sequence similarity to *B. anynana* En than to Inv from *D. melanogaster* (Diptera) (see supplementary file for sequence). The same phenomenon of gene conversion is also likely happening to Inv and Inv-like within Lepidoptera, as discussed above.

Overlapping and distinct expression patterns of *en* paralogs in the pupal wings of *B. anynana*

Studies on *en*-family genes in butterflies have been primarily carried out using the 4F11 monoclonal antibody that binds to a conserved domain of the translated En and Inv proteins of *D. melanogaster* (Patel et al., 1989). These En-family proteins are expressed in the posterior compartment and eyespot centers of larval and pupal wings (Banerjee and Monteiro, 2020b; Keys et al., 1999) and in the cells that will differentiate the orange ring of the eyespots in pupal wings of *B. anynana* (Brunetti et al., 2001), but until now it was not clear how many copies of *en-like* genes and transcripts were contributing to the observed multiple protein expression domains in this species. We aimed to find if one or more copies of *en*-like genes were exclusively associated with the novel eyespot color patterns and if these novel expression domains might have contributed to the preservation of *en* paralogs in the *B. anynana* genome.

*In-situ* hybridization against each *en* paralog in the present study indicates that they are expressed in different patterns in the pupal wings of *B. anynana*. The

protein expression patterns are likely the result of the expression of all three *en* paralogs (**Fig 4.3**). The expression in the posterior wing compartment and in the orange ring areas is specific to *en*, while all three paralogs are expressed in the eyespot centers.

Previous work established that both *en* and *inv* are expressed in the posterior compartment of butterflies during larval wing development (Banerjee and Monteiro, 2020b; Carroll et al., 1994), however, we show here that later in development only *en* is present in the posterior compartment of pupal wings. No data is currently available for the expression of *inv-like* during the larval stages. Previous studies using *in-situ* hybridization showed that *en* in *B. anynana* is homogeneously expressed in the posterior compartment, while *inv* is present in the upper posterior compartment in *B. anynana* (Banerjee and Monteiro, 2020b) and *Junonia coenia* (Carroll et al., 1994). In the pupal stage, we only observed the expression of *en* in the posterior compartment (Fig 3, d-f, Fig S2, a-c). The loss of *inv* in the posterior compartment during pupal wing development is most likely because this gene is no longer needed to perform its earlier larval function of setting up posterior veins (Banerjee and Monteiro, 2020b). Venation patterning happens during the early fifth instar (Banerjee and Monteiro, 2020b), where both *en* and *inv* are likely involved in activating other downstream diffusible signals such as Hedgehog and Bone Morphogenic Proteins that position the veins (Banerjee and Monteiro, 2020b). Late pupal expression of *en* in the wings of *D. melanogaster*, however, is involved in the maintenance of the posterior compartment, as mutations of *en* at these late stage leads to an enlarged posterior compartment with venation defects (Blair, 1992; Garcia-Bellido and Santamaria, 1972; Lawrence and Morata, 1976). The

maintenance of *en* expression in pupal wings of butterflies, suggests that this gene may be playing a similar role.

The expression of all three *B. anynana en*-family genes in the eyespot centers indicates that all three paralogs might have a redundant function in the differentiation or signaling from the eyespot center which sets the rings surrounding it, but this requires functional evidence (**Fig 4.3**) (Beldade and Brakefield, 2002; Monteiro, 2015). Many lineages of nymphalid butterflies lost expression of En-family protein in the eyespot center (as evidenced by lack of protein expression detected by the 4F11 antibody (Oliver et al., 2012)) and yet they develop eyespots. This suggests that En-family proteins might not be necessary for eyespot center differentiation and/or signaling (Oliver et al., 2012).

The expression of *en* in the orange ring area indicates that this gene alone might play a role in its differentiation, but again this requires functional validation. *B. anynana goldeneye* spontaneous mutants, where the orange ring extends inwards into the area of the central black scales in the eyespots, display an equally extended expression of En-family proteins into that area (Saenko et al., 2008) but it is still unclear whether *en*, or some other gene equally affected by the *goldeneye* mutation is actually responsible for the change in scale colors in the central disc.

Alternative 3' UTR splicing of *en* in *B. anynana*

The different isoforms of *en* with varying 3' UTR length could be involved in controlling the level of proteins produced and the localization of mRNA; and hence, they might also be playing a role in the differential expression of this paralog compared to *inv* and *inv-like*. Alternative 3' UTR lengths are achieved by the post-transcriptional process known as alternative cleavage and polyadenylation (APA) (Elkon et al., 2013). These 3' UTRs are known to control expression level and localization of proteins (Elkon et al., 2013; Mariella et al., 2019; Mayr and Bartel, 2009). Shorter 3' UTRs are usually translated at higher rates, leading to higher protein levels (Mayr and Bartel, 2009). Future studies could focus on understanding the spatial and temporal expression of the alternative 3' UTR of *en* to investigate if they have diverged from each other.

*en* paralogs and eyespot evolution

En-family proteins are differentially expressed in Nymphalid butterflies (Brunetti et al., 2001; Oliver et al., 2012). Eyespots are believed to have originated in a lineage sister to the sub-family Danainae, and now revised to have originated around 70 mya (Kawahara et al., 2019), via co-option which resulted in the expression of a set of genes in the eyespot centers, one of which includes En/Inv (Oliver et al., 2012). However, after the duplication, many lineages lost expression of En-family proteins in the eyespot centers (Oliver et al., 2012). *V. cardui,* for example, has expression in the orange rings but not in the eyespot centers (Brunetti et al., 2001), while the closely related *J. coenia* has expression of En-family proteins in the eyespot centers and the surrounding black scales, but not in the orange ring (Brunetti et al., 2001).

The *in-situ* hybridization data on the pupal wings of *B. anynana* (**Fig. 4.3**) reveal that two of the *en* paralogs (*inv* and *inv-like*) either lost two ancestral expression domains (posterior compartment and gold ring) or, alternatively, that one paralog (*en*) gained novel expression domains at these two locations after the duplication event. Given that *inv* is expressed in the posterior compartment of larval wings of *B. anynana*, but is absent in pupal wings, it might be more parsimonious to propose that *inv* might have had both larval and pupal wing expression but subsequently lost its pupal expression.

We propose two equally parsimonious scenarios for the evolution of expression domains of *en* paralogs (**Fig 4.4**). The two scenarios are not exclusive. In both scenarios, the ancestral copy of all the present paralogs in *B. anynana* already had expression in the posterior compartment of the wing because this is the pattern of expression for these genes in *D. melanogaster* in the pupal stage (Blair, 1992), which shared a common ancestor with Lepidoptera around 290 mya (Misof and et al., 2014) (Fig 4a). This would be facilitated if all copies shared a common set of regulatory elements, as is observed for the regulation of *en* and *inv* in *D. melanogaster* (Cheng et al., 2014; Gustavson et al., 1989). After eyespots originated, around 70 mya (Oliver et al., 2012) (this is a revised estimate based on a new phylogenetic dating study (Kawahara et al., 2019)), all paralogs gained a novel expression domain in the eyespot centers. This might have happened in a few different ways. A novel *cis*-regulatory element might have evolved in a genomic region that affects the expression of all *en*-copies simultaneously in the eyespot center (Scenario 1, Fig 4b), or a pre-existent *cis*-regulatory element that was already enhancing all copies simultaneously might have been re-used in a novel

location (eyespot centers) due to the co-option of a gene regulatory network to that novel location, containing *en/inv* as genes positioned within the co-opted network (Monteiro and Gupta, 2016; Monteiro and Podlaha, 2009) (Scenario 2, **Fig. 4.4c**). Subsequently, in both scenarios, the expression of *inv* and *inv-like* in the posterior compartment is lost (**Fig 4.4d and e**). Finally, either a new *cis*-regulatory element driving *en* expression in the orange ring is gained (Scenario 1, **Fig 4.4f**), or a pre-existent *cis*-regulatory element is reused for the novel *en* expression in the orange ring, again via gene network co-option (Scenario 2, **Fig 4.4g**). It is, however, also possible that the expression of *en* in association with the gold ring is gained in all three copies at once and subsequently lost in the *inv/inv-like* copies (not shown).

An important point to make here is that the timing of the first duplication event of *en* into *en* and *inv* paralogs happened around 480 mya or earlier (Misof and et al., 2014; Peel et al., 2006), and the duplication *inv* into *inv* and *inv-like* happened no more recently than around 130 mya (see discussion section above). Both of these duplications happened much earlier than the origin of eyespots, estimated to be around 70 mya (Kawahara et al., 2019; Oliver et al., 2012; Oliver et al., 2014). This suggests that the duplication and maintenance of *en*-like paralogs in the genome of *B. anynana* as well as in the genome of other Lepidoptera is not particularly well associated with the origin of eyespots.

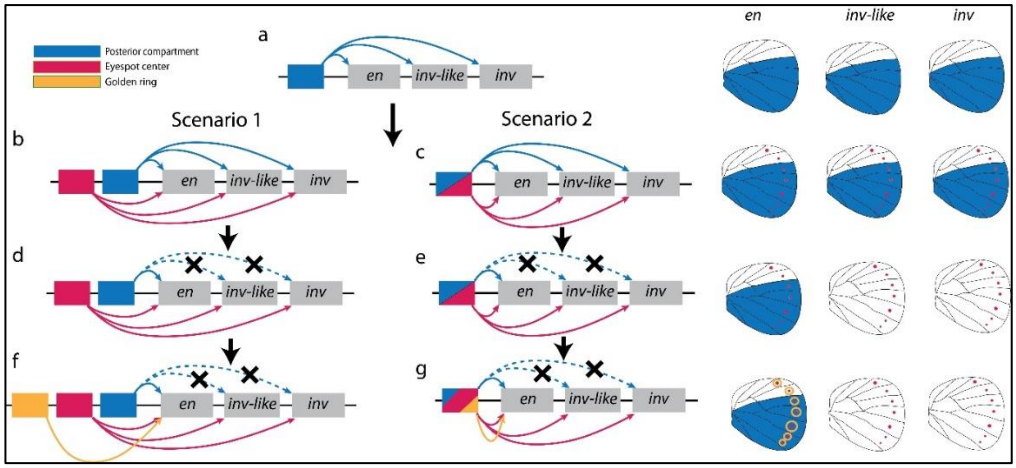



**Figure 4.4: Two alternative scenarios for the evolution of expression patterns of *en*-family genes.** (**a**) The ancestral state of all the three *en*-family genes likely had a common *cis*-regulatory element that derived expression of all three paralogs in the posterior compartment. (**b, d, and f**) Under Scenario 1, gain of novel *cis*-regulatory elements in the eyespot centers and in the orange rings resulted in the final patterns of gene expression we observe today. (**c, e, and g**) Under scenario 2, the same (or a restricted number) of ancestral *cis*-regulatory elements become redeployed to drive the novel patterns of gene expression in the eyespot centers and in the orange rings.

Future studies focused on the expression pattern of the *en* paralogs in model species such as *J. coenia* and *V. cardui*, as well as on the sequencing of open chromatin region surrounding the genes in *B. anynana* using techniques such as FAIRE and ATAC, will help us identify how the expression of *en*-like genes and their *cis*-regulatory elements diversified in the context of eyespot evolution. Functional studies will also be important to discover whether *en*-family members are actually involved in the regulation of eyespots in nymphalid butterflies.

## Conclusion

The present study provides the first evidence of the expression of multiple gene duplicates associated with the development of eyespots, a trait novel to nymphalid butterflies (Monteiro, 2015). By studying gene transcripts rather

than conserved domains of protein we found that *en* paralogs are expressed in both overlapping and distinct domains. We proposed two models for the evolution of the expression pattern in *en* paralogs, where either loss or gain of novel expression domains resulted in differential expression of the paralogs in the wings on *B. anynana*. Furthermore, we showed that the duplication of the three genes happened much earlier than the origin of eyespots. Hence, this study does not lend support to the idea that these particular gene duplicates were retained in Lepidopteran genomes due to their involvement in the development of this particular evolutionary novelty. Future studies focused on understanding the interaction of *en* paralogs with each other as well as other known genes that play roles in eyespot development will continue to illuminate the evo-devo of eyespot patterns.

**Authors contribution**

TDB wrote the manuscript with inputs from DR, and AM. TDB performed bioinformatic analysis, *in-situ* hybridization, and phylogenetic analysis. TDB and AM performed antibody staining's and DR found the splice variants of *en*. TDB and AM analysed and visualized the results. All authors read and agreed to the final version of the manuscript.

**Acknowledgement**

This work was supported by a Ministry of Education, Singapore grant MOE2015-T2-2-159 and the Department of Biological Sciences LHK fund GL 710221. TDB was supported by a Yale-NUS scholarship. We would like to thank Nipam Patel for the anti-Engrailed/Invected 4F11 antibody and Heidi Connahs for help with the transcriptome analysis. We would like to thank

DBS-CBIS confocal facility for access to confocal microscopes. The authors declare no conflict of interest.

**Supplementary materials**

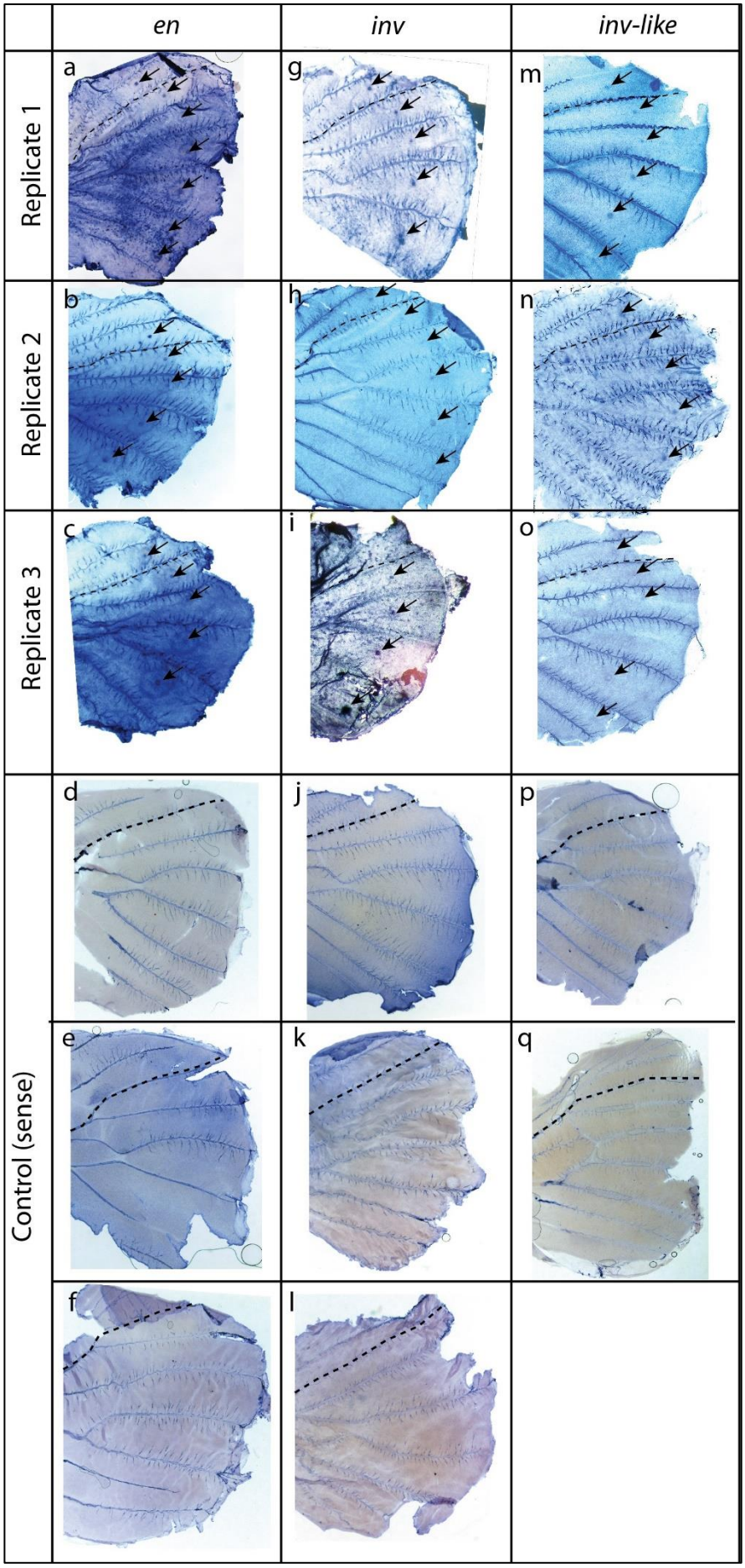

en
inv
inv-like
Replicate 1
Replicate 2
Replicate 3
Control (sense)
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q

**Figure S4.1. Replicates of *en*, *inv*, *inv-like* and control in-situ hybridization staining in *B. anynana* pupal wings.** (a-c) Pupal hindwings stained with probe against *en*. The transcript is present in the posterior compartment, in the eyespot centers and in the eyespot orange ring area. (g-i) Pupal hindwings stained with probe against *inv*. The transcript is present only in the eyespot centers (m-o) Pupal hindwings stained with probe against *inv-like*. The transcript is present only in the eyespot centers. (d-f) Pupal wings stained with *en* sense probe (control). (j-l) Pupal wings stained with *inv* sense probe. (p and q) Pupal wings stained with *inv-like* sense probe. Non-specific staining is observed in the veins and trachea. Anterior-posterior compartments are separated by dotted lines. Black arrows point to eyespot centers.

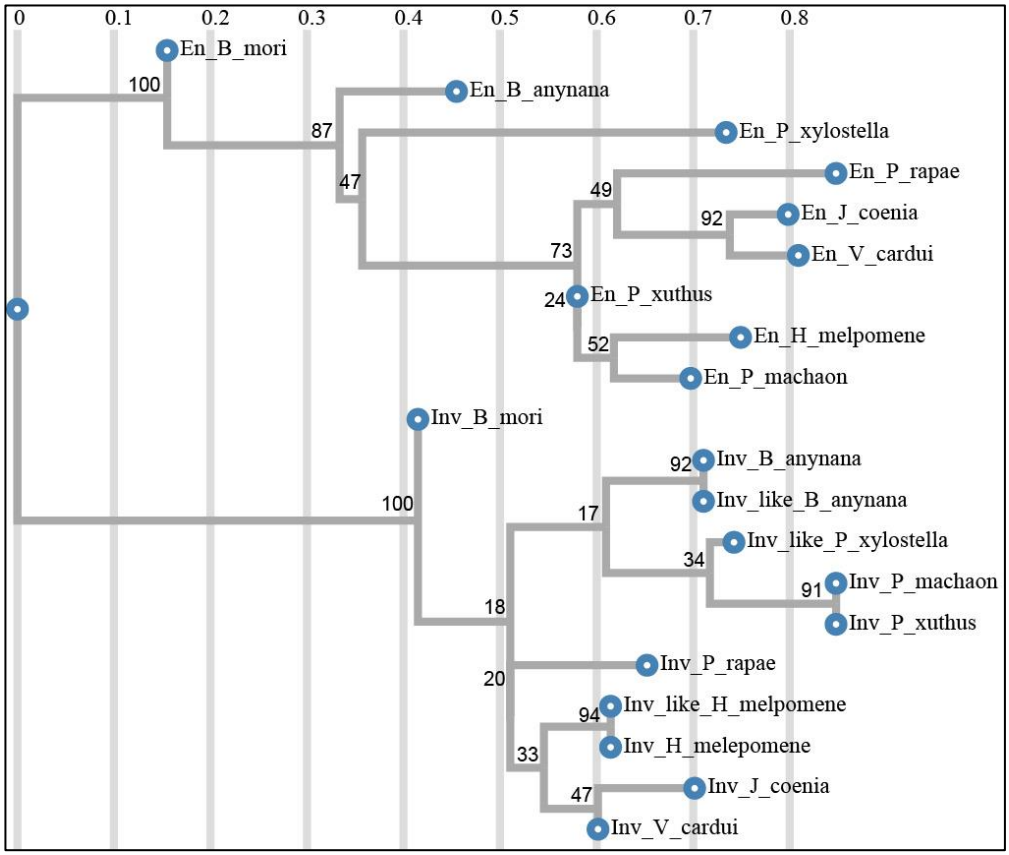



**Figure S4.2. Phylogenetic tree of En-family proteins in Lepidoptera created using PhyMl (Maximum likelihood. See methods in the main text for details)**. The tree clusters En together; and Inv and Inv-like together. Vertical lines indicate mean number of amino acid substitutions per site. Numbers above branches represent branch support calculated based on 100 bootstraps.

