Title
Making it clear: evolution, development and genetic basis of wing transparency in Lepidoptera

Author
Pomerantz, Aaron

Publication Date
2021

Peer reviewed|Thesis/dissertation

Making it clear: evolution, development and genetic basis of wing transparency in Lepidoptera

By

Aaron Ford Pomerantz

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Integrative Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Nipam H. Patel, Co-Chair
Professor Peter Sudmant, Co-Chair
Professor Craig T. Miller
Professor Noah K. Whiteman

Summer 2021

Abstract

Making it clear: evolution, development and genetic basis of wing transparency in Lepidoptera

by

Aaron Ford Pomerantz

Doctor of Philosophy in Integrative Biology

University of California, Berkeley

Professors Nipam H. Patel & Peter Sudmant, Co-Chairs

Pause for a moment, close your eyes, and picture a few of the most beautiful living organisms that come to your mind... What did you see? Perhaps the creatures were bright and shiny. Perhaps they were colorful and charismatic. Perhaps one of the beautiful creatures you pictured was a butterfly. Indeed, the diversity of colorful patterns in butterflies have captivated humans for centuries. Moreover, their wings have influenced studies in a variety of scientific fields, including evolutionary biology, ecology, and biophysics.

Lepidopteran wings are covered with thousands of flat overlapping scales, each one of which derives from a single cell. The scales on an adult are cuticular projections that serve as the unit of color for the wing. Each scale can generate color through pigmentation, which results from molecules that selectively absorb certain wavelengths, or due to light interacting with physical nanoarchitecture on the scales, known as structural color. Thus far, researchers have made progress in understanding genetic pathways responsible for pigment production and the early transcription factors and signaling molecules that demarcate wing pattern positions. It remains less clear, however, what precise genes and pathways give rise to an individual wing scale cell, how such a novel cell type evolved, or what factors modulate cuticular micro- and nanostructures that generate specific optical properties.

To better understand processes underlying scale and nanostructure development in Lepidoptera, my dissertation focuses on a unique optical strategy: wing transparency. The wings of butterflies and moths are typically covered with thousands of flat, overlapping scales that endow the wings with colorful patterns. Yet, numerous species of Lepidoptera have evolved highly transparent wings, which often possess scales of altered morphology and reduced size, and the presence of membrane surface nanostructures that dramatically reduce reflection. This trait has been interpreted as an adaptation in the context of camouflage, in which numerous lineages independently evolved transparent wings as a form of crypsis to reduce predation.

In order to unravel the biological processes of wing transparency, I engaged in an interdisciplinary collaboration (including the labs of Nipam Patel, Marianne Elias, Doris Gomez and Serge Berthier) at the interface of physics, developmental biology and evolutionary ecology. Working in parallel with our collaborators, we aimed to investigate the structure, development and evolution of wing transparency in butterflies and moths by implementing experimental and phylogenetic comparative methods. We revealed a diversity of structural features that underlie transparent wings, notably modifications of scale morphology, size, and density, and the presence of finely-tuned nanostructures on the surface of the wing membrane that generate anti-reflective properties. We were able to characterize developmental processes of wing micro- and nanostructure formation of glasswing butterflies that were raised in the field and in the lab, and additionally utilized museum specimens and data to identify correlations between light transmission (a quantitative measure of transparency) and structural features. Together, our results provide insight into the development, ecology and evolutionary history of terrestrial transparency within Lepidoptera, highlighting multiple lineages that have independently evolved clearwing phenotypes, as well as potential trade-offs related to thermoregulation, water repellency and predation pressure.

One of my main experimental systems became the so-called 'glasswing butterfly' *Greta oto*, which has thin, vertically oriented scales and nanopillars coating the wing membrane that enable omnidirectional anti-reflective properties. My collaborators and I employed a multitude of techniques, including confocal and electron microscopy, GC-MS, optical spectroscopy and analytical simulations to characterize wing development, comparing transparent and non-transparent wing regions. We found that during early wing development, scale precursor cell density was reduced in transparent regions, and cytoskeletal organization during scale growth differed between thin, bristle-like scale morphologies within transparent regions and flat, round scale morphologies within opaque regions. We also show that nanostructures on the wing membrane surface are composed of two layers: a lower layer of regularly arranged nipple-like nanostructures, and an upper layer of irregularly arranged wax-based nanopillars composed predominantly of long-chain n-alkanes. By chemically removing wax-based nanopillars, along with optical spectroscopy and analytical simulations, we demonstrate their role in generating anti-reflective properties. These findings provide insight into morphogenesis and composition of naturally organized microstructures and nanostructures, and may provide bioinspiration for new anti-reflective materials.