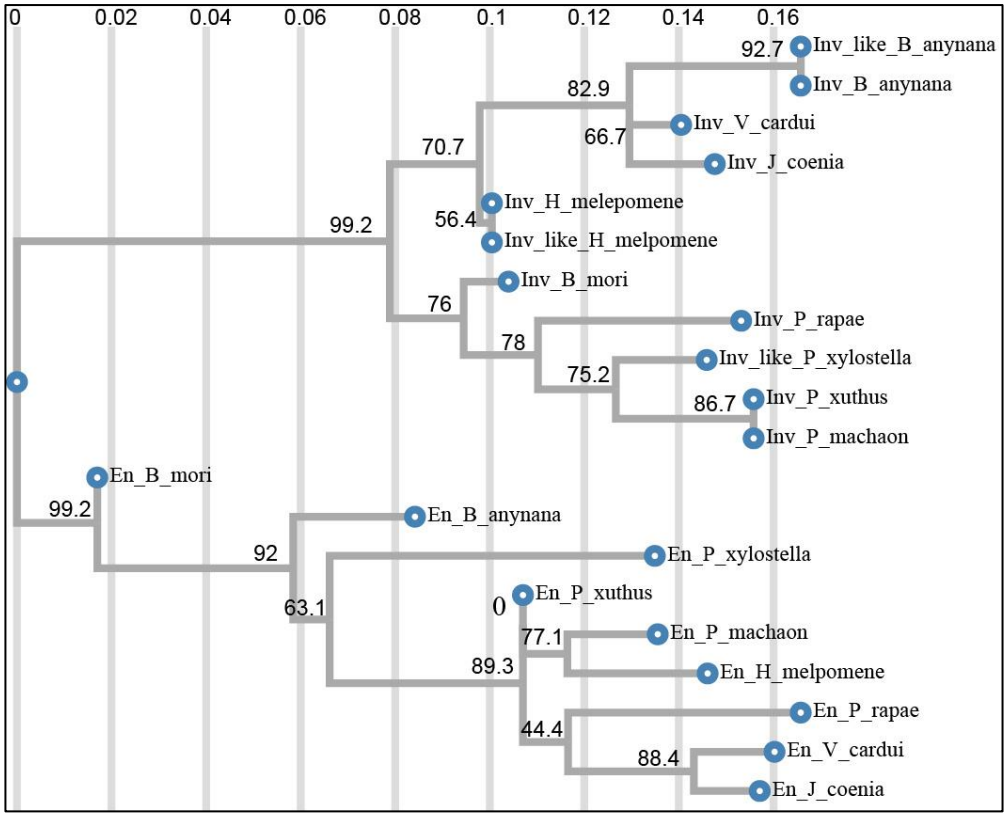



**Figure S4.3. Phylogenetic tree of En-family proteins in Lepidoptera created using fasttree (Maximum likelihood. See the methods section in the main text for details).** The tree clusters En together; and Inv and Inv-like together. Vertical lines indicate mean number of amino acid substitutions per site. Numbers above branches represents branch support.

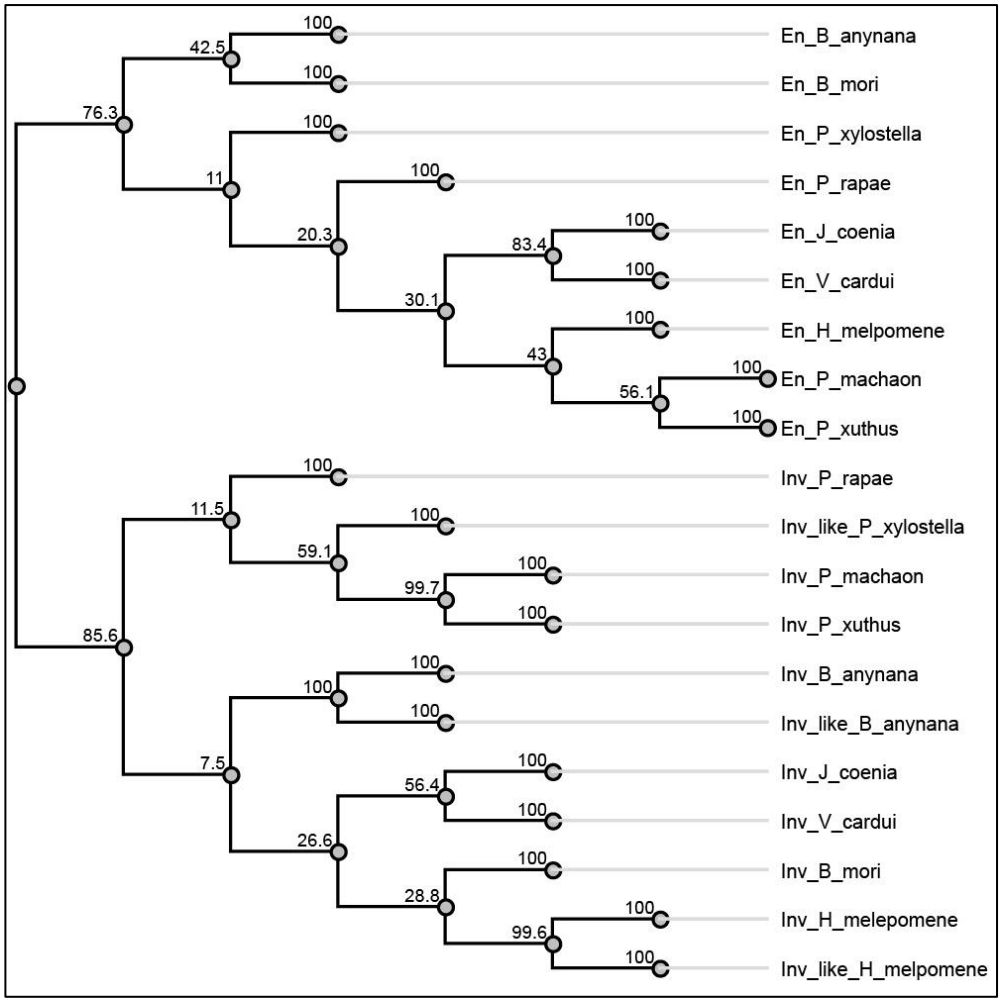



**Figure S4.4. Phylogenetic tree of En-family proteins in Lepidoptera created using Geneious tree builder (Jukes-Cantor as the genetic distance model, and UPGMA as tree build method).** The tree clusters En together; and Inv and Inv-like together. Numbers above branches represent branch support calculated based on 1000 bootstraps.

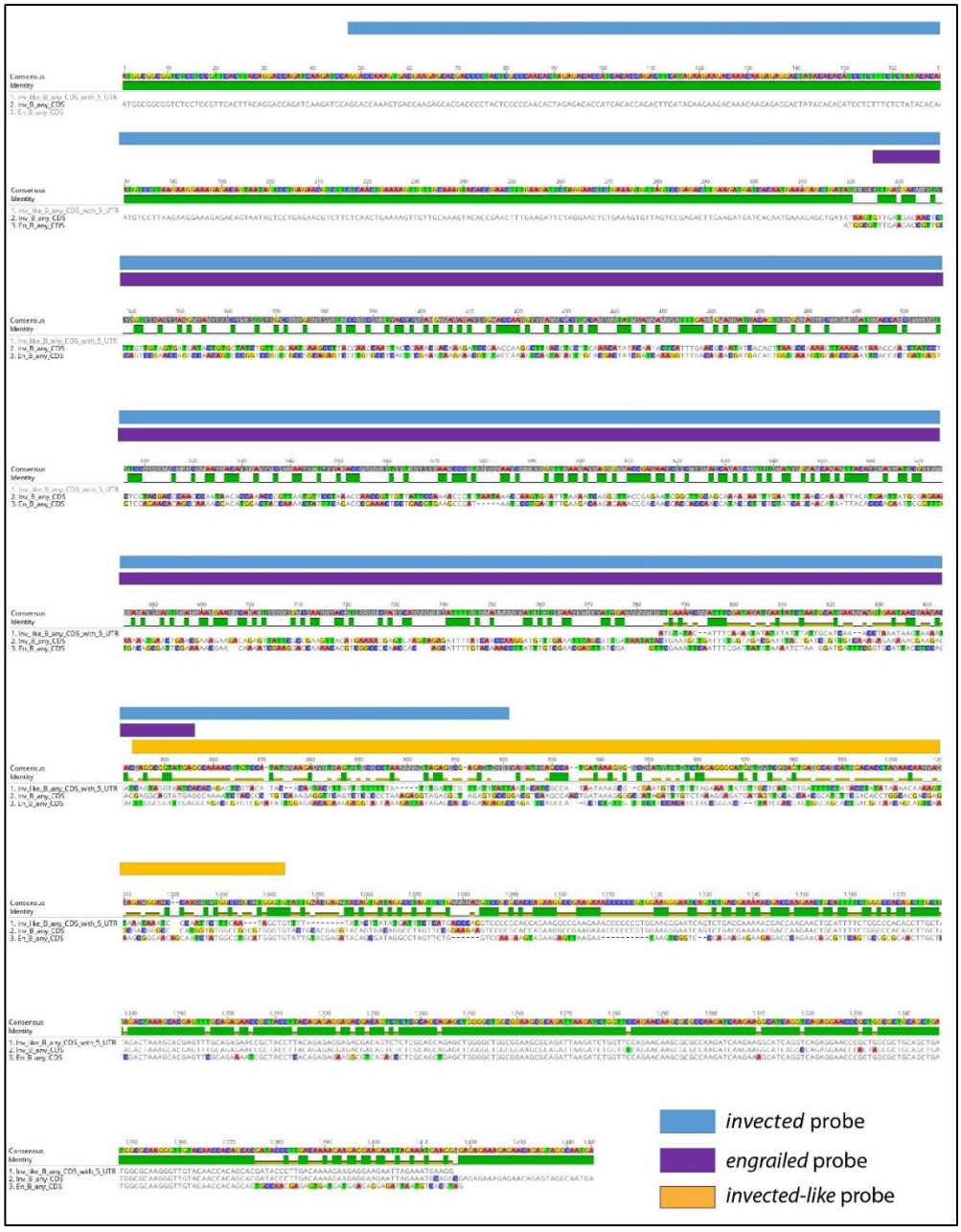



**Figure S4.5**. **DNA sequence alignment of *en*, *inv* and *inv-like* and the regions used for probe design**. Sequence used for *inv* probe has 24.9 % identical bps to that of the *en* sequence, and 36.03 % identical bps to that of the *inv-like* sequence. Sequence used for *inv-like* probe has 30.2 % identical bps to that of the *en* sequence, and 36.8 % identical bps to that of the *inv* sequence.

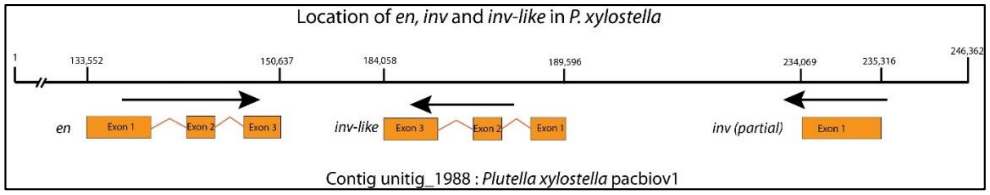



**Figure S4.6. Mapping of *en*, *inv* and *inv-like* in the genome of *P. xylostella* (pacbiov1).** Location of the transcripts was obtained from Lepbase.

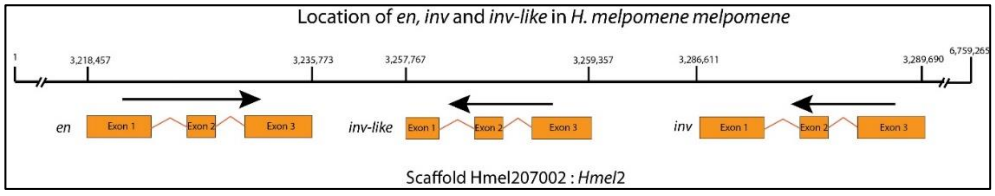



**Figure S4.7. Mapping of *en*, *inv* and *inv-like* in the genome of *H. melpomene melpomene* (Hmel2).** Location of the transcripts was obtained from Lepbase.

**Table S4.1. En scaffold features of *B. anynana, H. melpomene melpomene,* and *P. xylostella.***

| | En scaffold features | | | | | |
|---|---|---|---|---|---|---|
| | Distance between *en* and *inv* (bps) | Distance between *en* and *inv-like* (bps) | Distance between *inv* and *inv-like* (bps) | Orientation of *en* in the scaffold | Orientation of *inv* in the scaffold | Orientation of *inv-like* in the scaffold |
| *B. anynana* | 116,552 | 54,225 | 63,782 | Forward | Reverse | Reverse |
| *H. melpomene melpomene* | 71,233 | 40,900 | 31,923 | Forward | Reverse | Reverse |
| *P. xylostella* | 101,764 | 56,044 | 51,258 | Forward | Reverse | Reverse |

**Table S4.2. Similarity matrix of highly conserved regions (EH4 and EH5) of *en*, *inv* and *inv-like* nucleotide sequences in Fig S5.**

| | en_B_anynana_CDS | inv-like_B_anynana_CDS | inv_B_anynana_CDS |
|---|---|---|---|

| | | | |
|---|---|---|---|
| en_B_anynana_CDS | | 81.818% | 80.07% |
| inv-like_B_anynana_CDS | 81.818% | | 98.252% |
| inv_en3_B_anynana_CDS | 80.07% | 98.252% | |

**Table S4.3. Primer table**

| Sl. No. | Primer Name | Sequence | Description |
|---|---|---|---|
| 1. | en_insitu_F | TTGAAGACCGTTGCAGTCC | Forward primer to amplify *en* for *in-situ* hybridization |
| 2. | en_insitu_R | TAGATTGCTGTTCCCGCTTT | Reverse primer to amplify *en* for *in-situ* hybridization |
| 3. | inv_insitu_F | GGACCAAAGTGACGAAGAGC | Forward primer to amplify *inv* for *in-situ* hybridization |
| 4. | inv_insitu_R | TCCGGCACTCTAGCCTCTAC | Reverse primer to amplify *inv* for *in-situ* hybridization |
| 5. | inv-like_insitu_F | CAGTAGGTAATCACACAGACTCG | Forward primer to amplify *inv-like* for *in-situ* hybridization |
| 6. | inv-like_insitu_R | CACCTATTGAAAGAATTGG | Reverse primer to amplify *inv-like* for *in-situ* hybridization |
| 7. | ef1α_F | GTGGGCGTCAACAAAATGGA | Forward primer to amplify *ef1α* |
| 8. | ef1α_R | GCAAAAACAACGATGG | Reverse primer to amplify *enf1α* |

Sequence of *engrailed* used for *in-situ* hybridization (628 bps)

TTGAAGACCGTTGCAGTCCGAACCAGGCCAACAGCCCCGGTCCGGTCACC
GGCAGAGTCCCTGCGCCTCACTCCGAAGTAAGAAACGNGTACCAAAGTCA
ATACACTTGCACGACTATCGATCAAAGGTTTGACAGAACGATGACAGTGG
TGAAAGTGCAGCCGAATTCACCACCGATGAGTCCACTGACGTGAAGCCCA
TAATCCCTGAGTTTGAAGACAAGAGAAACCGACAACCACCACCAACCATA

CCCTTCTCTATCAGCAACATATTACACCCAGAATTCGGTTTGACAGCGATT
CGAAAAACGAACAAAATCGAAGGACCAAAACACGTCGGCCCCAACCACA
GCATTTTGTACAAACCTTATTTGTCGAACGAGTTATCGAGTTCGAAATTCA
ATTTCGATTATTTAAAATCTAAGGATGATTTCGGTGCATTACCTCCACTTG
GCGGTTTGAGGCAGACCGTGTCGAATATTGGAGAACAGAAGGAGGCACC
AAAGATTATAGAGCAGCAGAAGAGGCCAGATTCAGCCAGCTCTATTGTCT
CTTCCACATCTAGCGGGGCTTTATCGACGTGTGGCAGCACTGACGCCAAC
AGCAGTCAAAGCGGGAACAGCAATCTA

Sequence of *invected-like* used for *in-situ* hybridization (182 bps)

CAGTAGGTAATCACACAGACTCGTACATACACAATACTTTGTTTTTTTTTA
TTTGATTCGTTATTTATTAATACATCGCCATAATAAAGCGCGAATGTCTTT
TTAGAAAGTATGTTGCTCATAGTGATTTTCTATACCTATATAAAACAAAAG
TAAATAAATCCCAATTCTTTCAATAGGTG

Sequence of *invected* used for *in-situ* hybridization (873 bps)

GGACCAAAGTGACGAAGAGCACGACCCCTACTCGCCCAACACTAGAGAC
ACCATCACACCAGACTTCATAGAAGAAGACAAACAAGACAGGCCTATAC
ACACATCCTCTTTCTCCATACACAATGTCCTTAAGAAGGAAAGAGACAGT
AATAGTCCTGAGAACGTCTTCTCAACTGAAAAGTTGTTGCAAAGTACACC
GAACTTTGAAGATTCTAGGAACTCTGAAAGCGTTAGTCCGAGACTTGAAG
ATGATCACAATGAAAGAGCTGATATAAGTGTTGATGACAACTCTTGTTGT
AGTGATGATACTGTGCTATCTGTTGGCAATGAAGCCTTACCAACCAATTA
CCCAAACGACAAAGATCCGAACCAAGGCTTAACCTCCTTCAAACATATAC
AAACTCATTTGAACGCAATATCACAGTTAAGTCAAAATTTAAACATAAAC
CAACCAATCCTCCTACGACCCAACCCAATAACACCAAACCCGTTAATGTT
CCTAAACCAACCGTTGTTATTCCAAAACCCTTTAATAAACCAAGTGGATTT
AAAATCAGGGTTACCGAGAATCGGCTTGCAGCAAAACAATTTAAATTTGA
ACCAAAATTACATGAATTATGCGAGAAAAAATGAACTGAACGAAAGACG
ACAGAGTTATTCACCGAAGTTACATGAAAATGAGTCAAGTAGAGATTTTA
TTAACCAAGGATGTTTGAAATTTAGCATTGATAATATACTGAAAGCTGAT
TTTGGTAGACGAATTACTGATCCGTTGACAAAGAGAAAAACGAAGACGA
GGCAGTATGAGGCAAAATCTACCCCTGTCAAAGAGGTTCAGTCTCCCCCT
AAAGAGGTAGAGGCTAGAGTGCCGGA

Engrailed copies in *Bicyclus anynana* (highlighted colors represents different exons)

>En_B_anynana_CDS

ATGGCGTTTGAAGACCGTTGCAGTCCGAACCAGGCCAACAGTCCCGGTCCGGTCGCCGGCA
GAGTCCCTGCGCCTCACTCCGAAGTAAGAAACGTGTACCAAAGTCAATACACTTGCACGAC
TATCGATCAAAGGTTTGACAGAACGATGACAGTGGTGAAAGTGCAGCCGAATTCACCACCG
ATGAGTCCAGAACACAGCCAAAACGACATGGACTACCAAAACTATTTCAGACCCGAAACTC
CTGACGTGAAGCCCATAATCCCTGAGTTTGAAGACAAGAGAAACCGACAACCACCACCAAC
CATACCCTTCTCTATCAGCAACATATTACACCCAGAATTCGGTTTGACAGCGATTCGAAAAA
CGAACAAAATCGAAGGACCAAAACACGTCGGCCCCAACCACAGCATTTTGTACAAACCTTA
TTTGTCGAACGAGTTATCGAGTTCGAAATTCAATTTCGATTATTTAAAATCTAAGGATGATTT
CGGTGCATTACCTCCACTTGGCGGTTTGAGGCAGACCGTGTCGAATATTGGAGAACAGAAG
GAGGCACCAAAGATTATAGAGCAGCAGAAGAGGCCAGATTCAGCCAGCTCTATTGTCTCTT
CCACATCTAGCGGGGCTTTATCGACGTGTGGCAGCACTGACGCCAACAGCAGTCAAAGCGG
GAACAGCAATCTATGGCCTGCATGGGTGTATTGTACGAGATACAGCGATAGGCCTAGTTCT

GGTCCAAGAAGTAGAAGAGTTAAGAATAAGTCGGTCCCAGAAGAGAAGAGACCCAGAACA
GCGTTCAGTGCGGCGCAACTTGCTCGACTAAAGCACGAGTTCGCAGAAAATCGCTACCTCA
CAGAGAGAAGGCGTCAGACCCTCGCAGCTGAGCTGGGGCTGGCGGAAGCGCAGATTAAG
ATCTGGTTCCAGAACAAGCGCGCCAAGATCAAGAAAGCATCAGGTCAGAGGAACCCGCTG
GCGCTGCAGCTGATGGCGCAAGGGTTGTACAACCACAGCACTGCCAACGAGAGTGATGAT
GAAGAGGAGATTAATGTCACGTAG

>Inv_like_B_anynana_CDS_with_5_UTR

ATGTATACATTTCAAAATATATATTTATTGCATCAAACCTAAATAACTAAAATCAGTAGGT
AATCACACAGACTCGTACATACACAATACTTTGTTTTTTTTTATTTGATTCGTTATTTATTAAT
ACATCGCCATAATAAAGCGCGAATGTCTTTTTAGAAAGTATGTTGCTCATAGTGATTTTCTAT
ACCTATATAAAACAAAAGTAAATAAATCCCAATTCTTTCAATAGGTGTTTTTATACTTATATG
ATTTCTCATGACCCAGGTCCGCGCACCAGAAGGCCGAAGAAACCCCCCGTGGAAGGGAATC
AGTCTGACGAAAAACGACCAAGAACTGCATTTTCTGGGCCACAGCTTGCTAGACTAAAGCA
CGAGTTTGCAGAGAACCGCTACCTTACAGAGAGGAGACGACAGTCTCTCGCAGCAGAGCT
GGGGCTGGCGGAAGCGCAGATTAAGATCTGGTTCCAGAACAAGCGCGCCAAGATCAAGAA
GGCATCAGGTCAGAGGAACCCGCTGGCGCTGCAGCTGATGGCGCAAGGGTTGTACAACCA
CAGCACGATACCCTTGACAAAAGAAGAGGAAGAATTAGAAATGAAGG

>Inv_B_anynana_CDS

ATGGCGGCGGTCTCCTCCGTTCACTTACAGGACCAGATCAAGATCCAGGACCAAAGTGACG
AAGAGCACGACCCCTACTCGCCCAACACTAGAGACACCATCACACCAGACTTCATAGAAGA
AGACAAACAAGAGAGGACTATACACACATCCTCTTTCTCTATACACAATGTCCTTAAGAAGG
AAAGAGACAGTAATAGTCCTGAGAACGTCTTCTCAACTGAAAAGTTGTTGCAAAGTACACC
GAACTTTGAAGATTCTAGGAACTCTGAAAGTGTTAGTCCGAGACTTGAAGATGATCACAAT
GAAAGAGCTGATATAAGTGTTGATGACAACTCTTGTTGTAGTGATGATACTGTGCTATCTGT
TGGCAATGAAGCCTTACCAACCAATTACCCAAACGACAAAGATCCGAACCAAGGCTTAACC
TCCTTCAAACATATACAAACTCATTTGAACGCAATATCACAGTTAAGCCAAAACTTAAACATA
AACCAACCTATCCTCCTACGACCCAACCCAATAACACCAAACCCGTTAATGTTCCTAAACCAA
CCGTTGTTATTCCAAAACCCTTTAATAAACCAAGTGGATTTAAAATCAGGGTTACCGAGAAT
CGGCTTGCAGCAAAACAATTTGAATTTGAACCAAAATTACATGAATTATGCGAGAAAAAAT
GAACTGAACGAAAGAAGACAGAGTTATTCGCCGAAGTTACATGAAAACGAGTCAAGTAGA
GATTTTATCAACCAAGGATGTTTGAAATTTAGCATTGATAATATACTGAAAGCTGATTTTGG
TAGACGAATTACTGATCCGTTGACAAAGAGAAAAACGAAGACGAGGCAGTATGAGGCAAA
ATCTACCCCTGTCAAAGAGGTTCAGTCTCCCCCTAAAGAGGTAGAGGCTAGAGTGCCGGAC
GTCAAGCCAACTGATAAAGGGGCCATAGATTTGTCTAAAGGAGATGATAGTTGCAGCAAC
GCATCTTCGACACCTGGCACGACGAGCGACGGGCCCATGGTGTGGCCGGCGTGGGTGTAC
TGCACGAGGTACAGTGACAGGCCTAGTTCCGGAAGAAGTCCGCGCACCAGAAGGCCGAAG
AAACCCCCCGTGGAAGGGAATCAGTCTGACGAAAAACGACCAAGAACTGCATTTTCTGGGC
CACAGCTTGCTAGACTAAAGCACGAGTTTGCAGAGAACCGCTACCTTACAGAGAGGAGAC
GACAGTCTCTCGCAGCAGAGCTGGGGCTGGCGGAAGCGCAGATTAAGATCTGGTTTCAGA
ACAAGCGCGCCAAGATCAAGAAGGCATCAGGCCAGAGGAACCCACTAGCGCTGCAGCTGA
TGGCGCAAGGGTTGTACAACCACAGCACGATACCCTTGACAAAAGAAGAGGAAGAATTAG
AAATGCAGGCGAGAGAAAGAGAACAGAGTAGGCAATGA

>En_H_melpomene_lepbase_HMEL009342g1.t1

ATGGCGTTCGAAGACCGCTGCAGTCCGAATCAAGCGACGAGTCCTGGCCCAGTTACTGGTA
GAGTCCCAGCGCCGCACGCGAGCTCAGCAAATATGACTTTGTACACGCAAAACCAGTACAC
TTGCACCACTATTGATTCACGGTTTGACAGATCGAACATGACGGTGATTAAAGTTCAGCCTA
ACTCGCCGCCACCGAGCCCTGATCACAGCCAAGAGATGGACTACCAAAACTACTACCGCCC
AGAAACACCTGATGTGAAACCGGTTATCCCTGATGACAGATTCGATTTGGATAAAAGAAGT
CGACAGCAAGCGGCTCCTTCGATTGCGTTTTCGATAACTAACATCCTCCACCCTGAATTTGG
GTTGAATGCTATAAAGAAAACAAGCAAAATTGAAGGTCCCAAATCAGTTGGGCCAAATCAC
AGTATTTTGTACAAGCCATACGATTTGTCGAAGGAATATTCAAAGTTTAACTTCGATTATATA
AAATCCAAAGATGACTTTAGTGCTTTACCGCCGCTGGGTGGTTTACGGCAGACTGTGTCAA
ATATTGGGGAAGTTCAGAAGGAGAAGGAGATACCAAAGATCCAGGAAGTACAGAAACGA
CCAGATTCAGCCAGTTCTATAGTTTCTTCGGCGTCCAGCGGGCTGTCCACTTGTGGGAGTAC
GGATACGAATAGCCAGACTTCTCAAGGCAATAGTTCTTTATGGCCGGCGTGGGTCTACTGC
ACCAGATATAGCGATCGTCCAAGTTCTGGTCCAAGAAGCAGAAGGCTAAAGAAGAAAATG
AATCCAGAAGAGAAGAGGCCAAGAACGGCGTTCAGCGCAGCACAACTTGGTAGACTAAAG
CACGAGTTTGCGGAAAATCGCTATCTGACAGAACGTAGAAGACAGGCCCTGGCTACAGAG
TTAGGTCTGGCTGAAGCCCAGATCAAAATCTGGTTCCAGAATAAGCGGGCTAAAATCAAGA
AAGCTACTGGTCAAAGGAACCCATTGGCCTTACAGCTTATGGCCCAAGGACTCTATAACCAT
CGCACTGCGACAGAGAGTGATGAAGAAGAAGAGATTAGTGTTACGTAA

>Inv_H_melpomene _from _Lepbase_and _transcriptome _
HMEL009343g2.t1_Hme-CYT021116

ATGGCGGCGGTCTCCTCCGTTCATTTGCAAGATCAGATCAAAATTCAGGACAACAGCGATG
ATGAAACGGAACCCTATTCACCTAACACCAGAGACACAATCAGTCCATACCACGAACCTGA
AAAGCTCGAAGAGAGACCAATACATACATCGTCTTTTTCAATTCACAATGTTTTAAAGAAGG
AAAGAGAGAGTAATAGCCCTGAAAATGTTTTTTCTACAGATAAGCTATTGCAGAACACTCCA
AATTTCGAGGAGGGTTCCAGAAATTCGGAAAGTGTCAGTCCTAGATTGGACGACGATCACA
CAGACAGGGCTGATATAAGTGTAGATGATAACTCTTGCTGTAGTGACGATACAGTGTTATC
AGTAGGAAATGAGGCCCCAGCTTATAACCAGCACGAGAAAAATGATGCCCAGCTCACCTCC
TTTAAGCACATTCAAACGCACCTTAACGCGATATCGCAGCTTAGTCAAAATTTAAACATAAA
CCAACCAATTCTCTTAAGGCCCAATCCGATAACACCCAATCCATTAATGTTCCTCAACCAGCC
ATTGCTATTCCAAAATCCTTTGATGAGTCAAGTGGATCTCAAATCGGGTATGCCAAGAATGT
CGATACCACAAAACAGTTTGAACATAAATCAACAATTTAATATGAACTTTGTGAGAAAAAAT
CAGGAACCGTTGCACAGGATCGATGAAAATCGACGGCAAAATCAAAATTATTCACCAAAAT
CCCCAGAAAATGAATCCGAAAGGGATTTTATTAATCAGAGTTGTTTGAAATTTAGCATAGAT
AATATATTGAAAGCAGACTTCGGTAGGCGTATCACAGATCCGTTGACGAAACGGAAAACGT
TGAAGTCTAGGCAGTGTGATGTGAAGCCTACGCCTGTTAAGGAAGTGGCGTCGCCACCTAA
GGATATAGAAGCACGGGTACCGGAAGTGAAGCCTACGGATAAAGGTGCTATTGACCTTTCT
AAGGGCGACGATAGTGGTAGTAATGCATCTTCTACCCCGGGAACAACCAGCGACGGTCCA
ATGGTGTGGCCAGCGTGGGTGTACTGTACGAGGTACAGCGATCGGCCTAGTTCCGGACGA
AGTCCTCGCACCAGAAGACCTAAAAAGCCACCAGGAGATGGTGGTTCGGCAGACGAGAAA
AGACCTAGAACTGCATTTTCTGGCCCTCAACTAGCTAGATTAAAGCACGAATTCGCAGAAAA
TCGATATCTAACAGAAAGAAGAAGACAGTCTCTGGCTGCGGAACTAGGTCTGGCTGAAGCC
CAAATAAAAATATGGTTTCAGAATAAACGGGCAAAGATCAAAAAGGCATCTGGACAAAGA
AACCCATTGGCCTTACAACTTATGGCCCAAGGCCTTTACAATCACAGTACAATACCCTTGAC
AAAGGAAGAAGAAGAACTAGAAATGAAAGCAAGAGAAAGAGAACAG AGTCGACAGTAA

>Inv_like_H_melpomene_Lepbase_HMEL009343g1.t1

ATGGCGCAGTTGGACAGCCTCTCAATGTTAATGTGTTGGTTTAAGGAAAGTCTTTTTTTAAA
CTTTGTTTCGATCTTAGGTCCTCGCACCAGAAGACCTAAAAAGCCACCAGGAGATGGTGGTT
CGGCAGACGAGAAAAGACCTAGAACTGCATTTTCTGGCCCTCAACTAGCTAGATTAAAGCA
CGAATTCGCAGAAAATCGATATCTAACAGAAAGAAGAAGACAATCTCTGGCGGCGGAACT
AGGTCTGGCTGAAGCCCAAATAAAAATATGGTTTCAGAATAAACGGGCAAAGATCAAAAA
GGCATCTGGACAAAGAAACCCATTGGCCTTACAACTTATGGCCCAAGGCCTTTACAATCACA
GTACAATACCCTTGACAAAGGAAGAAGAAGAACTAGAAATGAAAGCAAGAGAAAGAGAA
CAGAGTCGACAGTAA

>En_P_xylostella_CDS_pacbiov1_g11934.t1

ATGGCCTTTGAGGACCGTTGCAGCCCGAGCCAGGCTAACAGCCCGGGGCCTGTCAGCGGG
AGAGTGCCGGCGCCTCACGCGGAGAACCTCATGAGCTTCTGCCAGCCGAGCCAGTACACCT
GCACCACCATCGAGCCGAGGTATGAGAGGAACCAGCCGTCGCCAATGACCATTGTGAAAG
TCCAGCCAGCTTCCCCGCCCTCCAGTCCCATGGAGACTCACGGCGACATGGATAACTACCAG
AACTTCTACCCGGAACGGCCAGAAACACCAGACGTGAAGCCCGTGATAAACCCAGTTCAAG
AAGACGCAAGATTCAACGATCGGTTCAGAAACCAGTGCCAGCAAGTGACTGTACCGATCAC
GTTCTCAATTAGCAACATTCTACACCCAGAATTCGGTTCTGGCGCATTGCGGAAAACCAACA
AGATTGAAGGACCGAAACCCGTCGGGCCTAACCACAGCATTCTTTACAAGCCATACGACCT
ATCCAAACCGGGGAGCGTTGCTGAGCACCAAAAAGGTTACAGCTTCGATTATTTGAAGACA
AAAGAGTCGAGTCACGACTTCAACCGACTACCACCGCTTGGGGGGTTGAGGCAGACCGTAT
CTCAGATTGGGGAGGTTCAGAAAGAGAAGGAGGTTGTTAGACCCATTGAGCCTCAAAGAA
GGCCGGATTCTGCGAGTTCCATTGTTTCTTCGACGTCGAGTGGGGCTCCCTCCACGTGTGGC
AGCACCGACGCCGGCAGCCAGGGGCAAGGGAACGCGAACTTGTGGCCAGCCTGGGTGTA
CTGCACCCGATACAGCGACCGACCCAGTTCTGGTCCTAGAAGTAGAAGGTCAAAGAAGAA
GTCCCCAAGTGCCGCAGAAGAGAAGAGACCGAGAACAGCGTTCAGCGCGGCGCAACTCAA
CAGGTTGAAGCACGAGTTCGCGGAGAACCGCTACCTGACGGAGCGGCGGCGCCAGGCGCT
GGCCGCGGAGCTGGGCCTGGCCGAGGCGCAGATCAAGATCTGGTTCCAGAACAAGCGAGC
CAAGATCAAGAAGGCCAACGGACAGAGGAACCCACTAGCCTTACAACTCATGGCTCAAGG
GCTCTACAACCACACGACAGCCACTGAAAGCGACGAAGAGGAAGAAATTAGTGTGACGTA
A

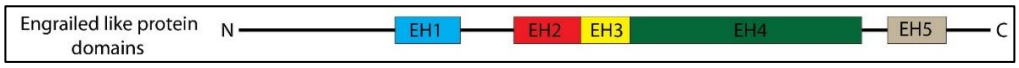
>Inv_like_P_xylostella_CDS_pacbiov1_g11936.t1

ATGAATGGTCCGCGCACCCGGCGACCCAAGAAGCCCCCCGGAGAGGTGCAGAACGACGAG
AAGCGACCCCGCACAGCCTTCTCCGGGCCACAGCTGGCCAGGCTAAAGCACGAGTTCGCG
GAGAACCGCTACCTGACGGAGCGGCGGCGCCAGGCGCTGGCCGCGGAGCTGGGCCTGGC
CGAGGCGCAGATCAAGATCTGGTTCCAGAACAAGCGAGCCAAGATCAAGAAGGCATCAGG
ACAGAGGAACCCACTCGCCTTACAGCTGATGGCACAAGGACTGTACAACCACAGTACGGTA
CCGTTGACGAGAGAAGAGGAGGAGTTGGAAATGAAGGCGAGAGAGAGGGATGCAGCTA
GGGGCGCGCCTCCGTAG

>Inv_P_xylostella_CDS_partial_pacbiov1_g11937.t1

ATGGCGGCGGTCTCCACCACGGTGCACCACGTGCAGCCAATCAAGACCTTAGAGCAGAGC
GACGAAGAGCCATATTCCCCCAACACGCGAGACACCACCACACCCGACTACCAAGAAGACG
AAAAGACCGAGGAAAGACCTCTCACCTCTTCTTTCTCCATACACAACGTCCTTAAGAAAGAA
AGGGACAGCAATAGCCCTGACAATGTCTTTTCAACGGACAAGCTTCTCCAAAGCACTCCGA
ATTTCGACCAATCATCGCCCAGAAACTCCCAAGACAGTCTGAACAGTCCACGGTTTGACAAC
GAAGATCACAATTCAAATTCAGAATCTGTCAGAGCGGAAATCAGTGTAGACGATGACGTGT

CATGCTGTAGCGATGATACTGTCTTGTCTGTTGGCAACGAAGCGCAACTTTCTTATAATTCG
CCGTCGGAAAGCGCGAAATCCTTGCCAGCCAGCCCTGGAACTCCTCCTAACAACCTCACCTC
GTTCAAACATATACAAACTCATTTGAACGCAATCTCACAGCTAAGTCAGAACATTAACATGA
ACCAGCCATTATTATTACGCCCCAACCCGATCTCACCTAACCCGCTTATGTTCTTAAACCAAC
CTCATTTGTTATTTCAAAACCCTCTCCTCGGCCAGGATGTGAAAGGAGGTCTGTTACCGAGC
AGAATCCCGATCCCTAATCCGAATTTAAACTTAAGTCCACCTCCTTCGAACCCATTCGGGCA
GTTGAACTTTTTGCGCAACCCGAAGACCCAAGAGTTGCGCAACAACCGAAACTTGGACGAA
AACCGCCGCTACCAGCCGAAATCCCCAGAAAACGAGTCCCACTTAAACATCCTCAACCAAA
GCTGTCTAAAATTCAGCATAGACAACATCCTAAAAGCGGATTTCGGGCGTCGAATAACGGA
TCCTTTGACGAAGCGGAAGAATTTCAAAAATCGAACGGAAGCGCCGAAGAGCTGTCCAAA
GGAGGTCGCGAAAGAGGTTCCTTGTTCCGAGGCGCGGGTGCCGGATGTGAAAGGGGGTG
CTATAGACTTGTCGAAGGTGGATGAAGGGGCCAGTAACCAAGGCAGTACCACCAGCAGTA
CCAGTGGCGATGGCCCCATGGTGTGGCCAGCCTGGGTCTACTGCACCCGCTACAGCGACCG
ACCTAGTTCTGGTAAGTTATTTTCTTTCCAGCCAGTGTTCTATGATCAAGGATTTACAAAAAA
GAGGGCAGCGGCCTTAGAATAA




EH1 (Groucho-binding domain), EH2 (Extra denticle binding domain), EH3, EH4 (Homeodomain), EH5 (Repression domain)

En Specific domain  Inv specific domain

>En_B_anynana

MAFEDRCSPNQANSPGPVAGRVPAPHSEVRNVYQSQYTCTTIDQRFDRTMTV
VKVQPNSPPMSPEHSQNDMDYQNYFRPETPDVKPIIPEFEDKRNRQPPPTIPFSI
SNILHPEFGLTAIRKTNKIEGPKHVGPNHSILYKPYLSNELSSSKFNFDYLKSK
DDFGALPPLGGLRQTVSNIGEQKEAPKIIEQQKRPDSASSIVSSTSSGALSTCGS
TDANSSQSGNSNLWPAWVYCTRYSDRPSSGPRSRRVKNKSVPEEKRPRTAFS
AAQLARLALKHEFAENRYLTERRRQTLAAELGLAEAQIKIWFQNKRAKIKKA
SGQRNPLALQLMAQGLYNHSTANESDDEEEINVT

>Inv_like_B_anynana

MISHDPGPRTRRPKKPPVEGNQSDEKRPRTAFSGPQLARLKHEFAENRYLTER
RRQSLAAELGLAEAQIKIWFQNKRAKIKKASGQRNPLALQLMAQGLYNHSTI
PLTKEEEELEMK

> Inv_B_anynana

MAAVSSVHLQDQIKIQDQSDEEHDPYSPNTRDTITPDFIEEDKQERTIHTSSFSI
HNVLKKERDSNSPENVFSTEKLLQSTPNFEDSRNSESVSPRLEDDHNERADIS
VDDNSCCSDDTVLSVGNEALPTNYPNDKDPNQGLTSFKHIQTHLNAISQLSQ
NLNINQPILLRPNPITPNPLMFLNQPLLFQNPLINQVDLKSGLPRIGLQQNNLNL
NQNYMNYARKNELNERRQSYSPKLHENESSRDFINQGCLKFSIDNILKADFG
RRITDPLTKRKTKTRQYEAKSTPVKEVQSPPKEVEARVPDVKPTDKGAIDLSK
GDDSCSNASSTPGTTSDGPMVWPAWVYCTRYSDRPSSGRSPRTRRPKKPPVE
GNQSDEKRPRTAFSGPQLARLKHEFAENRYLTERRRQSLAAELGLAEAQIKI
WFQNKRAKIKKASGQRNPLALQLMAQGLYNHSTIPLTKEEEELEMQARERE
QSRQ

>En_B_mori

MAFEDRCSPSQANSPGPVTGRVPAPHAETLAYSPQSQYTCTTIESKYERGSPN
MTIVKVQPDSPPPSPGRGQNEMEYQDYYRPETPDVKPHFSREEQRFELDRSR
GQRLQPTTPVAFSINNILHPEFGLNAIRKTNKIEGPKPIGPNHSILYKPYDLSKP
DLSKYGFDYLKSKETSDCNALPPLGGLRETVSQIGERLSRDREPPKSLEQQKR
PDSASSIVSSTSSGAVSTCGSSDASSIQSQSNPGQLWPAWVYCTRYSDRPSSGP
RSRRVKKKAAPEEKRPRTAFSGAQLARLKHEFAENRYLTERRRQSLAAELGL
AEAQIKIWFQNKRAKIKKASGQRNPLALQLMAQGLYNHSTVTESDDEEEINV
T

>Inv_B_mori

MAAVSAHMQDIKIQDQSDDDPYSPNTRDTTSPECHDDEKSEDISIRSSSFSIHN
VLKKERDNSSPDNVFSTEKLLENTPNFDDRNTESVSPVVEVNEREISVDDGNS
CCSDDTVLSVGNEAPVSNYEEKASQNTHQELTSFKHIQTHLSAISQLSQNMN
VAQPLLLRPSPINPNPIMFLNQPLLFQSPILSQDLKGMPNRQTANVISPTFGLNF
GMRLKANHETRTRSDENRYSKPEESRDYINQNCLKFSIDNILKADFGRRITDP
LHKRKVKTRYEAKPAPAKDTAAFAPKLDEARVPDIKTPDKAGAIDLSKDDSG
SNSGSTSGATSGDSPMVWPAWVYCTRYSDRPSSGRSPRTRRPKKPPGDTASN
DEKRPRTAFSGPQLARLKHEFAENRYLTERRRQSLAAELGLAEAQIKIWFQN
KRAKIKKASGQRNPLALQLMAQGLYNHSTVPLTKEEEELEMKARERERELK
NRC

>En_H_melpomene

MAFEDRCSPNQATSPGPVTGRVPAPHASSANMTLYTQNQYTCTTIDSRFDRS
NMTVIKVQPNSPPPSPDHSQEMDYQNYYRPETPDVKPVIPDDRFDLDKRSRQ
QAAPSIAFSITNILHPEFGLNAIKKTSKIEGPKSVGPNHSILYKPYDLSKEYSKF
NFDYIKSKDDFSALPPLGGLRQTVSNIGEVQKEKEIPKIQEVQKRPDSASSIVSS
ASSGLSTCGSTDTNSQTSQGNSSLWPAWVYCTRYSDRPSSGPRSRRLKKKMN
PEEKRPRTAFSAAQLGRLKHEFAENRYLTERRRQALATELGLAEAQIKIWFQ
NKRAKIKKATGQRNPLALQLMAQGLYNHRTATESDEEEEISVT

>Inv_like_H_melpomene

MAQLDSLSMLMCWFKESLFLNFVSILGPRTRRPKKPPGDGGSADEKRPRTAF
SGPQLARLKHEFAENRYLTERRRQSLAAELGLAEAQIKIWFQNKRAKIKKAS
GQRNPLALQLMAQGLYNHSTIPLTKEEEELEMKAREREQSRQ

>Inv_H_melepomene

MAAVSSVHLQDQIKIQDNSDDETEPYSPNTRDTISPYHEPEKLEERPIHTSSFSI
HNVLKKERESNSPENVFSTDKLLQNTPNFEEGSRNSESVSPRLDDDHTDRADI
SVDDNSCCSDDTVLSVGNEAPAYNQHEKNDAQLTSFKHIQTHLNAISQLSQN
LNINQPILLRPNPITPNPLMFLNQPLLFQNPLMSQVDLKSGMPRMSIPQNSLNI
NQQFNMNFVRKNQEPLHRIDENRRQNQNYSPKSPENESERDFINQSCLKFSID
NILKADFGRRITDPLTKRKTLKSRQCDVKPTPVKEVASPPKDIEARVPEVKPT
DKGAIDLSKGDDSGSNASSTPGTTSDGPMVWPAWVYCTRYSDRPSSGRSPRT
RRPKKPPGDGGSADEKRPRTAFSGPQLARLKHEFAENRYLTERRRQSLAAEL
GLAEAQIKIWFQNKRAKIKKASGQRNPLALQLMAQGLYNHSTIPLTKEEEELE
MKAREREQSRQ

> En_P_xuthus

MAFEDRCSPNQATSPGPVSGRVPAPHAESPVGYRPPSQYTCTTIDGRFDRGTP
NITMIKVQPSSPPPSPDRHNEMEYQNYYRPETPDVKPIINPDDRFEHDKRRQQP
TAPIAFSISNILHPEFGLSAIRKTNKIEGPKPAGPNHSMLYKPYDLSKQNEFQK
YSFDYLKSKEPSDFNSLPPLGGLRQTVSQIGELVSKEKEVQKVVEQRRPDSAS
SLVSSASSGAPSTCSTDATSQASGNSALWPAWVYCTRYSDRPSSGPRSRRMK
KKMNPEEKRPRTAFSAAQLARLKHEFAENRYLTERRRQALAAELGLAEAQIK
IWFQNKRAKIKKATGQRNPLALQLMAQGLYNHSTATESDEDEEEISVT

> Inv_P_xuthus

MAAVSTTVHLQDTRIKIQDQSDEDHEPYSPNTRDTISPEYNEEDKFEDRTIHTS
SFSIHNVLKKERDTNSPENVFSTDKLLQNTPNFDDGQNSRTSESVSPRLEDDD
TIGNEYRGEISVDDENSCCSDDTVLSVGNEAPVSSYHDNEKSSPDTSQGLSSF
KHIQTHLNAISQLSQNLNVNQPLLLRPSPITPNPLMFLNQPLLFQNPLNQVDLK
SGVIPNRMPMPQNNLNLNTQHFGLNFGLRMKNQTMETRHRSDENRRVNYST
PKSPDNESGRDFINQNCLKFSIDNILKADFGRRITDPLTKRKSLKSRSFEAKVP
VKEAPLPSKPEVLEARVPEIKPIDKGAIDLSKSDDGSSNQSSTTSTTTGENGPM
VWPAWVYCTRYSDRPSSGRSPRTRRPKKPPGEGPTSDEKRPRTAFSGPQLAR
LKHEFAENRYLTERRRQALAAELGLAEAQIKIWFQNKRAKIKKASGQRNPLA
LQLMAQGLYNHSTVPLTKEEEELEMKARAREKELQNRH

>En_P_machaon

MAFEDRCSPNQATSPGPVSGRVPAPHAESPVGYRPPSQYTCTTIDARYDRGM
PNITMIKVQPSSPPPSPDRHHEMEYQNYYRPETPDVKPVINPDDRFEPDKRRQ
QPTAPIAFSISNILHPEFGLSAIRKTNKIEGPKPAGPNHSMLYKPYDLSKQNEY
QKYSFDYLKSKEPSDFNSLPPLGGLRQTVSQIGELVSKDKEVQKVAEQRRPDS
ASSLVSSASSGAPSTCSTDAPSQASGNSALWPAWVYCTRYSDRPSSGPRSRR
MKKNMNPEEKRPRTAFSAAQLARLKHEFAENRYLTERRRQALATELGLAEA
QIKIWFQNKRAKIKKASGQRNPLALQLMAQGLYNHSTATESDEDDEEISVT

>Inv_P_machaon

MAAVSTTVHLQDTRIKIQDHSDEDHEPYSPNTRDTISPEYNEEDKSEDRTIHTS
SFSIHNVLKKERDTNSPENVFSTDKLLQNTPNFDDSQNSRTSESVSPRLEDDDT
NGNEYRAEISVDDENSCCSDDTVLSVGNEAPVSSYHDNEKSSPDPSHGLSSFK
HIQTHLNAISQLSQNLNVNQPLLLRPSPITPNPLMFLNQPLLFQNPLNQVDLKS
GVIPNRMPMPQSNLNLNTQHFGLNFGLRMKNQTMETRHRSDENRRVNYVTP
KSPDNESGRDFINQNCLKFSIDNILKADFGRRITDPLTKRKSLKSRSFEAKVPV
KEASLPSKPEVLEARVPEVKPTDKGAIDLSKSDDGSSNQSSTSTTTGENGPMV
WPAWVYCTRYSDRPSSGRSKYGTGSRQQRPRTRRPKKPPGEGPTSDEKRPRT
AFSGPQLARLKHEFAENRYLTERRRQALAAELGLAEAQIKIWFQNKRAKIKK
ASGQRNPLALQLMAQGLYNHSTVPLTKEEEELEMKARAREKELQNRH

>En_P_rapae

MAFEDRPSEIRVPGSYQNQYTCTTIQYRSIKTSNSPPPYCYEDVKPVITIPFSIN
NILQPDFGLNAIRKTKLQLRGQRTVSSEAKEAKIAEQRRPDSASSVVSSSSGST
STDTNSQSGSGTLWPAWVYCTRYSDRPSSGPRSRRMKKRITPEEKRPRTAFSA
AQLQRLKHEFAENRYLTERRRQALAAELGLAEAQIKIWFQNKRAKIKKATGH
RNPLAMQLMAQGLYNHSTANESDDDEEISVT

>Inv_P_rapae

MAAVAPIPGQINPENSDDEPAPDDERQVTPFSIHDVLKKERETSPENVFATDK
LLQSTPNFEDASNPESPRLDDTEISVDDNSCSSNDTVLSVGNENPVNDELSFK
HIQTHLNAISQLSHFNTPLLLRPNPIPMFLNQSFYPQVKYQQNEERREDTYRRE
QKEERDFNPNLKFSIDNILKADFGRRITDPLKIRKKRFDGKKEFKEISQVRQEV
RQEVRQEVRTDVKEKGAIDLSKADDSPASSGGTSNSEGPMVWPAWVYCTRY
SDRPSSGPRTRKPKKPPGDTNNDEKRPRTAFSGPQLARLKHEFAENRYLTERR
RQALAAELGLAEAQIKIWFQNKRAKIKKASGHRNPLALQLMAQGLYNHSTIP
LTKEEEELEMKARAREQARQ

>En_partial_J_coenia

LNAIRKTNKIEGPKAVGPNHSILYKPYEVSKEYSKFNYDYLKSKDDFNPLPPL
GGLKQTVSNIGEVHKEVKSVEVQKRPDSASSMVSSTSSGLSTCGSTDNSQTSQ
GNANLWPAWVYCTRYSDRPSSGPRSRRVQKKMNPEEKRPRTAFSAAQLARL
KHEFAENRYLTERRRQALASELGLAEAQIKIWFQNKRAKIKKATGQRNPLAM
QLMAQGLYNHSTVTESDDEEEINVT

>Inv_partial_J_coenia

MVWPAWVYCTRYSDRPSSGRSPRTRRPKKPPGDGNPTDEKRPRTAFSGPQLA
RLKHEFAENRYLTERRRHTLAAELGLAEAQIKIWFQNKRAKIKKASGQRNPL
ALQLMAQGLYNHSTIPLTKEEEELEMKARAREQNRQ

>En_P_xylostella

MAFEDRCSPSQANSPGPVSGRVPAPHAENLMSFCQPSQYTCTTIEPRYERNQP
SPMTIVKVQPASPPSSPMETHGDMDNYQNFYPERPETPDVKPVINPVQEDARF
NDRFRNQCQQVTVPITFSISNILHPEFGSGALRKTNKIEGPKPVGPNHSILYKPY
DLSKPGSVAEHQKGYSFDYLKTKESSHDFNRLPPLGGLRQTVSQIGEVQKEK
EVVRPIEPQRRPDSASSIVSSTSSGAPSTCGSTDAGSQGQGNANLWPAWVYCT
RYSDRPSSGPRSRRSKKKSPSAAEEKRPRTAFSAAQLNRLKHEFAENRYLTER
RRQALAAELGLAEAQIKIWFQNKRAKIKKANGQRNPLALQLMAQGLYNHTT
ATESDEEEEISVT

>Inv_like_P_xylostella

MNGPRTRRPKKPPGEVQNDEKRPRTAFSGPQLARLKHEFAENRYLTERRRQA
LAAELGLAEAQIKIWFQNKRAKIKKASGQRNPLALQLMAQGLYNHSTVPLTR
EEEELEMKARERDAARGAPP

>Inv_P_xylostella

MAAVSTTVHHVQPIKTLEQSDEEPYSPNTRDTTTPDYQEDEKTEERPLTSSFSI
HNVLKKERDSNSPDNVFSTDKLLQSTPNFDQSSPRNSQDSLNSPRFDNEDHNS
NSESVRAEISVDDDVSCCSDDTVLSVGNEAQLSYNSPSESAKSLPASPGTPPN
NLTSFKHIQTHLNAISQLSQNINMNQPLLLRPNPISPNPLMFLNQPHLLFQNPL
LGQDVKGGLLPSRIPIPNPNLNLSPPPSNPFGQLNFLRNPKTQELRNNRNLDEN
RRYQPKSPENESHLNILNQSCLKFSIDNILKADFGRRITDPLTKRKNFKNRTEA
PKSCPKEVAKEVPCSEARVPDVKGGAIDLSKVDEGASNQGSTTSSTSGDGPM
VWPAWVYCTRYSDRPSSGKLFSFQPVFYDQGFTKKRAAALE

>En_V_cardui

MAFEDRCSPNQANSPGPVTGRVPAPHANPGNMSYCPPSQYTCTTIDSRFDRS
NMTVVKVQPNSPPPSPEHNQNEMDYQNYYRPDTPDVKPVLPADRFDPDKRN
RQQPPASIAFSISNILHPEFGLNAIRKTNKIEGPKAVGPNHSILYKPYDLSKEYS

KFNFEYLKSKDDFGALPPLGGLRQTVSNIGEIQKEKEISKSMEAQKRPDSASS
MVSSTSSGALSTCGSTDTNSQTSQGNANLWPAWVYCTRYSDRPSSGPRSRRV
KKKMHPEEKRPRTAFSAAQLARLKHEFAENRYLTERRRQALAGELGLAEAQI
KIWFQNKRAKIKKATGQRNPLAMQLMAQGLYNHSTVTESDEEEEISVT

>Inv_V_cardui

MAAVSSAHLQDQIKIQDHSDDDQDPYSPNTRDTISPDYHDEEKQEERPLHTSS
FSIHNVLKKERETNSPENVFSTDKLLQNTPNFEETSRNSESVSPRIDDEHTDRA
DISVDDNSCCSDDTVLSVGNEAPVSSYHSEKSQDTQGLTSFKHIQTHLNAISQ
LSQNLNINQPILLRPNPITPNPLMFLNQPLLFQNPLMNQVDLKSGIPRMALPQN
SLNLNQQFNMNFVRKSQESLNRVDENRRQNHNVSPKSPDNVSERDFINQNCL
KFSIDNILKADFGRRITDPLTKRKTIKSRQFEVKASPVKETSPAKEVEARVPEV
KPGDKGAIDLSKGDDSGSNASSTPGTASDGPMVWPAWVYCTRYSDRPSSGR
SKYGARSGTQRPRTRRPKKPPGDGTPTDEKRPRTAFSGPQLARLKHEFAENR
YLTERRRQSLAAELGLAEAQIKIWFQNKRAKIKKASGQRNPLALQLMAQGLY
NHSTIPLTKEEEELEMKAREREQSNRQ

>En_D_melanogaster

MALEDRCSPQSAPSPITLQMQHLHHQQQQQQQQQQQMQHLHQLQQLQQLH
QQQLAAGVFHHPAMAFDAAAAAAAAAAAAAAHAHAAALQQRLSGSGSPA
SCSTPASSTPLTIKEEESDSVIGDMSFHNQTHTTNEEEEAEEDDDIDVDVDDTS
AGGRLPPPAHQQQSTAKPSLAFSISNILSDRFGDVQKPGKSMENQASIFRPFEA
SRSQTATPSAFTRVDLLEFSRQQQAAAAAATAAMMLERANFLNCFNPAAYP
RIHEEIVQSRLRRSAANAVIPPPMSSKMSDANPEKSALGSLCKAVSQIGQPAAP
TMTQPPLSSSASSLASPPPASNASTISSTSSVATSSSSSSSGCSSAASSLNSSPSSR
LGASGSGVNASSPQPQPIPPPSAVSRDSGMESSDDTRSETGSTTTEGGKNEMW
PAWVYCTRYSDRPSSGPRYRRPKQPKDKTNDEKRPRTAFSSEQLARLKREFN
ENRYLTERRRQQLSSELGLNEAQIKIWFQNKRAKIKKSTGSKNPLALQLMAQ
GLYNHTTVPLTKEEEELEMRMNGQIP

>Inv_D_melanogaster

MSTLASTRPPPLKLTIPSLEEAEDHAQERRAGGGGQEVGKMHPDCLPLPLVQP
GNSPQVREEEEDEQTECEEQLNIEDEEVEEEHDLDLEDPASCCSENSVLSVGQ
EQSEAAQAALSAQAQARQRLLISQIYRPSAFSSTATTVLPPSEGPPFSPEDLLQ
LPPSTGTFQEEFLRKSQLYAEELMKQQMHLMAAARVNALTAAAAGKQLQM
AMAAAAVATVPSGQDALAQLTATALGLGPGGAVHPHQQLLLQRDQVHHHH
HMQNHLNNNENLHERALKFSIDNILKADFGSRLPKIGALSGNIGGGSVSGSST
GSSKNSGNTNGNRSPLKAPKKSGKPLNLAQSNAAANSSLSFSSSLANICSNSN
DSNSTATSSSTTNTSGAPVDLVKSPPPAAGAGATGASGKSGEDSGTPIVWPA
WVYCTRYSDRPSSGRSPRARKPKKPATSSSAAGGGGGGVEKGEAADGGGVP
EDKRPRTAFSGTQLARLKHEFNENRYLTEKRRQQLSGELGLNEAQIKIWFQN
KRAKLKKSSGTKNPLALQLMAQGLYNHSTIPLTREEEELQELQEAASAAAAK
EPC

Engrailed isoforms (highlighted colors represents different exons)

>En_a_B_anynana

AAAAGTCGACCGGTGNCCGCCGCCGCGAGCGCATCGTGTTGTCAAAAACTCGTTT
GACAGCTTGACAGTTCGGGGATATTTGAATTTGCTAAGTGATGTATTAACATTCAN
CAATTTGAGGACTATACCCGAGAACGTTTTTGAAGGTGTTTTTTAAAGACTTTTTT
GTAATAAAAATGGCGTTTGAAGACCGTTGCAGTCCGAACCAGGCCAACAGCCCCG
GTCCGGTCACCGGCAGAGTCCCTGCGCCTCACTCCGAAGTAAGAAACGNGTACCA

AAGTCAATACACTTGCACGACTATCGATCAAAGGTTTGACAGAACGATGACAGTG
GTGAAAGTGCAGCCGAATTCACCACCGATGAGTCCACTGACGTGAAGCCCATAAT
CCCTGAGTTTGAAGACAAGAGAAACCGACAACCACCACCAACCATACCCTTCTCT
ATCAGCAACATATTACACCCAGAATTCGGTTTGACAGCGATTCGAAAAACGAACA
AAATCGAAGGACCAAAACACGTCGGCCCCAACCACAGCATTTTGTACAAACCTTA
TTTGTCGAACGAGTTATCGAGTTCGAAATTCAATTTCGATTATTTAAAATCTAAGG
ATGATTTCGGTGCATTACCTCCACTTGGCGGTTTGAGGCAGACCGTGTCGAATATT
GGAGAACAGAAGGAGGCACCAAAGATTATAGAGCAGCAGAAGAGGCCAGATTCA
GCCAGCTCTATTGTCTCTTCCACATCTAGCGGGGCTTTATCGACGTGTGGCAGCAC
TGACGCCAACAGCAGTCAAAGCGGGAACAGCAATSKACTGGCCTGCWTGGGTGT
ATTGYACGAGATACAGCGATAGGCCTAGTTCTGGTCCAAGAAGTAGAAGAGTTAA
GAATAAGTCGGTCCCAGAAGAGAAGAGACCCAGAACAGCGTTCAGTGCGGCGCA
ACTTGCTCGACTAAAGCACGAGTTCGCAGAAAATCGCTACCTCACAGAGAGAAGG
CGTCAGACCCTCGCAGCTGAGCTGGGGCTGGCGGAAGCGCAGATTAAGATCTGGT
TCCAGAACAAGCGCGCCAAGATCAAGAAAGCATCAGGTCAGAGGAACCCGCTGG
CGCTGCAGCTGATGGCGCAAGGGTTGTACAACCACAGCACTGCCAACGAGAGTG
ATGATGAAGAGGAGATTAATGTCACGTAGGGTTGTGCCCAGTGACAGCCCAGTGG
ATATGACCTCTGCCTCCGAATCGAAGCCAAACGTCATATCCACTGGGCTATCAGT
ATTGCTTACTTCTAATATATTAATAAATCAAAATCATACAGTTACAGATGTTGTAT
ATGCAGTGGCATAAGCTCATCGATAAGGGTATGCACTAGTAAAAAAAAAAAA

>En_b_B_anynana

AAAAGTCGACCGGTGNCCGCCGCCGCGAGCGCATCGTGTTGTCAAAAACTCGTTT
GACAGCTTGACAGTTCGGGGATATTTGAATTTGCTAAGTGATGTATTAACATTCAN
CAATTTGAGGACTATACCCGAGAACGTTTTTGAAGGTGTTTTTTAAAGACTTTTTT
GTAATAAAAATGGCGTTTGAAGACCGTTGCAGTCCGAACCAGGCCAACAGCCCCG
GTCCGGTCACCGGCAGAGTCCCTGCGCCTCACTCCGAAGTAAGAAACGNGTACCA
AAGTCAATACACTTGCACGACTATCGATCAAAGGTTTGACAGAACGATGACAGTG
GTGAAAGTGCAGCCGAATTCACCACCGATGAGTCCAGAACACAGCCAAAACGAC
ATGGACTACCAAAACTATTTCAGACCCGAAACTCCTGACGTGAAGCCCATAATCC
CTGAGTTTGAAGACAAGAGAAACCGACAACCACCACCAACCATACCCTTCTCTAT
CAGCAACATATTACACCCAGAATTCGGTTTGACAGCGATTCGAAAAACGAACAAA
ATCGAAGGACCAAAACACGTCGGCCCCAACCACAGCATTTTGTACAAACCTTATT
TGTCGAACGAGTTATCGAGTTCGAAATTCAATTTCGATTATTTAAAATCTAAGGAT
GATTTCGGTGCATTACCTCCACTTGGCGGTTTGAGGCAGACCGTGTCGAATATTGG
AGAACAGAAGGAGGCACCAAAGATTATAGAGCAGCAGAAGAGGCCAGATTCAGC
CAGCTCTATTGTCTCTTCCACATCTAGCGGGGCTTTATCGACGTGTGGCAGCACTG
ACGCCAACAGCAGTCAAAGCGGGAACAGCAATSKACTGGCCTGCWTGGGTGTAT
TGYACGAGATACAGCGATAGGCCTAGTTCTGGTCCAAGAAGTAGAAGAGTTAAG
AATAAGTCGGTCCCAGAAGAGAAGAGACCCAGAACAGCGTTCAGTGCGGCGCAA
CTTGCTCGACTAAAGCACGAGTTCGCAGAAAATCGCTACCTCACAGAGAGAAGGC
GTCAGACCCTCGCAGCTGAGCTGGGGCTGGCGGAAGCGCAGATTAAGATCTGGTT
CCAGAACAAGCGCGCCAAGATCAAGAAAGCATCAGGTCAGAGGAACCCGCTRGC
KMTGCARCTGATGGCKCAARGGTTGTACAACCACAGCACTGCCAACGAGAGTGA
TGATGAAGAGGAGATTAATGTCACGTAGGGTTGTGCCCAGTGACAGCCCAGTGGA
TATGACCTCTGCCTCCGATTCCGGAGGGTGTGGGTTCAAATCCAGTCCAGAACAT
GCACTTCCAACTTTTCAATTGTCTTTTCATTTTAAGAAATTAAATATCACGTGTCTC
AAATGGTGAAGGAAAACATCGTGAGGAAACCTGCATACCAGAGAATTTTCTCAAT
TGTCTGCGTGTGTGAAGTCTGCCAATCCGCATTGGGCCAGCGTGGTGGACTATTG
GCCTAACCCCTCTCATTCAGAGAGGAGACTAGAGCTCAGCAGTGAGCCGAATATG
GGTTGATAATGATGAGGGTTGTGCCCACCTACCCAGTTTAGAATGACTTGACACG
TGTAAGCGTCATGTAGAAATGTATTCACGTACGTAAATGGTGGCCCTGGGTCTTA
CTTGTAATCTAGAATTGGGTTAGTTGACTCAAATCAAATTTAATGCAAAAAAAAA
AAA

>En_c_ B_anynana

AAAAGTCGACCGGTGNCCGCCGCCGCGAGCGCATCGTGTTGTCAAAAACTCGTTT
GACAGCTTGACAGTTCGGGGATATTTGAATTTGCTAAGTGATGTATTAACATTCAN
CAATTTGAGGACTATACCCGAGAACGTTTTTGAAGGTGTTTTTTAAAGACTTTTTT
GTAATAAAAATGGCGTTTGAAGACCGTTGCAGTCCGAACCAGGCCAACAGCCCCG
GTCCGGTCACCGGCAGAGTCCCTGCGCCTCACTCCGAAGTAAGAAACGTGTACCA
AAGTCAATACACTTGCACGACTATCGATCAAAGGTTTGACAGAACGATGACAGTG
GTGAAAGTGCAGCCGAATTCACCACCGATGAGTCCAGAACACAGCCAAAACGAC
ATGGACTACCAAAACTATTTCAGACCCGAAACTCCTGACGTGAAGCCCATAATCC
CTGAGTTTGAAGACAAGAGAAACCGACAACCACCACCAACCATACCCTTCTCTAT
CAGCAACATATTACACCCAGAATTCGGTTTGACAGCGATTCGAAAAACGAACAAA
ATCGAAGGACCAAAACACGTCGGCCCCAACCACAGCATTTTGTACAAACCTTATT
TGTCGAACGAGTTATCGAGTTCGAAATTCAATTTCGATTATTTAAAATCTAAGGAT
GATTTCGGTGCATTACCTCCACTTGGCGGTTTGAGGCAGACCGTGTCGAATATTGG
AGAACAGAAGGAGGCACCAAAGATTATAGAGCAGCAGAAGAGGCCAGATTCAGC
CAGCTCTATTGTCTCTTCCACATCTAGCGGGGCTTTATCGACGTGTGGCAGCACTG
ACGCCAACAGCAGTCAAAGCGGGAACAGCAATCTATGGCCTGCATGGGTGTATTG
TACGAGATACAGCGATAGGCCTAGTTCTGGTCCAAGAAGTAGAAGAGTTAAGAAT
AAGTCGGTCCCAGAAGAGAAGAGACCCAGAACAGCGAACAGTGCGGCGCAACTT
GCTCGACTAAAGCACGAGTTCGCAGAAAATCGCTACCTCACAGAGAGAAGGCGT
CAGACCCTCGCAGCTGAGCTGGGGCTGGCGGAAGCGCAGATTAAGATCTGGTTCC
AGAACAAGCGCGCCAAGATCAAGAAAGCATCAGGTCAGAGGAACCCGCTGGCGC
TGCAGCTGATGGCGCAAGGGTTGTACAACCACAGCACTGCCAACGAGAGTGATG
ATGAAGAGGAGATTAATGTCACGTAGGGTTGTGCCCAGTGACAGCCCAGTGGATA
TGACCTCTGCCTCCGATTCCGGAGGGTGTGGGTTCAAATCCAGTCCAGAACATGC
ACTTCCAACTTTTCAATTGTCTTTTCATTTTAAGAAATTAAATATCACGTGTCTCAA
ATGGTGAAGGAAAACATCGTGAGGAAACCTGCATACCAGAGAATTTTCTCAATTG
TCTGCGTGTGTGAAGTCTGCCAATCCGCATTGGGCCAGCGTGGTGGACTATTGGC
CTAACCCCTCTCATTCTGAGAGGAGACTAGAGCTCAGCAGTGAGCCGAATATGGG
TTGATAATGATGAGGGTTGTGCCCACCTACCCAGTTTAGAATGACTTGACACGTGT
AAGCGTCATGTAGAAATGTATTACGTACGTAAATGGTGGCCCTGGGTCTTACTTGT
AATCTAGAATTGGGTAGTTGACTCACATCACATTTAATATGAAAAAATATTAATTT
CATGGAATAAGGATCCGAAATATATCGATTTTAATTAAAAAAAGTCAATGTTTTT
ATAGAAAACATAATTTAATTTTTAATTTTTCTAAACGTCAAAAATTCTGTTNTCCC
ATTTTNGTNGACGTTCCCNATGGTTACNATTTGATTGTCCCTTACAGGCAATTAAC
GGNCGAAATAATTTGATTAACTGGNTGNTTGTCATAAAAAAAATTTATTCTTNAN
GNTTTCCNTNGTCNCCAAANAGTATGNANTANTNTTTTTTTTCTAGNTAGAACCAA
WGGTKCATTYWTTATMCKKGGGGAWWAGGNGAATNARCYNTTTRGMCCCTKTT
MAMMCGGGTTCACCYGSSSKAYTSSRSTGANWACTTTCTYAGMMKKTKMTAARK
TCNRACACMYAWAMMCMMTWYCCYACTTACCCTACCCTACCCCTACCCTACCC
CTACCTTTTAAGCCCGAGATCTTTCCAACGAATGCAAAACCGTGGAAGTCGGTTG
GTGCGTTCTAGAGTTATAGCATCAGGAAGGAAAACCCGACTTATTTTTATATATTA
GATTACAGGAGATTTTGAATTAAATTTCCTTTGGAACGAAATCTATCATTTTAATA
ACATGCGGGTAGTTCAATAAATAAAAAATGACACATCATGGTAATATATTTCTAA
GCACAAAATAGTTCCTGTACTTTAAATAATATTATATTAATAAATAACTCTTTGTT
ATCAGTAAGCCCTCATTCGCACGAGAGTTTTTTTAACGGACGTTAGAAAAACGTT
CAAATACACCGAATGCATTCCCAAGTGTATGTTCACACGACAGAGTTTTTAAAAC
ACGACGTTTTTTTTTCGAGGTGTATTGCATTTTCCATTTTGGGCGTTGGAAATAGAC
CTTCGTTTAACTTCATACTAAAAAAACTCTCGTGCGAATGAGAGCTTAGCGTCAA
ATGAGATACCTTGTGAAGAATCGTCCAATAGCGTCAGAGTTGAAAATAGATAGAT
AATTGAAATATAGCTTTCTCTTTTCTCCCAAAGCAACATAAGAGATAGCGATATTT
TTGATGTTCTTCATTAACTGTTTATCTCTCATCTCTAGTTACTCGTTTATCTTGGTG
CGCGTTTAAGTATGCCTTTTAGACACAAAAAAAAAAAAA

## Chapter 5: Conclusions and Future Directions

My work illustrates the mechanisms underlying two specific types of biological patterning using *Bicyclus anynana* as a model organism. The first pattern I explored during my Ph.D. journey is the arrangement of hollow cuticular tubes in the wings of butterflies called veins. The second is the development of concentric color pattern on the wings of butterflies called eyespots which play an important role in predator deflection and sexual signaling.

In the first chapter, I provided a brief introduction to two main theories of biological patterning in nature that are involved in governing vast arrays of biological patterns. The first theory is known as reaction-diffusion and involves at least two diffusible molecules that interact with each other to produce patterns from a homogeneous state. The second theory is called positional information and is a form of reaction-diffusion except a single morphogen is usually involved and no reaction with another morphogen takes place. Here molecules diffusing from a source provide information to a field of cells by regulating the expression of downstream targets responding at different concentration thresholds. I also reviewed how the combination of these two theories can work in synergy to produce patterns such as the eyespots of butterflies and embryos in humans.

In the second chapter, I studied the molecular mechanisms underlying the specific arrangement of veins in the wings of butterflies by examining the expression and function of an orthologous set of genes previously implicated in venation patterning of *Drosophila melanogaster* wings. I suggest

mechanisms by which complex venation patterns present in the ancestral wings of insects might have simplified into less complex patterns over the course of evolution. My main findings are that butterflies have more expression domains of genes that provide critical boundaries for vein differentiation. Derived insects such as *D. melanogaster* lost some of these domains and as a result, lost veins. Further simplification in venation occurred due to the loss of vein developmental programs and loss of vein maintenance genes at the same gene expression boundaries. These findings contrast with what has been believed in the scientific community where simplified venation was thought to be just due to loss of vein specific genes or fusion of veins when the wing develops.

In the third chapter, I explored the involvement of the eye development gene *optix* in the eyespots of butterflies and its interaction with other genes involved in eyespot development. Optix has been considered a hotspot gene where allelic variants regulate a variety of red pigmentation patterns in a wide array of *Heliconius* butterflies. In this chapter, I showed how Optix is involved in the pigmentation and formation of scale ultrastructure of the eyespot orange ring and how it interacts with another eyespot gene *spalt*. The results point to the presence of a positional information model defining the rings of the eyespot which leads to the expression of *optix* in the outer orange ring where it determines the identity of the orange scales by changing both their pigmentation and scale ultrastructure.

In the fourth chapter, I explored the involvement of gene duplicates in patterning the eyespots. Evolutionary innovations such as beetle horns, flowers, vision in humans, and chambers of hearts have often been associated

with the appearance of novel gene duplicates and hence, their retention in genomes. Eyespots are one such evolutionary innovation. Over 180 genes have been shown to be associated with the development of eyespots, many of which have multiple gene duplicates. However, these gene duplicates and their association with the development of eyespots have never been studied before. In this chapter, I explored the homeodomain transcription factor *engrailed* and its paralogs *invected* and *invected-like* in the wings of *Bicyclus anynana* butterflies. The differential expression of the paralogs indicates an ancestral expression in the posterior compartment followed by a gain of expression in the eyespots. By dating the origin of both duplication events, however, I concluded that they predate the origin of eyespots by at least 60 mya, and hence the data does not support the retention of the multiple *en* gene duplicates in the genome via their involvement with the novel eyespot evolutionary innovation.

The work presented in this thesis, however, is far from complete and would require further exploration to enlighten our understanding of these amazing patterns in nature. Below are few of the future directions where research could shape our understanding of every data chapter:

**Chapter 2: Molecular mechanisms underlying simplification of venation pattern in holometabolous insects**

Exploring the expression domains and function of unexplored genes involved in the mechanisms of positional information in *B. anynana* butterflies.

This chapter presented expressional and functional data on a few of the genes involved in venation patterning in *B. anynana*. However, there are dozens of

genes that remained unexplored, and if examined might further our understanding of venation patterns in insects. A few of the genes involved in venation in *D. melanogaster* that can be explored in future projects include:

i) Optomotor-blind (Omb)- *omb* is a downstream target of Dpp and is involved in the formation of the Cu1 (L5) vein (Cook et al., 2004).

ii) Brinker (Brk)- Brinker is a repressor of Dpp signaling (Campbell and Tomlinson, 1999b) and along with Omb forms the boundary of the Cu1 (L5) vein (Cook et al., 2004).

iii) Knirps (Kni) – Kni is involved the formation of the R2+3 (L2) vein (Lunde et al., 1998). The position of this vein is determined by the interaction of Spalt, Optix, and Aristaless (Martín et al., 2017).

iv) Abrupt (Abt) and Iroquis complex (Iro) – Both these genes are involved in the formation of the Cu1 (L5) vein (Cook et al., 2004; Gömez-Skarmeta et al., 1996).

v) Vein – Vein is involved the formation of R4+5 (L3) and M1 (L4) veins where it activates Egfr signaling (Schnepp et al., 1996).

<u>Exploring a reaction-diffusion system in venation patterning in *B. anynana* butterflies</u>

Recent reports on insect venation discuss the possibility of a reaction-diffusion mechanism involved in patterning the branches and cross veins of insects. A report on the planthopper *Orosanga japonicus* indicate that a mechanism of positional information followed by reaction-diffusion is involved in the branching venation pattern in this species (Yoshimoto and Kondo, 2012).

Furthermore, a report on dragonflies and damselflies indicates that a mechanism of reaction-diffusion is involved in patterning the cross veins in these insects (Hoffmann et al., 2018a). In *B. anynana* we observe cross veins as well as vein branching.

A reaction-diffusion mechanism might be involved in venation patterning in *B. anynana* butterflies. In *B. anynana* venation we observe a mechanism of positional information similar to that in *D. melanogaster* as explained in chapter 2. This mechanism, however, fails to explain all the features of venation in *B. anynana* such as Radial vein branching in the forewing. I am planning to explore the possibility of a reaction-diffusion mechanism also being involved. A reaction-diffusion mechanism, however, requires at least two diffusible molecule that can interact with each other. My first quest will be to search for these two molecules (or signaling modules).

The two likely signaling modules involved in a reaction-diffusion system that patterns the veins in *B. anynana* are WNTs and BMPs. I observed the presence of Armadillo (Arm: The signal transducer of WNT signaling) in a homogeneous state in the early larval wings, which during later stage gets concentrated along the veins, in the eyespot center, and the wing margin (**Fig 5.1A**). While, *decapentaplegic* (*dpp*) (Dpp: a ligand of BMP signaling) is present only in a single stripe along the anterior-posterior boundary during an early stage of wing development, but is present everywhere in the wing blade except for the veins, in the eyespot center, and the wing margin during later stages (**Fig 5.1B**). This complementary pattern is similar to that of a substrate-depletion reaction-diffusion system (**Fig 5.2**) where the WNT module can acts as a substrate and the BMP module as an activator.

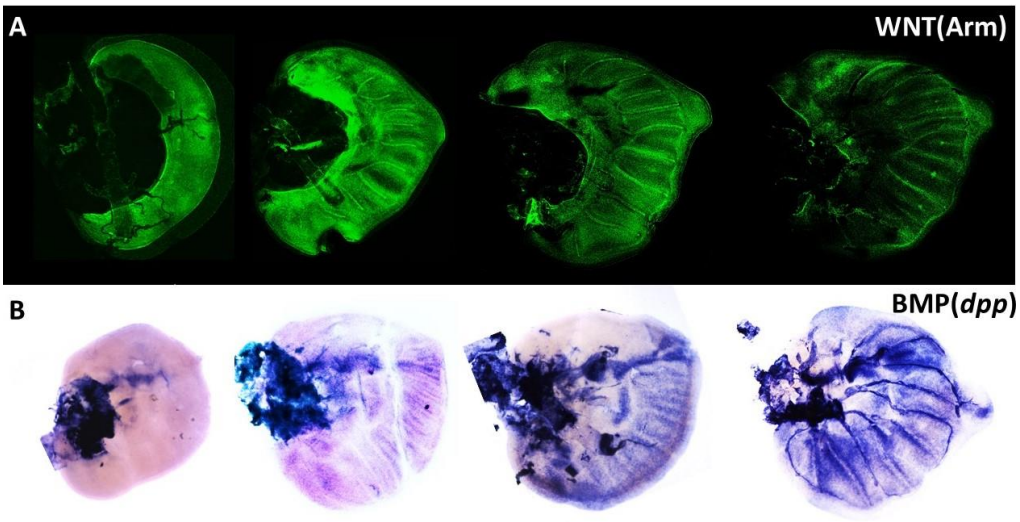



**Figure 5.1. Expression of Armadillo (Arm) and *decapentaplegic* (*dpp*) during the *B. anynana* larval wing development.** (**A**) Arm is more homogeneous during the early larval wing development, while later in development it gets focused along the veins, in the eyespot center and along the wing margin. (**B**) *dpp* is present only in a single stripe along the anterior-posterior boundary during an early stage of wing development but during a later stage is present everywhere in the wing blade except for the veins, in the eyespot center, and the wing margin.

Inhibition of BMP and WNT signaling using drug antagonist results in loss of venation pattern (**Fig. 5.2**). Culturing the wings for one day under the presence of the BMP inhibitor Dorsomorphin (Yu et al., 2008) resulted in loss of venation pattern and a more homogeneous expression of Rhomboid (Rho: A vein maker gene) (**Fig. 5.2C**). Similar loss of venation pattern is observed with the WNT inhibitor drug iCRT3 (Lee et al., 2013) along with disruptions of cell adhesion. This is most likely due to reduced production of Arm which also acts as a cell adhesion molecule (**Fig. 5.2D**) (Peifer et al., 1993). WT wings had an intact Rho expression pattern, while DMSO treated wings had reduced expression of Rho (**Fig. 5.2A and B**).

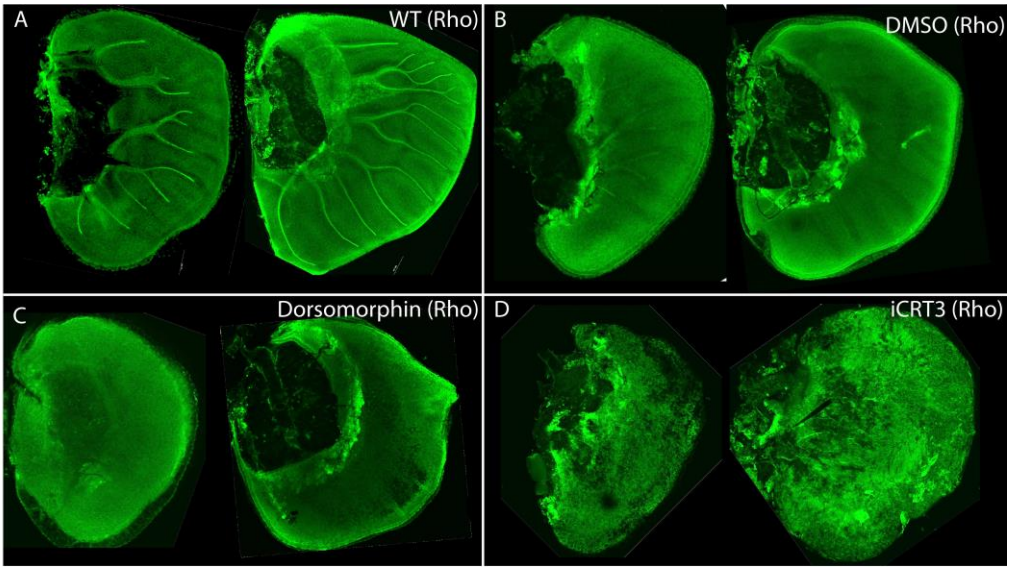



**Figure 5.2. Expression of the vein marker gene Rhomboid (Rho) in WT, DMSO, Dorsomorphin and iCRT3 treated wings**. (**A**) WT expression of Rho shows clear expression along the veins with intact venation pattern. (**B**) DMSO treated wings have the venation pattern but low levels of Rho along the veins. (**C**) Wings treated with Dorsomorphin (in DMSO) have lost their venation pattern. (**D**) Wings treated with iCRT3 have lost their venation pattern and also lost cell to cell adhesion, likely due to reduced expression of Arm in the intercellular space.

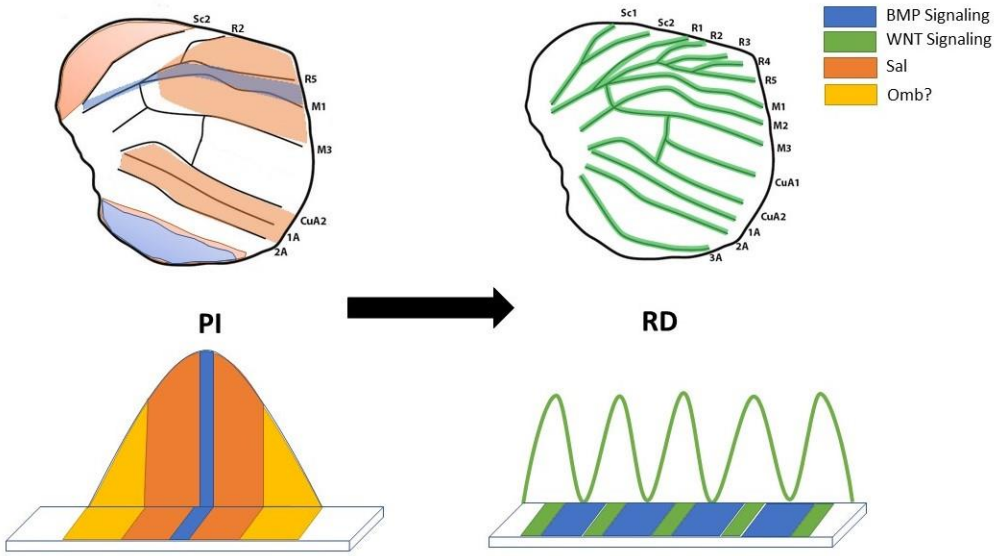



**Figure 5.3: Reaction-diffusion + positional information model to explain venation patterning in *B. anynana.*** In this model PI is upstream of RD. The initial veins are positioned by PI based on different concentration thresholds of the *dpp* BMP morphogen. Once the positioning of the veins is setup, RD acts on the initial veins and produces the bifurcation and additional veins from the initial positional information. This model is similar to what has been proposed for *Orosanga japonicus* (Yoshimoto and Kondo, 2012).

The results I have obtained until now hint at the possibility that both positional information and reaction-diffusion might be involved in *B. anynana* wing venation (**Fig. 5.3**). In a recent paper, Green and Sharpe illustrated models to describe the interaction between positional information and reaction-diffusion. One for these models include positional information upstream of reaction-diffusion where the positional information sets up the initial boundaries upon which the reaction-diffusion acts (Green and Sharpe, 2015). I believe such a system might be in play in *B. anynana* venation as well. This project, however, needs further exploration and will be of interest for a future research project.

**Chapter 3: Optix: An eye development gene paints the eyespots of *Bicyclus anynana* butterflies via a possible positional information mechanism.**

Function of Dpp in forming the rings of an eyespot

In chapter 3, Dpp has been proposed as the morphogen secreted from the center of the eyespot. The chapter proposes how different concentrations of Dpp can activate downstream genes such as *spalt* and *optix* in the black and orange ring cells respectively. This, however, needs experimental verification.

One way I am pursuing the involvement of Dpp in eyespot ring differentiation is via bead implantations in the developing pupal wing of *B. anynana*. In this experiment agarose beads will be hybridized with the BMP (Dpp) inhibitor Dorsomorphin which is capable of blocking Dpp binding sites (Yu et al., 2008). The beads will then be implanted near the eyespot centers. If the

colored rings fails to form and the downstream genes fail to express, then the results will indicate that Dpp plays a role in activating the genes responsible for the formation of the eyespot rings. Beads can also be hybridized with Dpp recombinant protein. If the implanted beads, in eyespot free regions of the wing, lead to the formation of ectopic eyespots, then this will confirm that Dpp is sufficient for eyespot ring differentiation.

Studying the interaction of Optix and Spalt with other genes involved in the development of eyespot such as Engrailed and Distal-less

Chapter 3 explores the interaction of Optix expressed in the orange ring scale cells with Spalt expressed in black scale cells. There are however, other genes such as Engrailed co-expressed with Optix (**Fig. 5.4**) and Distal-less co-expressed with Spalt in the black scales (Brakefield et al., 1996; Brunetti et al., 2001). It is still unclear how the expressions of these genes are controlled and would be interesting to knock-out one of the genes and look at the expression pattern of the others using immunostaining.

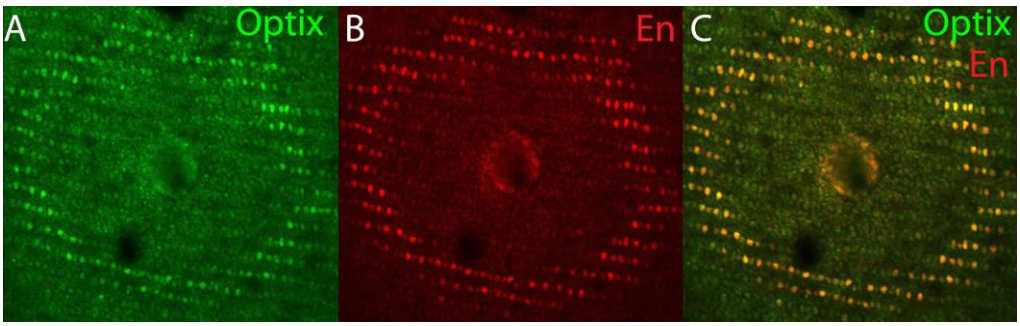



**Figure 5.4. Co-expression of Optix and Engrailed (En) in the orange ring of the eyespots**. Both Optix and En are present in the cells that will form the orange scales of the eyespot.

Studying the cis-elements of Optix to find the regions controlling expression at different domains on the wing of *B. anynana* and other butterfly species

Optix is expressed in many different regions of the wing such as the upper anterior and lower posterior domain of the larval wing of *B. anynana*; orange ring of the eyespot, silver scales, and beige transversal band during pupal wing development of *B. anynana*; and along the veins during the pupal wing development of *Pieris canidia* (**Fig. 5.5**). Studies have proposed that a gene can be controlled in the spatial-temporal domains either via different cis-elements or via a single pleotropic cis-element (Carroll, 2008; Sabarís et al., 2019). Future studies can look for the open chromatin region around Optix to see which cis-element/s are controlling the expression in these domains of the wing within *B. anynana* and in different species of butterflies.

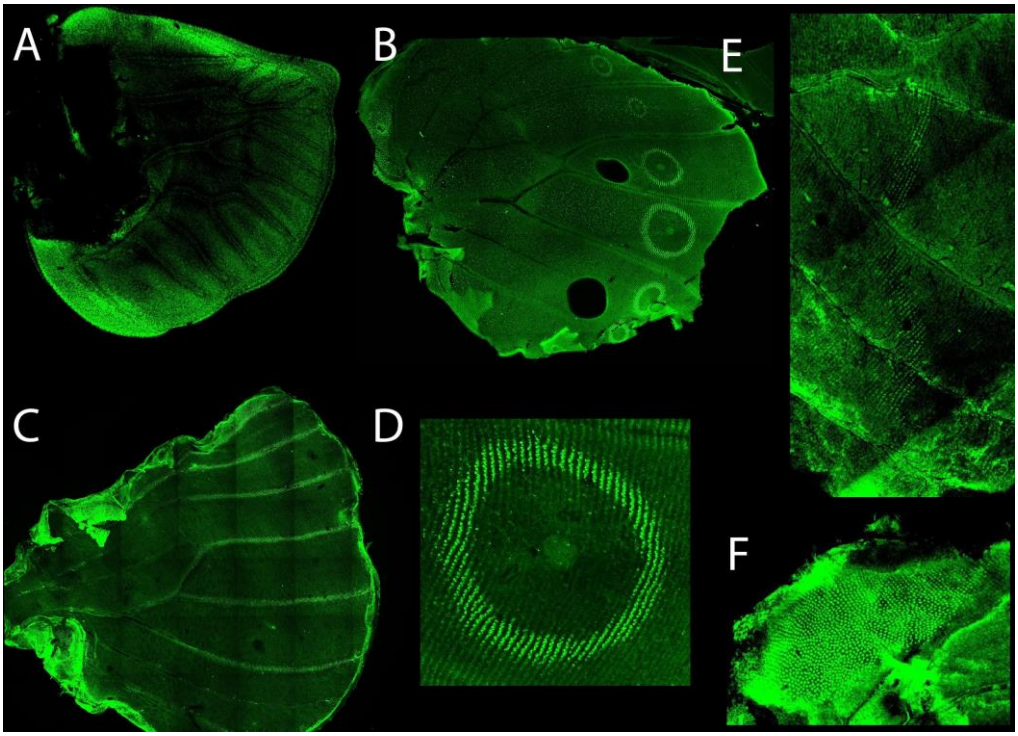



**Figure 5.5. Expression of Optix in the larval wing of B. anynana and pupal wings of *B. anynana* and *P. canidia*.** (**A**) During the larval stage Optix is expressed in the upper anterior and lower posterior domain of *B. anynana*. (**B**) Expression of Optix a 48 hrs old pupal wing of *B. anynana* where it is expressed in the (**D**) eyespot, (**E**) beige transversal band, and (**F**) the silver (coupling) scales. (**C**) Expression of Optix in a 24 hrs pupal wing of *P. canidia* showing expression along the veins.

**Chapter 4: Expression of multiple *engrailed* family genes in eyespots of *Bicyclus anynana* butterflies does not implicate the duplication events in the evolution of this morphological novelty**

Exploring the function of *engrailed* (*en*), *invected* (*inv*) and *invected-like* (*inv-like*)

Chapter 4 explores the expression domains of *en, inv* and *inv-like* in the pupal wings of *B. anynana* where they are differentially expressed. The function of these genes, however, remained unexplored and hence their role in the formation of eyespots is still unclear. A project can be designed to study the function of these genes using CRISPR-Cas9 and the role they play in the differentiation of eyespots.

Exploring the function of different isoforms of *en*

The data from chapter 4 indicate that there are multiple isoforms of *en* with different lengths of the 3' UTR. Studies have shown that isoforms with different length play important roles in many biological processes (Elkon et al., 2013). The transcript of *en* is expressed in the eyespot centers, in the orange ring and in the posterior compartment of the pupal wing. It would be interesting to study if these different isoforms play any role in localizing *en* in these different regions of the wing.

Studying the possible common cis-elements that might be controlling the expression domains of *en*, *inv*, and *inv-like*

The discussion section in chapter 4 I proposed a model which involves different or a common cis-elements that can control expression of the paralogs in different domains of the wing. The regulation of gene expression is an

important field in biological research and future work could focus on exploring these cis-elements and the role they play in controlling the expression of the different *engrailed* gene copies. This work can further our understanding of gene regulation around a morphological novelty that reuses old genes.

Exploring the expression domains of *en*, *inv*, and *inv-like* in other butterfly species with eyespots

The chapter looked at the expression domains of the paralogs only in one species of butterfly *B. anynana*. Butterflies such as *J. coenia* and *V. cardui* have different domains where En/Inv (using 4F11 antibody) is expressed (Brunetti et al., 2001). The evolution of these genes and how they might have acquired the differential expression domains in different regions of the eyespot will require further exploration in other species of butterflies with eyespots.

Biological patterning in living systems has remained a topic of prime interest in the scientific community. Starting from systems as simple as bacterial and fungal colonies to as advanced as human tissues, biological patterning has seen remarkable advances in the recent few decades. The use of model systems with a plethora of tools available at their disposal relieves much of the hurdles researchers face to understand and answer how biological systems become patterned. In the present thesis, I used the butterfly *Bicyclus anynana* as a model system to illustrate how we can use the tools available in this simple system (compared to vertebrates) to answer complex questions such as the patterning of veins (a structure with functions similar to both bones and veins in vertebrates) and color patterns on wing tissues.

In the second chapter of my thesis, I showed how evolution has tweaked the expressed pattern of genes to simplify venation on the wings of flies when compared to those of butterflies. The third chapter shows the presence of components of the same venation gene network to be involved in patterning of the colorful eyespots on the wings, illustrating a remarkable example of how nature is using the same gene regulatory networks to pattern different domains of an organism. Biological complexity often increases when genes duplicate in genomes and copies either loose ancestral functions or gain new functions. I examined how the paralogs of a venation patterning gene have become associated with eyespot patterning in the fourth chapter. Here, genes involved in venation patterning are also expressed in different and novel domains in eyespots, providing further insight into the reuse of old genes in the development of novel traits.

The data presented in the thesis provides insight into our understanding of pattern formation across biological tissues. Morphogens such as Dpp (BMP ligand) and Wg (WNT ligand) play roles in patterning myriads of systems across the living world. Some examples include patterning of bones in mice limbs (Raspopovic et al., 2014), embryogenesis in humans (Tewary et al., 2017), hair follicles in mice (Sick et al., 2006), segmentation in insects (Green and Sharpe, 2015) to name a few. The combined action of patterning mechanism such as positional information and reaction-diffusion proposed in the patterning of eyespots in chapter 3 has also been shown to pattern human embryos (Tewary et al., 2017) indicating that simple systems, such as eyespots, can provide insights on patterning in vertebrate systems. *Engrailed* and its paralogs studied in chapter 4 have roles in many different aspects of

animal development such as the invertebrate embryogenesis and development of neurons. Understanding the expression pattern and the interaction of these copies will enhance our understanding of more complex systems and perhaps someday would be of use for much broader applications such as in human development.

## BIBLIOGRAPHY


**Abbasi, R. and Marcus, J. M.** (2017). A new A-P compartment boundary and organizer in holometabolous insect wings. *Sci. Rep.* **7**, 16337.

**Abouheif, E. and Wray, G.** (2002). Evolution of the gene network underlying wing polymorphism in ants. *Science (80-. ).* **297**, 249–252.

**Abzhanov, A. and Kaufman, T. C.** (2000). Evolution of distinct expression patterns for engrailed paralogues in higher crustaceans (Malacostraca). *Dev. Genes Evol.* **210**, 493–506.

**Akiyama, T. and Gibson, M. C.** (2015). Decapentaplegic and growth control in the developing Drosophila wing. *Nature* **527**, 375–8.

**Al Khatib, A., Siomava, N., Iannini, A., Posnien, N. and Casares, F.** (2017). Specific expression and function of the Six3 optix in Drosophila serially homologous organs. *Biol. Open* **6**, 1155–1164.

**Alexandre, C., Lecourtois, M. and Vincent, J.** (1999). Wingless and Hedgehog pattern Drosophila denticle belts by regulating the production of short-range signals. *Development* **126**, 5689–5698.

**Allena, R., Muñoz, J. J. and Aubry, D.** (2013). Diffusion-reaction model for Drosophila embryo development. *Comput. Methods Biomech. Biomed. Engin.* **16**, 235–248.

**Arun, A., Baumlé, V., Amelot, G. and Nieberding, C. M.** (2015). Selection and validation of reference genes for qRT-PCR expression analysis of candidate genes involved in olfactory communication in the butterfly bicyclus anynana. *PLoS One* **10**, 1–17.

**Ball, P.** (2015). Forging patterns and making waves from biology to geology: a commentary on Turing (1952) “The chemical basis of morphogenesis.” *Philos. Trans. R. Soc. B Biol. Sci.* **370**, 20140218–20140218.

**Banerjee, T. and Monteiro, A.** (2018). *CRISPR-Cas9 Mediated Genome Editing in Bicyclus anynana Butterflies*.

**Banerjee, T. Das and Monteiro, A.** (2020a). Dissection of Larval and Pupal Wings of Bicyclus anynana Butterflies. *Methods Protoc.* **3**, 5.

**Banerjee, T. Das and Monteiro, A.** (2020b). Molecular mechanism underlying venation patterning in butterflies. *bioRxiv* 1–23.

**Banerjee, T. Das, Ramos, D. and Monteiro, A.** (2020). Expression of multiple engrailed family genes in eyespots of Bicyclus anynana butterflies does not implicate the duplication events in the evolution of this morphological novelty. *Front. Ecol. Evol.* **8**, 1–12.

**Bangi, E. and Wharton, K.** (2006). Dpp and Gbb exhibit different effective ranges in the establishment of the BMP activity gradient critical for Drosophila wing patterning. *Dev. Biol.* **295**, 178–193.

**Baratte, S. and Bonnaud, L.** (2009). Evidence of Early Nervous Differentiation and Early Catecholaminergic Sensory System during Sepia officinalis Embryogenesis. **549**, 539–549.

**Barrio, R. and De Celis, J. F.** (2004). Regulation of spalt expression in the Drosophila wing blade in response to the Decapentaplegic signaling pathway. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 6021–6026.

**Beldade, P. and Brakefield, P. M.** (2002). The genetics and evo-devo of butterfly wing patterns. *Nature* **3**, 442–452.

**Beldade, P. and Peralta, C. M.** (2017). Developmental and evolutionary mechanisms shaping butterfly eyespots. *Curr. Opin. Insect Sci.* **19**, 22–29.

**Bhardwaj, S., Prudic, K. L., Bear, A., Dasgupta, M., Wasik, B. R., Tong, X., Cheong, W. F. and Wenk, M. R.** (2018). Sex Differences in 20-Hydroxyecdysone Hormone Levels Control Sexual Dimorphism in Bicyclus anynana Wing Patterns. *Mol. Biol. Evol.* **35**, 465–472.

**Bhardwaj, S., Jolander, L. S. H., Wenk, M. R., Oliver, J. C., Frederik Nijhout, H. and Monteiro, A.** (2020). Origin of the mechanism of phenotypic plasticity in satyrid butterfly eyespots. *Elife* **9**, 1–13.

**Biehs, B., Sturtevant, M. a and Bier, E.** (1998). Boundaries in the Drosophila wing imaginal disc organize vein-specific genetic programs. *Development* **125**, 4245–4257.

**Bier, E.** (2000). Drawing lines in the Drosophila wing: Initiation of wing vein development. *Curr. Opin. Genet. Dev.* **10**, 393–398.

**Blair, S. S.** (1992). Engrailed expression in the anterior lineage compartment of the developing wing blade of Drosophila. *Development* **115**, 21–33.

**Blair, S. S.** (2007). Wing vein patterning in Drosophila and the analysis of intercellular signaling. *Annu. Rev. Cell Dev. Biol.* **23**, 293–319.

**Braendle, C., Miura, T., Bickel, R., Shingleton, A. W., Kambhampati, S. and Stern, D. L.** (2003). Developmental origin and evolution of bacteriocytes in the aphid-Buchnera symbiosis. *PLoS Biol.* **1**,.

**Brakefield, P. M., Gates, J., Keys, D., Kesbeke, F., Wijngaarden, P. J., Monteiro, a, French, V. and Carroll, S. B.** (1996). Development, plasticity and evolution of butterfly eyespot patterns. *Nature* **384**, 236–242.

**Brockington, S. F., Yang, Y., Gandia-Herrero, F., Covshoff, S., Hibberd, J. M., Sage, R. F., Wong, G. K. S., Moore, M. J. and Smith, S. A.** (2015). Lineage-specific gene radiations underlie the evolution of novel betalain pigmentation in Caryophyllales. *New Phytol.* **207**, 1170–1180.

**Brook, W. J., Diaz-Benjumea, F. J. and Cohen, S. M.** (1996). Organizing spatial pattern in limb development. *Annu. Rev. Cell Dev. Biol.* **12**, 161–180.

**Brown, S. J., Patel, N. H. and Denell, R. E.** (1994). Embryonic Expression

of the Single Tribolium engrailed Homology. *Dev. Genet.* **15**, 7–18.

**Brunetti, C. R., Selegue, J. E., Monteiro, A., French, V., Brakefield, P. M. and Carroll, S. B.** (2001). The generation and diversification of butterfly eyespot color patterns. *Curr. Biol.* **11**, 1578–1585.

**Campbell, G. and Tomlinson, A.** (1999a). Transducing the Dpp Morphogen Gradient in the Wing of Drosophila. *Cell* **96**, 553–562.

**Campbell, G. and Tomlinson, A.** (1999b). Transducing the Dpp Morphogen Gradient in the Wing of Drosophila: Regulation of Dpp Targets by brinker. *Cell* **96**, 553–562.

**Carroll, S. B.** (2008). Evo-Devo and an Expanding Evolutionary Synthesis: A Genetic Theory of Morphological Evolution. *Cell* **134**, 25–36.

**Carroll, S. B., Gates, J., Keys, D. N., Paddock, S. W., Grace, E. F., Selegue, J. E. and Williams, J. A.** (1994). Pattern Formation and Eyespot Determination in Butterfly Wings. *Science (80-. ).* **265**, 109–114.

**Challis, R. J., Kumar, S., Dasmahapatra, K. K., Jiggins, C. D. and Blaxter, M.** (2016). Lepbase: The Lepidopteran genome database. *bioRxiv* 056994.

**Cheng, Y., Brunner, A. L., Kremer, S., DeVido, S. K., Stefaniuk, C. M. and Kassis, J. A.** (2014). Co-regulation of invected and engrailed by a complex array of regulatory sequences in Drosophila. *Dev. Biol.* **395**, 131–143.

**Chintapalli, R. T. V. and Hillyer, J. F.** (2016). Hemolymph circulation in insect flight appendages: physiology of the wing heart and circulatory flow in the wings of the mosquito *Anopheles gambiae*. *J. Exp. Biol.* **219**, 3945–3951.

**Combes, S. A.** (2003). Flexural stiffness in insect wings I. Scaling and the influence of wing venation. *J. Exp. Biol.* **206**, 2979–2987.

**Comstock, J.H. Needham, J. G.** (1898). THE WINGS OF INSECTS. *Am. Nat.* **XXXII**, 231–257.

**Comstock, J. . H. . and Needham, J. . G. .** (1898). The Wings of Insects. *Am. Soc. Nat.* **32**, 43–48.

**Conant, G. C. and Wolfe, K. H.** (2008). Turning a hobby into a job: How duplicated genes find new functions. *Nat. Rev. Genet.* **9**, 938–950.

**Condron, B. G., Patel, N. H. and Zinn, K.** (1994). Engrailed Controls Glial/Neuronal Cell Fate Decisions At the Midline of the Central Nervous System. *Neuron* **13**, 541–554.

**Conley, C. a, Silburn, R., Singer, M. a, Ralston, a, Rohwer-Nutter, D., Olson, D. J., Gelbart, W. and Blair, S. S.** (2000). Crossveinless 2 contains cysteine-rich domains and is required for high levels of BMP-like activity during the formation of the cross veins in Drosophila. *Development* **127**, 3947–3959.

**Connahs, H., Rhen, T. and Simmons, R. B.** (2016). Transcriptome analysis of the painted lady butterfly, Vanessa cardui during wing color pattern development. *BMC Genomics* **17**, 1–16.

**Connahs, H., Tlili, S., van Creij, J., Loo, T. Y. J., Banerjee, T. Das, Saunders, T. E. and Monteiro, A.** (2019). Activation of butterfly eyespots by Distal-less is consistent with a reaction-diffusion process. *Development* **146**, 1–12.

**Cook, O., Biehs, B. and Bier, E.** (2004). brinker and optomotor-blind act coordinately to initiate development of the L5 wing vein primordium in Drosophila. *Development* **131**, 2113–24.

**Couso, J. P., Bishop, S. A. and Martinez Arias, A.** (1994). The wingless signalling pathway and the patterning of the wing margin in Drosophila. *Development* **120**, 621–636.

**Damen, W. G. M.** (2002). Parasegmental organization of the spider embryo implies that the parasegment is an evolutionary conserved entity in arthropod embryogenesis. *Development* **129**, 1239–1250.

**Daniels, E. V., Murad, R., Mortazavi, A. and Reed, R. D.** (2014). Extensive transcriptional response associated with seasonal plasticity of butterfly wing patterns. *Mol. Ecol.* **23**, 6123–6134.

**Davey, J., Chouteau, M., Barker, S. L., Maroja, L., Baxter, S. W., Simpson, F., Joron, M., Mallet, J., Dasmahapatra, K. K. and Jiggins, C. D.** (2015). Major Improvements to the Heliconius melpomene Genome Assembly Used to Confirm 10 Chromosome Fusion Events in 6 Million Years of Butterfly Evolution. *G3* **6**, 695–708.

**De Celis, J. F.** (2003). Pattern formation in the Drosophila wing: The development of the veins. *BioEssays* **25**, 443–451.

**De Celis, J. F. and Diaz-Benjumea, F. J.** (2003). Developmental basis for vein pattern variations in insect wings. *Int. J. Dev. Biol.* **47**, 653–663.

**Dion, E., Monteiro, A. and Yew, J. Y.** (2016). Phenotypic plasticity in sex pheromone production in Bicyclus anynana butterflies. *Sci. Rep.* **6**, 1–13.

**Doyle, J. J. and Coate, J. E.** (2019). Polyploidy, the nucleotype, and novelty: The impact of genome doubling on the biology of the cell. *Int. J. Plant Sci.* **180**, 1–52.

**Eaton, S. and Kornberg, T. B.** (1990). Repression of ci-D in posterior compartments of Drosophila by engrailed. *Genes Dev.* **4**, 1068–1077.

**Economou, A. D., Ohazama, A., Porntaveetus, T., Sharpe, P. T., Kondo, S., Basson, M. A., Gritli-Linde, A., Cobourne, M. T. and Green, J. B. A.** (2012). Periodic stripe formation by a Turing mechanism operating at growth zones in the mammalian palate. *Nat. Genet.* **44**, 348–351.

**Elkon, R., Ugalde, A. P. and Agami, R.** (2013). Alternative cleavage and polyadenylation: Extent, regulation and function. *Nat. Rev. Genet.* **14**, 496–506.

**Ferguson, L., Lee, S. F., Chamberlain, N., Nadeau, N., Joron, M., Baxter, S., Wilkinson, P., Papanicolaou, A., Kumar, S., Kee, T. J., et al.** (2010). Characterization of a hotspot for mimicry: Assembly of a butterfly wing transcriptome to genomic sequence at the HmYb/Sb locus. *Mol. Ecol.* **19**, 240–254.

**Fjose, A., Mcginnis, W. J. and Gehring, W. J.** (1985). Isolation of a homoeo box-containing gene from the engrailed region of Drosophila and the spatial distribution of its transcripts. *Nature* **I**,.

**Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. and Postlethwait, J.** (1999). Preservation of Duplicate Genes by Complementary , Degenerative Mutations. *Genetics* **151**, 1531–1545.

**Foronda, D., Pérez-Garijo, A. and Martín, F. A.** (2009). Dpp of posterior origin patterns the proximal region of the wing. *Mech. Dev.* **126**, 99–106.

**Fristrom, D., Gotwals, P., Eaton, S., Kornberg, T. B., Sturtevant, M. and Fristrom, J. W.** (1994). blistered : a gene required for vein / intervein formation in wings of Drosophila. **2671**, 2661–2671.

**Fujioka, M., Yusibova, G. L., Patel, N. H., Brown, S. J. and Jaynes, J. B.** (2002). The repressor activity of Even-skipped is highly conserved, and is sufficient to activate engrailed and to regulate both the spacing and stability of parasegment boundaries. *Development* **129**, 4411–21.

**Funakoshi, Y., Minami, M. and Tabata, T.** (2001). mtv shapes the activity gradient of the Dpp morphogen through regulation of thickveins. *Development* **128**, 67–74.

**Gantz, V. M.** (2015). The mutagenic chain reaction: from Evo-Devo to active genetics. *Calif. Digit. Libr. (University California, San Diego)*.

**Garcia-bellido, A. and Celis, J. F. De** (1992). Developmental genetics of the venation pattern: Origin of Wing Veins. *Annu. Rev. Genet.* **1940**, 28.

**Garcia-Bellido, A. and Santamaria, P.** (1972). Developmental analysis of the wing disc in the mutant engrailed of Drosophila melanogaster. *Genetics* **72**, 87–104.

**Gibert, J. M., Mouchel-Vielh, E., Quéinnec, E. and Deutsch, J. S.** (2000). Barnacle duplicate engrailed genes: Divergent expression patterns and evidence for a vestigial abdomen. *Evol. Dev.* **2**, 194–202.

**Gierer, A. and Meinhardt, H.** (1972). A theory of biological pattern formation. *Kybernetik* **12**, 30–39.

**Gömez-Skarmeta, J. L., Del Corral, R. D., De La Calle-Mustienes, E., Ferrés-Marcó, D. and Modolell, J.** (1996). Araucan and Caupolican, two members of the novel iroquois complex, encode homeoproteins that control proneural and vein-forming genes. *Cell* **85**, 95–105.

**Green, J. B. A. and Sharpe, J.** (2015). Positional information and reaction-diffusion: two big ideas in developmental biology combine. *Development* **142**, 1203–1211.

**Green, J. B. A. and Smith, J. C.** (1991). Growth factors as morphogens: do gradients and thresholds establish body plan? *Trends Genet.* **7**, 245–250.

**Grieder, N. C., Morata, G., Affolter, M. and Gehring, W. J.** (2009). Spalt major controls the development of the notum and of wing hinge primordia of the Drosophila melanogaster wing imaginal disc. *Dev. Biol.* **329**, 315–326.

**Grimm, S. and Pflugfelder, G. O.** (1996). Control of the gene optomotor-blind in Drosophila wing development by decapentaplegic and wingless. *Science* **271**, 1601–1604.

**Guichard, a, Biehs, B., Sturtevant, M. a, Wickline, L., Chacko, J., Howard, K. and Bier, E.** (1999). rhomboid and Star interact synergistically to promote EGFR/MAPK signaling during Drosophila wing vein development. *Development* **126**, 2663–2676.

**Guillén, I., Mullor, J. L., Capdevila, J., Sánchez-Herrero, E., Morata, G. and Guerrero, I.** (1995). The function of *engrailed* and the specification of *Drosophila* wing pattern. *Development* **121**, 3447–3456.

**Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O.** (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321.

**Gustavson, E., Goldsborough, A. S., Ali, Z. and Kornberg, T. B.** (1989). The Drosophila engrailed and invected Genes: Partners in Regulation, Expression and Function. **906**, 893–906.

**Han, K. and Manley, J. L.** (1993). Functional domains of the Drosophila Engrailed protein. *EMBO J.* **12**, 2723–2733.

**Harrison, L. G. and Tan, K. Y.** (1988). Where may Reaction-Diffusion Mechanisms be Operating in Metameric Patterning of Drosophila Embryos ? **8**, 118–124.

**Hepker, J., Blackman, R. K. and Holmgren, R.** (1999). Cubitus interruptus is necessary but not sufficient for direct activation of a wing-specific decapentaplegic enhancer. *Development* **126**, 3669–3677.

**Hoffmann, J., Donoughe, S., Li, K., Salcedo, M. K. and Rycroft, C. H.** (2018a). A simple developmental model recapitulates complex insect wing venation patterns. *Proc. Natl. Acad. Sci.* 201721248.

**Hoffmann, J., Donoughe, S., Li, K., Salcedo, M. K. and Rycroft, C. H.** (2018b). A simple developmental model recapitulates complex wing venation patterns in insects. *Proc. Natl. Acad. Sci.* **115**, 630–681.

**Huang, Y., Hatakeyama, M. and Shimmi, O.** (2018). Wing vein development in the sawfly Athalia rosae is regulated by spatial transcription of Dpp/BMP signaling components. *Arthropod Struct. Dev.* **47**, 408–415.

**Huerta-Cepas, J., Serra, F. and Bork, P.** (2016). ETE 3: Reconstruction, Analysis, and Visualization of Phylogenomic Data. *Mol. Biol. Evol.* **33**,

1635–1638.

**Hui, C. C., Matsuno, K., Ueno, K. and Suzuki, Y.** (1992). Molecular characterization and silk gland expression of Bombyx engrailed and invected genes. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 167–171.

**Huq, M., Bhardwaj, S. and Monteiro, A.** (2019). Male bicyclus anynana butterflies choose females on the basis of their ventral UV-reflective eyespot centers. *J. Insect Sci.* **19**,.

**Ingham, P. W. and Fietz, M. J.** (1995). Quantitative effects of hedgehog and decapentaplegic activity on the patterning of the Drosophila wing. *Curr. Biol.* **5**, 432–440.

**Jiggins, C. D., Wallbank, R. W. R. and Hanly, J. J.** (2017). Waiting in the wings: What can we learn about gene co-option from the diversification of butterfly wing patterns? *Philos. Trans. R. Soc. B Biol. Sci.* **372**,.

**Jung, H. S., Francis-West, P. H., Widelitz, R. B., Jiang, T. X., Ting-Berreth, S., Tickle, C., Wolpert, L. and Chuong, C. M.** (1998). Local inhibitory action of BMPs and their relationships with activators in feather formation: Implications for periodic patterning. *Dev. Biol.* **196**, 11–23.

**Kaba, D., Berté, D., Ta, B. T. D., Tellería, J., Solano, P. and Dujardin, J. P.** (2017). The wing venation patterns to identify single tsetse flies. *Infect. Genet. Evol.* **47**, 132–139.

**Kawahara, A. Y., Plotkin, D., Espeland, M., Meusemann, K., Toussaint, E. F. A., Donath, A., Gimnich, F., Frandsen, P. B., Zwick, A., dos Reis, M., et al.** (2019). Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 22657–22663.

**Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., et al.** (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649.

**Keys, D. N., Lewis, D. L., Selegue, J. E., Pearson, B. J., Goodrich, L. V., Johnson, R. L., Gates, J., Scott, M. P. and Carroll, S. B.** (1999). Recruitment of a hedgehog Regulatory Circuit in Butterfly Eyespot Evolution. *Science (80-. ).* **283**, 532–534.

**Koch, P. B. and Nijhout, H. F.** (2002). The role of wing veins in colour pattern development in the butterfly Papilio xuthus (Lepidoptera: Papilionidae). *Eur. J. Entomol.* **99**, 67–72.

**Koch, P. B., Merk, R., Reinhardt, R. and Weber, P.** (2003). Localization of ecdysone receptor protein during colour pattern formation in wings of the butterfly Precis coenia (Lepidoptera: Nymphalidae) and co-expression with Distal-less protein. *Dev. Genes Evol.* **212**, 571–584.

**Kosman, D., Small, S. and Reinitz, J.** (1998). Rapid preparation of a panel of

polyclonal antibodies to Drosophila segmentation proteins. *Dev. Genes Evol.* **208**, 290–294.

**Kukalova-Peck, J.** (1978). Origin and evolution of insecte wings and their relation to metamerphosis as documented by fossil record. *J. Morphol.* **156**, 53–125.

**Kunte, K., Zhang, W., Tenger-Trolander, A., Palmer, D. H., Martin, A., Reed, R. D., Mullen, S. P. and Kronforst, M. R.** (2014). doublesex is a mimicry supergene. *Nature* **507**, 229–232.

**Lawrence, P. A. and Morata, G.** (1976). Compartments in the wing of Drosophila: A study of the engrailed gene. *Dev. Biol.* **50**, 321–337.

**Lawrence, P. A. and Struhl, G.** (1996). Morphogens, compartments, and pattern: Lessons from Drosophila? *Cell* **85**, 951–961.

**Lawrence, P. A., Casal, J., Celis, J. F. de and Morata, G.** (2017). A refutation to "A new A-P compartment boundary and organizer in holometabolous insect wings." *Sci. Rep.* **7**, 16337.

**Lee, E., Madar, A., David, G., Garabedian, M. J., Gupta, R. Das and Logan, S. K.** (2013). Inhibition of androgen receptor and β-catenin activity in prostate cancer. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 15710–15715.

**Lewis, J. J., Geltman, R. C., Pollak, P. C., Rondem, K. E., Van Belleghem, S. M., Hubisz, M. J., Munn, P. R., Zhang, L., Benson, C., Mazo-Vargas, A., et al.** (2019). Parallel evolution of ancient, pleiotropic enhancers underlies butterfly wing pattern mimicry. *Proc. Natl. Acad. Sci. U. S. A.*

**Li, Z., Tiley, G., Galuska, S., Reardon, C., Kidder, T., Rundell, R. and Barker, M. S.** (2018). Multiple large-scale gene and genome duplications during the evolution of hexapods. *Proc. Natl. Acad. Sci.* **115**, 253609.

**Liao, D.** (1999). Concerted evolution: Molecular mechanism and biological implications. *Am. J. Hum. Genet.* **64**, 24–30.

**Livraghi, L., Hanly, J. J., Loh, L. S., Ren, A. and Warren, I. A.** (2020). The gene cortex controls scale colour identity in Heliconius. 1–30.

**Logan, C., Hanks, M. C., Noble-Topham, S., Nallainathan, D., Provart, N. J. and Joyner, A. L.** (1992). Cloning and sequence comparison of the mouse, human, and chicken engrailed genes reveal potential functional domains and regulatory regions. *Dev. Genet.* **13**, 345–358.

**Lunde, K., Biehs, B., Nauber, U. and Bier, E.** (1998). The knirps and knirps-related genes organize development of the second wing vein in Drosophila. *Development* **125**, 4145–4154.

**Lynch, M., O'Hely, M., Walsh, B. and Force, A.** (2001). The probability of preservation of a newly arisen gene duplicate. *Genetics* **159**, 1789–1804.

**Macias-Munoz, A., Smith, G., Monteiro, A. and Briscoe, A. D.** (2016). Transcriptome-Wide Differential Gene Expression in Bicyclus anynana

Butterflies: Female Vision-Related Genes Are More Plastic. *Mol. Biol. Evol.* **33**, 79–92.

**Manzanares, M., Marco, R. and Garesse, R.** (1993). Genomic organization and developmental pattern of expression of the engrailed gene from the brine shrimp Artemia. *Development* **118**, 1209–1219.

**Marie, B. and Bacon, J. P.** (2000). Two engrailed-related genes in the cockroach: cloning, phylogenetic analysis, expression and isolation of splice variants. *Dev. Genes Evol.* **210**, 436–48.

**Marie, B. and Blagburn, J. M.** (2003). Differential roles of engrailed paralogs in determining sensory axon guidance and synaptic target recognition. *J. Neurosci.* **23**, 7854–7862.

**Mariella, E., Marotta, F., Grassi, E., Gilotto, S. and Provero, P.** (2019). The length of the expressed 3′ UTR is an intermediate molecular phenotype linking genetic variants to complex diseases. *Front. Genet.* **10**, 1–20.

**Martin, A. and Reed, R. D.** (2010). wingless and aristaless2 define a developmental ground plan for moth and butterfly wing pattern evolution. *Mol. Biol. Evol.* **27**, 2864–2878.

**Martin, A. and Reed, R. D.** (2014). Wnt signaling underlies evolution and development of the butterfly wing pattern symmetry systems. *Dev. Biol.* **395**, 367–378.

**Martin, A., McCulloch, K. J., Patel, N. H., Briscoe, A. D., Gilbert, L. E. and Reed, R. D.** (2014). Multiple recent co-options of Optix associated with novel traits in adaptive butterfly wing radiations. *Evodevo* **5**, 1–13.

**Martín, M., Ostalé, C. M. and 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. *Dev.* **144**, 3168–3176.

**Matsuda, S. and Affolter, M.** (2017). Dpp from the anterior stripe of cells is crucial for the growth of the Drosophila wing disc. *Elife* **6**, 1–9.

**Matsuda, S., Yoshiyama, N., Künnapuu-Vulli, J., Hatakeyama, M. and Shimmi, O.** (2013). Dpp/BMP transport mechanism is required for wing venation in the sawfly Athalia rosae. *Insect Biochem. Mol. Biol.* **43**, 466–473.

**Matsuoka, Y. and Monteiro, A.** (2019). Hox genes are essential for the development of novel serial homologous eyespots on the wings of Bicyclus anynana butterflies. *bioRxiv* 814848.

**Mayr, C. and Bartel, D. P.** (2009). Widespread Shortening of 3′UTRs by Alternative Cleavage and Polyadenylation Activates Oncogenes in Cancer Cells. *Cell* **138**, 673–684.

**Meindhardt, H.** (2012). *The Algorithmic Beauty of Sea Shells*.

**Menshykau, D., Kraemer, C. and Iber, D.** (2012). Branch mode selection

during early lung development. *PLoS Comput. Biol.* **8**,.

**Merzendorfer, H. and Zimoch, L.** (2003). Chitin metabolism in insects: Structure, function and regulation of chitin synthases and chitinases. *J. Exp. Biol.* **206**, 4393–4412.

**Misof, B. and et al.** (2014). Phylogenomics resolves the timing and pattern of insect evolution. *Science (80-. ).* **346**, 763–767.

**Mohler, J., Seecoomar, M., Agarwal, S., Bier, E. and Hsai, J.** (2000). Activation of knot (kn) specifies the 3–4 intervein region in the Drosophila wing. *Development* **127**, 55–63.

**Monteiro, A.** (2015). Origin, Development, and Evolution of Butterfly Eyespots. *Annu. Rev. Entomol.* **60**, 253–271.

**Monteiro, A. and Gupta, M. D.** (2016). *Identifying Coopted Networks and Causative Mutations in the Origin of Novel Complex Traits*. 1st ed. Elsevier Inc.

**Monteiro, A. and Podlaha, O.** (2009). Wings, horns, and butterfly eyespots:How do complex traits evolve? *PLoS Biol.* **7**, 0209–0216.

**Monteiro, A. and Prudic, K. L.** (2010). Multiple approaches to study color pattern evolution in butterflies. *Trends Evol. Biol.* **2**, 7–15.

**Monteiro, A., Glaser, G., Stockslager, S., Glansdorp, N. and Ramos, D.** (2006). Comparative insights into questions of lepidopteran wing pattern homology. *BMC Dev. Biol.* **6**, 52.

**Monteiro, A., Chen, B., Ramos, D. M., Oliver, J. C., Tong, X., Guo, M., Wang, W. K., Fazzino, L. and Kamal, F.** (2013). Distal-Less Regulates Eyespot Patterns and Melanization in Bicyclus Butterflies. *J. Exp. Zool. Part B Mol. Dev. Evol.* **320**, 321–331.

**Monteiro, A., Tong, X., Bear, A., Liew, S. F., Bhardwaj, S., Wasik, B. R., Dinwiddie, A., Bastianelli, C., Cheong, W. F., Wenk, M. R., et al.** (2015). Differential Expression of Ecdysone Receptor Leads to Variation in Phenotypic Plasticity across Serial Homologs. *PLoS Genet.* **11**, 1–20.

**Müller, H. A. J., Samanta, R. and Wieschaus, E.** (1999). Wingless signaling in the Drosophila embryo: Zygotic requirements and the role of the frizzled genes. *Development* **126**, 577–586.

**Müller, P., Rogers, K. W., Jordan, B. M., Lee, J. S., Robson, D., Ramanathan, S. and Schier, A. F.** (2012). Differential diffusivity of nodal and lefty underlies a reaction-diffusion patterning system. *Science (80-. ).* **336**, 721–724.

**Nederbragt, A. J., Loon, E. Van and Dictus, W. J. A. G.** (2002). Expression of Patella vulgata Orthologs of engrailed and dpp-BMP2 / 4 in Adjacent Domains during Molluscan Shell Development Suggests a Conserved Compartment Boundary Mechanism. **355**, 341–355.

**Nel, A., Roques, P., Nel, P., Prokop, J. and Steyer, J. S.** (2007). The earliest holometabolous insect from the Carboniferous: A “crucial” innovation

with delayed success (Insecta Protomeropina Protomeropidae). *Ann. la Soc. Entomol. Fr.* **43**, 349–355.

**Nijhout, H. F.** (1978). Wing pattern formation in Lepidoptera: A model. *J. Exp. Zool.* **206**, 119–136.

**Nijhout, H. F.** (1980). Pattern formation on lepidopteran wings: Determination of an eyespot. *Dev. Biol.* **80**, 267–274.

**Nowell, R. W., Elsworth, B., Oostra, V., Zwaan, B. J., Wheat, C. W., Saastamoinen, M., Saccheri, I. J., van't Hof, A. E., Wasik, B. R., Connahs, H., et al.** (2017). A high-coverage draft genome of the mycalesine butterfly Bicyclus anynana. *Gigascience* **6**, 1–7.

**Oliver, J. C., Tong, X. L., Gall, L. F., Piel, W. H. and Monteiro, A.** (2012). A Single Origin for Nymphalid Butterfly Eyespots Followed by Widespread Loss of Associated Gene Expression. *PLoS Genet.* **8**,.

**Oliver, J. C., Beaulieu, J. M., Gall, L. F., Piel, W. H. and Monteiro, A.** (2014). Nymphalid eyespot serial homologues originate as a few individualized modules. *Proc. R. Soc. B Biol. Sci.* **281**, 20133262–20133262.

**Olson, E. N.** (2006). Gene regulatory networks in the evolution and development of the heart. *Science* **313**, 1922–7.

**Organista, M. F. and De Celis, J. F.** (2013). The Spalt transcription factors regulate cell proliferation, survival and epithelial integrity downstream of the Decapentaplegic signalling pathway. *Biol. Open* **2**, 37–48.

**Özsu, N. and Monteiro, A.** (2017). Wound healing, calcium signaling, and other novel pathways are associated with the formation of butterfly eyespots. *BMC Genomics* **18**, 1–14.

**Özsu, N., Chan, Q. Y., Chen, B., Gupta, M. Das and Monteiro, A.** (2017). Wingless is a positive regulator of eyespot color patterns in Bicyclus anynana butterflies. *Dev. Biol.* **429**, 177–185.

**Patel, N. H., Martin-Blanco, E., Coleman, K. G., Poole, S. J., Ellis, M. C., Kornberg, T. B. and Goodman, C. S.** (1989). Expression of engrailed proteins in arthropods, annelids, and chordates. *Cell* **58**, 955–968.

**Patil, S. and Magdum, S.** (2017). Insight into wing venation in butterflies belonging to families Papilionidae , Nymphalidae and Pieridae from Dang Dist Gujarat , India. *J. Entomol. Zool. Stud.* **5**, 1596–1607.

**Peel, A. D., Telford, M. J. and Akam, M.** (2006). The evolution of hexapod engrailed-family genes: evidence for conservation and concerted evolution. *Proc. R. Soc. B Biol. Sci.* **273**, 1733–1742.

**Peifer, M., Orsulic, S., Sweeton, D. and Wieschaus, E.** (1993). A role for the Drosophila segment polarity gene armadillo in cell adhesion and cytoskeletal integrity during oogenesis. *Development* **118**, 1191–1207.

**Peltenburg, L. T. and Murre, C.** (1996). Engrailed and Hox homeodomain proteins contain a related Pbx interaction motif that recognizes a common

structure present in Pbx. *EMBO J.* **15**, 3385–3393.

**Peterson, M. D., Popadić, A. and Kaufman, T. C.** (1998). The expression of two engrailed-related genes in an apterygote insect and a phylogenetic analysis of insect engrailed-related genes. *Dev. Genes Evol.* **208**, 547–557.

**Phillips, R. and Whittle, J.** (1993). wingless expression mediates determination of peripheral nervous system elements in late stages of Drosophila wing disc development. *Development*.

**Prakash, A. and Monteiro, A.** (2016). Molecular mechanisms of secondary sexual trait development in insects. *Curr. Opin. Insect Sci.* **17**, 40–48.

**Prakash, A. and Monteiro, A.** (2018). *apterous A* specifies dorsal wing patterns and sexual traits in butterflies. *Proc. R. Soc. B Biol. Sci.* **285**, 20172685.

**Prakash, A. and Monteiro, A.** (2019). Doublesex mediates the development of sex - specific pheromone organs in Bicyclus butterflies via multiple mechanisms. *bioRxiv*.

**Price, M. N., Dehal, P. S. and Arkin, A. P.** (2009). Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650.

**Prokop, J. and Ren, D.** (2007). New significant fossil insects from the Upper Carboniferous of Ningxia in northern China (Palaeodictyoptera, Archaeorthoptera). *Eur. J. Entomol.* **104**, 267–275.

**Prudic, K. L., Stoehr, A. M., Wasik, B. R. and Monteiro, A.** (2015). Eyespots deflect predator attack increasing fitness and promoting the evolution of phenotypic plasticity. *Proc. R. Soc. London B Biol. Sci.* **282**, 20141531.

**Quinlan, M. C. and Gibbs, A. G.** (2006). Discontinuous gas exchange in insects. *Respir. Physiol. Neurobiol.* **154**, 18–29.

**Quiring, R., Walldorf, U., Kloter, U. and Gehring, W. J.** (1994). Homology of the eyeless gene of drosophila to the small eye gene in mice and aniridia in humans. *Science (80-. ).* **265**, 785–789.

**Raftery, L. A., Sanicola, M., Blackman, R. K. and Gelbart, W. M.** (1991). The relationship of decapentaplegic and engrailed expression in Drosophila imaginal disks: do these genes mark the anterior-posterior compartment boundary? *Development* **113**, 27–33.

**Ralston, A. and Blair, S. S.** (2005). Long-range Dpp signaling is regulated to restrict BMP signaling to a crossvein competent zone. *Dev. Biol.* **280**, 187–200.

**Raspopovic, J., Marcon, L., Russo, L. and Sharpe, J.** (2014). (Suppl)Digit patterning is controlled by a Bmp-Sox9-Wnt Turing network modulated by morphogen gradients. *Science (80-. ).* **345**, 566–570.

**Rauser, C. L. and Rutowski, R. L.** (2003). Male-specific structures on the

wings of the gulf fritillary butterfly, Agraulis vanillae (Nymphalidae). *J. Lepid. Soc.* **57**, 279–283.

**Ray, R. P. and Wharton, K. A.** (2001). Context-dependent relationships between the BMPs gbb and dpp during development of the Drosophila wing imaginal disk. *Development* **128**, 3913–3925.

**Reed, R. D. and Gilbert, L. E.** (2004). Wing venation and Distal-less expression in Heliconius butterfly wing pattern development. 628–634.

**Reed, R. D. and Serfas, M. S.** (2004). Butterfly Wing Pattern Evolution Is Associated with Changes in a Notch/Distal-less Temporal Pattern Formation Process. *Curr. Biol.* **14**, 1159–1166.

**Reed, R. D., Chen, P. H. and Frederik Nijhout, H.** (2007). Cryptic variation in butterfly eyespot development: The importance of sample size in gene expression studies. *Evol. Dev.* **9**, 2–9.

**Reed, R. D., Papa, R., Martin, A., Hines, H. M., Kronforst, M. R., Chen, R., Halder, G., Nijhout, H. F. and Mcmillan, W. O.** (2011). optix Drives the Repeated Convergent Evolution of Butterfly Wing Pattern Mimicry. *Science (80-. ).* **333**, 1137–1141.

**Robertson, K. a and Monteiro, A.** (2005). Female Bicyclus anynana butterflies choose males on the basis of their dorsal UV-reflective eyespot pupils. *Proc. Biol. Sci.* **272**, 1541–1546.

**Roch, F., Baonza, F., Martín-Blanco, E., García-Bellido, A., Baonza, A., Martin-Blanco, E. and Garcia-Bellido, A.** (1998). Genetic interactions and cell behaviour in blistered mutants during proliferation and differentiationof the Drosophila wing. *Development* **125**, 1823–1832.

**Sabarís, G., Laiker, I., Preger-Ben Noon, E. and Frankel, N.** (2019). Actors with Multiple Roles: Pleiotropic Enhancers and the Paradigm of Enhancer Modularity. *Trends Genet.* **35**, 423–433.

**Saenko, S. V, French, V., Brakefield, P. M. and Beldade, P.** (2008). Conserved developmental processes and the formation of evolutionary novelties: examples from butterfly wings. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **363**, 1549–1555.

**Saenko, S. V., Marialva, M. S. P. P. and Beldade, P.** (2011). Involvement of the conserved Hox gene Antennapedia in the development and evolution of a novel trait. *Evodevo* **2**, 9.

**Sanderson, A. R., Kirby, R. M., Johnson, C. R. and Yang, L.** (2006). Advanced Reaction-Diffusion Models for Texture Synthesis. *J. Graph. Tools* **11**, 47–71.

**Sanson, B., Alexandre, C., Fascetti, N. and Vincent, J. P.** (1999). Engrailed and hedgehog make the range of wingless asymmetric in Drosophila embryos. *Cell* **98**, 207–216.

**Schachat, S. R. and Brown, R. L.** (2015). Color pattern on the forewing of Micropterix (Lepidoptera: Micropterigidae): Insights into the evolution of wing pattern and wing venation in moths. *PLoS One* **10**, 1–16.

**Schnepp, B., Grumbling, G., Donaldson, T. and Simcox, A.** (1996). Vein is a novel component in the *Drosophila* epidermal growth factor receptor pathway with similarity to the neuregulins. *Genes Dev.* **10**, 2302–2313.

**Schwartz, C., Locke, J., Nishida, C. and Kornberg, T.** (1995). Analysis of cubitus interruptus regulation in Drosophila embryos and imaginal disks. *Development* **121**, 1625–1635.

**Schweitzer, R., Howes, R., Smith, R., Shilo, B. Z. and Freeman, M.** (1995). Inhibition of drosophila EGF receptor activation by the secreted protein argos. *Nature* **376**, 699–702.

**Seimiya, M. and Gehring, W. J.** (2000). The Drosophila homeobox gene optix is capable of inducing ectopic eyes by an eyeless-independent mechanism. *Development* **127**, 1879–1886.

**Sekimura, T. and Nijhout, H. F.** (2017). *Diversity and Evolution of Butterfly Wing Patterns*. Springer nature.

**Shimmi, O., Ralston, A., Blair, S. S. and O'Connor, M. B.** (2005). The crossveinless gene encodes a new member of the Twisted gastrulation family of BMP-binding proteins which, with Short gastrulation, promotes BMP signaling in the crossveins of the Drosophila wing. *Dev. Biol.* **282**, 70–83.

**Shimmi, O., Matsuda, S. and Hatakeyama, M.** (2014). Insights into the molecular mechanisms underlying diversified wing venation among insects. *Proc. Biol. Sci.* **281**, 20140264.

**Sick, S., Reinker, S., Timmer, J. and Schlake, T.** (2006). WNT and DKK Determine Hair Follicle Spacing Through a Reaction-Diffusion Mechanism. **314**, 1447–1451.

**Simcox, A. A., Grumbling, G., Schnepp, B., Bennington-Mathias, C., Hersperger, E. and Shearn, A.** (1996). Molecular, phenotypic, and expression analysis of vein, a gene required for growth of the Drosophila wing disc. *Dev. Biol.* **177**, 475–489.

**Stamatakis, A.** (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313.

**Stark, J., Bonacum, J., Remsen, J. and DeSalle, R.** (1999). THE EVOLUTION AND DEVELOPMENT OF DIPTERAN WING VEINS: A Systematic Approach. *Annu. Rev. Entomol.* **44**, 97–129.

**Stoehr, A. M., Walker, J. F. and Monteiro, A.** (2013). Spalt expression and the development of melanic color patterns in pierid butterflies. *Evodevo* **4**, 6.

**Strigini, M. and Cohen, S. M.** (1999a). Formation of morphogen gradients in the Drosophila wing. *Semin Cell Dev Biol* **10**, 335–344.

**Strigini, M. and Cohen, S. M.** (1999b). Formation of morphogen gradients in the Drosophila wing. *Semin Cell Dev Biol* **10**, 335–344.

**Sturtevant, M. a, Biehs, B., Marin, E. and Bier, E.** (1997). The spalt gene

links the A/P compartment boundary to a linear adult structure in the Drosophila wing. *Development* **124**, 21–32.

**Sun, P., Mhatre, N., Mason, A. C. and Yack, J. E.** (2018). In that vein : inflated wing veins contribute to butterfly hearing Author for correspondence : **2017**,.

**Szuperak, M., Salah, S., Meyer, E. J., Nagarajan, U., Ikmi, A. and Gibson, M. C.** (2011). Feedback regulation of Drosophila BMP signaling by the novel extracellular protein Larval Translucida. *Development* **138**, 715–724.

**Tabata, T., Eaton, S. and Kornberg, T. B.** (1992). The Drosophila hedgehog gene is expressed specifically in posterior compartment cells ana is a target of engrailed regulation. *Genes Dev.* **6**, 2635–2645.

**Tanimoto, H., Itoh, S., Ten Dijke, P. and Tabata, T.** (2000). Hedgehog creates a gradient of DPP activity in Drosophila wing imaginal discs. *Mol. Cell* **5**, 59–71.

**Tewary, M., Ostblom, J., Prochazka, L., Zulueta-Coarasa, T., Shakiba, N., Fernandez-Gonzalez, R. and Zandstra, P. W.** (2017). A stepwise model of reaction-diffusion and positional information governs self-organized human peri-gastrulation-like patterning. *Dev.* **144**, 4298–4312.

**Thayer, R. C., Allen, F. I. and Patel, N. H.** (2020). Structural color in Junonia butterflies evolves by tuning scale lamina thickness. *Elife* **9:e52187**, 1–21.

**Theißen, G.** (2001). Development of floral organ identity: Stories from the MADS house. *Curr. Opin. Plant Biol.* **4**, 75–85.

**Thompson, J. D., Higgins, D. G. and Gibson, T. J.** (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.

**Tolkunova, E. N., Fujioka, M., Kobayashi, M., Deka, D. and Jaynes, J. B.** (1998). Two Distinct Types of Repression Domain in Engrailed: One Interacts with the Groucho Corepressor and Is Preferentially Active on Integrated Target Genes. *Mol. Cell. Biol.* **18**, 2804–2814.

**Tomoyasu, Y., Wheeler, S. R. and Denell, R. E.** (2005). Ultrabithorax is required for membranous wing identity in the beetle Tribolium castaneum. *Nature* **433**, 643–647.

**Turing, A. M.** (1952). The chemical basis of morphogenesis. *Philos. Trans. ofthe R. Soc. ofLondon. Ser. B, Biol. Sci.* **237**, 37–72.

**Van de Wetering, M., Cavallo, R., Dooijes, D., Van Beest, M., Van Es, J., Loureiro, J., Ypma, A., Hursh, D., Jones, T., Bejsovec, A., et al.** (1997). Armadillo coactivates transcription driven by the product of the Drosophila segment polarity gene dTCF. *Cell* **88**, 789–799.

**Waddington, C. H.** (1940). The genetic control of wing development in Drosophila. *J. Genet.* **41**, 75–113.

**Wagner, G. P. and Lynch, V. J.** (2010). Hox cluster duplications and the opportunity for evolutionary novelties. *Curr. Biol.* **20**, 14603–14606.

**Wang, D., Li, J., Liu, S., Zhou, H., Zhang, L., Shi, W. and Shen, J.** (2017). spalt is functionally conserved in Locusta and Drosophila to promote wing growth. *Nat. Publ. Gr.* 1–9.

**Weatherbee, S. D., Nijhout, H. F., Grunert, L. W., Halder, G., Galant, R., Selegue, J. and Carroll, S.** (1999). Uitrabithorax function in butterfly wings and the evolution of insect wing patterns. *Curr. Biol.* **9**, 109–115.

**Westerman, E. L., VanKuren, N. W., Massardo, D., Tenger-Trolander, A., Zhang, W., Hill, R. I., Perry, M., Bayala, E., Barr, K., Chamberlain, N., et al.** (2018). Aristaless Controls Butterfly Wing Color Variation Used in Mimicry and Mate Choice. *Curr. Biol.* **28**, 3469-3474.e4.

**Whitington, P. M., Meier, T. and King, P.** (1991). Segmentation, neurogenesis and formation of early axonal pathways in the centipede, Ethmostigmus rubripes (Brandt). *Roux's Arch. Dev. Biol.* **199**, 349–363.

**Wolpert, L.** (1969). Positional information and the spatial pattern of cellular differentiation. *J. Theor. Biol.* **25**, 1–47.

**Wolpert, L.** (1971). Positional information and pattern formation. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **295**, 441–50.

**Woronik, A., Tunström, K., Perry, M. W., Neethiraj, R., Stefanescu, C., Celorio-Mancera, M. de la P., Brattström, O., Käkelä, R., Hill, J., Lehmann, P., et al.** (2018a). A homeobox gene, BarH-1, underlies a female alternative life-history strategy. *bioRxiv*.

**Woronik, A., Stefanescu, C., Käkelä, R., Wheat, C. W. and Lehmann, P.** (2018b). Physiological differences between female limited, alternative life history strategies: The Alba phenotype in the butterfly Colias croceus. *J. Insect Physiol.* **107**, 257–264.

**Yang, L., Meng, F., Ma, D., Xie, W. and Fang, M.** (2013). Bridging Decapentaplegic and Wingless signaling in Drosophila wings through repression of naked cuticle by Brinker. *Development* **140**, 413–422.

**Yokoyama, S. and Yokoyama, R.** (1989). Molecular evolution of human visual pigment genes. *Mol. Biol. Evol.* **6**, 186–197.

**Yoshimoto, E. and Kondo, S.** (2012). Wing vein patterns of the Hemiptera insect Orosanga japonicus differ among individuals. *Interface Focus* **2**, 451–456.

**Yu, P. B., Hong, C. C., Sachidanandan, C., Babitt, J. L., Deng, D. Y., Hoyng, S. A., Lin, H. Y., Bloch, K. D. and Peterson, R. T.** (2008). Dorsomorphin inhibits BMP signals required for embryogenesis and iron metabolism. *Nat. Chem. Biol.* **4**, 33–41.

**Zattara, E. E., Busey, H. A., Linz, D. M., Tomoyasu, Y. and Moczek, A. P.** (2016). Neofunctionalization of embryonic head patterning genes facilitates the positioning of novel traits on the dorsal head of adult

beetles. *Proc. R. Soc. B Biol. Sci.* **283**,.

**Zecca, M., Basler, K. and Struhl, G.** (1995). Sequential organizing activities of engrailed, hedgehog and decapentaplegic in the Drosophila wing. *Development*.

**Zecca, M., Basler, K. and Struhl, G.** (1996). Direct and long range action of a DPP morphogen gradient. *Cell* **87**, 833–844.

**Zhang, J.** (2003). Evolution by gene duplication: An update. *Trends Ecol. Evol.* **18**, 292–298.

**Zhang, J., Rosenberg, H. F. and Nei, M.** (1998). Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 3708–3713.

**Zhang, S., Zhang, J. S., Zhao, J. and He, C.** (2015). Distinct subfunctionalization and neofunctionalization of the B-class MADS-box genes in Physalis floridana. *Planta* **241**, 387–402.

**Zhang, L., Martin, A., Perry, M. W., van der Burg, K. R. L., Matsuoka, Y., Monteiro, A. and Reed, R. D.** (2017a). Genetic Basis of Melanin Pigmentation in Butter fl y Wings. *Genetics* **205**, 1537–1550.

**Zhang, L., Mazo-Vargas, A. and Reed, R. D.** (2017b). Single master regulatory gene coordinates the evolution and development of butterfly color and iridescence. *Proc. Natl. Acad. Sci.* **114**, 10707–10712.

# APPENDIX

**Table A.1. *In-situ* hybridization Buffers**

| Buffers | Chemicals | Amount |
|---|---|---|
| 10X PBS (500 ml)<br>* Sterilize by autoclaving. | $K_2HPO_4$ | 5.34 g |
| | $KH_2PO_4$ | 2.64 g |
| | NaCl | 40.9 g |
| | DEPC treated $H_2O$ | To 500 ml |
| 1X PBST (50 ml) | 1X PBS | 50 ml |
| | Tween® 20 | 50 µl |
| 20X SSC (1000 ml)<br>*Adjust the pH to 7.0 with 1M HCl and sterilize by autoclaving. | NaCl | 175.3 g |
| | Trisodium citrate | 88.2 g |
| | DEPC treated $H_2O$ | Till 1000 ml |
| Pre-hybridization buffer (40 ml) | Formamide | 20 ml |
| | 20X SSC | 10 ml |
| | DEPC treated water | 10 ml |
| | TWEEN20 | 40 µl |
| Hybridization buffer (40 ml) | Formamide | 20 ml |
| | 20X SSC | 10 ml |
| | DEPC treated water | 10 ml |
| | TWEEN20 | 40 µl |
| | Salmon sperm | 40 µl |
| | Glycine (100mg/ml) | 40 µl |

| | | |
|---|---|---|
| Block buffer (50 ml) | 1X PBS | 50 ml |
| | TWEEN20 | 50 µl |
| | BSA | 0.1 gm |
| Alkaline phosphatase buffer (20 ml) | Tris-HCl (pH 8.0) | 2 ml |
| | NaCl (5M) | 400 µl |
| | $MgCl_2$ (200mM) | 250 µl |
| | DEPC treated water | Till 20 ml |
| | TWEEN20 | 20 µl |

**Table A.2. Immunohistochemistry Buffers**

| Buffers | Chemicals | Amount |
|---|---|---|
| Fix buffer (30 ml) | M PIPES pH 6.9 (500 mM) | 6 ml |
| | mM EGTA pH 6.9 (500mM) | 60 µl |
| | % Triton x-100 (20 %) | 1.5 ml |
| | mM $MgSO_4$ (1M) | 60 µl |
| | 37% Formaldehyde | 55 µl per 500 µl of buffer |
| | $dH_2O$ | 22.4 ml |
| Block buffer (40 ml) | 50 mM Tris pH 6.8 (1 M) | 2 ml |
| | 150 mM NaCl (5 M) | 1.2 ml |
| | 0.5% IGEPAL (NP40) (20%) | 1 ml |
| | 5 mg/ml BSA | 0.2 gr |
| | H2O | 35.8 ml |
| Wash buffer (200 ml) | 50mM Tris pH 6.8 (1 M) | 10 ml |
| | 150 mM NaCl (5 M) | 6 ml |
| | 0.5% IGEPAL (20 %) | 5 ml |

| | | |
|---|---|---|
| | 1 mg/ml BSA | 0.2 gr |
| | $dH_2O$ | 179 ml |
| Mounting media | Tris-HCl (pH 8.0) | 20 mM |
| | N-propyl gallate | 0.5% |
| | Glycerol | 60% |

**Publications**


Banerjee, T.D.; Monteiro, A. Molecular mechanisms underlying simplification of venation patterns in holometabolous insects. Development dev.196394 (2020).

Banerjee, T.D.; Ramos, D.; A. Monteiro, Expression of multiple engrailed family genes in eyespots of Bicyclus anynana butterflies does not implicate the duplication events in the evolution of this morphological novelty. *Front. Ecol. Evol*. 8, 1–12 (2020).

Banerjee, T.D.; Monteiro, A. Dissection of Larval and Pupal Wings of *Bicyclus anynana* Butterflies. *Methods Protoc*. 3, 5 (2020).

Banerjee, T.D.; Monteiro, A. Molecular mechanism underlying venation patterning in butterflies. *bioRxiv* 1–23 (2020).

Connahs H.; Tlili S.; van Creij J.; T. Y. J. Loo; Banerjee T.D.; Saunders T. E.; Monteiro, A. Activation of butterfly eyespots by Distal-less is consistent with a reaction-diffusion process. *Development*. **146**, dev169367 (2019).

Banerjee, T.D.; Monteiro, A. CRISPR-Cas9 Mediated Genome Editing in *Bicyclus anynana* Butterflies. *Methods Protoc.* (2018) *1*, 16.