Additionally, I undertook a comparative transcriptomic analysis to identify molecular pathways involved in scale cell development in the buckeye butterfly *Junonia coenia* and the giant silkmoth *Antheraea polyphemus*. I also investigated differential expression between two regions within the wing of *A. polyphemus*: a region we refer to as a transparent 'window' in which scale cells do not develop, and an adjacent region that undergoes canonical scale development. I then applied fluorescent *in situ* hybridization and CRISPR/Cas9 induced knockouts to characterize the spatiotemporal expression and function of genes involved in scale cell development. Comparative RNA-seq between *J. coenia* and *A. polyphemus* uncovered genes with similar expression levels during early pupal wing development and scale precursor differentiation, including proneural, cell cycle, and *Notch* signaling factors. At later pupal stages, when scale cell projections are forming and maturing, I identified genes with

similar expression levels related to cytoskeletal organization, melanization, cuticle formation, and chitin-synthesis. Using stage-specific transcriptomic analysis followed by *in situ* hybridization, I uncover a suite of genes that likely play conserved roles in scale cell patterning and morphogenesis in butterflies and moths. I identified two *achaete scute* homologs (*ASH1*, *ASH2*) expressed at the scale cell precursor stage and loss of function of *ASH2* resulted in the loss of scale cells. In contrast, loss of function of the *Notch* receptor led to overproduction and dense clusters of scale cells, likely due to improper lateral inhibition during scale precursor cell differentiation. I also identified that the 'window' scaleless region in *A. polyphemus* is associated with high expression levels of Wnt ligands, including *wingless*, and the bHLH transcription factor *hairy*, a negative regulator of sensory bristles, revealing how putative co-option of neurogenesis regulatory factors could contribute to scale cell patterning in Lepidoptera.

Finally, I lay out a new and easy-to-follow protocol for portable, rapid, field-deployable amplicon sequencing through the use of new miniaturized lab equipment, which can be beneficial for biodiversity exploration and educational programs. Human-mediated environmental change is depleting biodiversity faster than it can be characterized, while invasive species cause agricultural damage, threaten human health, and disrupt native habitats. Consequently, the application of effective approaches for rapid surveillance and identification of biological samples is increasingly important to inform conservation efforts. Taxonomic assignments have been greatly advanced using sequence-based applications, such as DNA barcoding, a diagnostic technique that utilizes polymerase chain reaction (PCR) and DNA sequence analysis of standardized genetic regions. However, in many biodiversity hotspots, endeavors are often hindered by a lack of genomic infrastructure and funding for biodiversity research and restrictions on the transport of biological samples. A promising development is the advent of low-cost, miniaturized scientific equipment. Such tools can be assembled into functional laboratories to carry out genetic analyses *in situ*, at local institutions, field stations, or classrooms.

Dedication

To my mother, Deborah Ford. I love and miss you every day.

Table of Contents

List of Figures

Acknowledgements

I am grateful to many people and institutions for their support, mentorship, encouragement, and training. This includes many talented undergraduate students who volunteered their time and creativity on projects, including Yuriko Kishi, Raymundo Picos, Johnny On, Cody Limber, Cassidy Quilalang, Madelyn Frank, Kia Torab, Shrey Saretha and Katie Sanko. My fellow Patel and Sudmant lab mates, including Ryan Null, Erin Jarvis, Heather Bruce, Dennis Sun, Rachel Thayer, Jaap van Krugten, Damien Gailly, Kyle DeMarr, Jennifer McCarthy, Sophia Kelly, Gabrielle Jerz, Rohit Kolora, Manny Vazquez, Isabel Serrano, Francisca Catalan, Stacy Li, and Alma Halgren, were (and will remain) wonderful and supportive colleagues and friends. I would like to give special thanks to Manu Prakash for connecting me with Arnaud Martin on Twitter, who led me to the Patel lab and taught me everything I know about genome-editing in Lepidoptera. I am extremely grateful to Melanie Mcclure, Luca Livraghi, Oscar Paneso, Rémi Mauxion, Owen McMillan and Fred Gagnon of Magic Wings Butterfly Conservatory and Gardens for their assistance with rearing and obtaining glasswing butterflies for research. I thank the wonderful communities in the Department of Integrative Biology and the Marine Biological Laboratory, Hitomi Asahara and the UC Berkeley DNA Sequencing Facility for technical guidance with genomic experiments, Neil Tsutsui for assistance with GC-MS experiments, Rosemary Gillespie for her unwavering support and encouragement with science communication efforts, and Mark Jenkinson for helping to transform hands-on educational genomics programs from an idea into reality. I am grateful to Tropical Herping, Rainforest Expeditions, Field Projects International and the Inkaterra Guides Field Station for their support to carry out portable DNA sequencing experiments in the tropics. I want to thank Josh Cassidy and KQED's Deep Look for helping to produce beautiful and educational digital media on our work. I am indebted to the members of my thesis committee, Nipam Patel, Peter Sudmant, Craig Miller, and Noah Whiteman, who not only helped guide me into the fascinating world of genomics and developmental biology but were also always available to provide their mentorship, stimulating intellectual conversations and encouragement that helped me through the highs and lows. Last but not least, I am grateful for the love and support of my incredible family and friends - I could not have done this without you. Thank you all.

Introduction to the work

In this dissertation I will present the findings of my doctoral research on the evolution, development, and genetic basis of wing transparency in butterflies and moths. Additionally, I provide a new protocol for long-range DNA amplicon sequencing using miniaturized laboratory equipment. What follows are four main chapters, beginning with an introduction summarizing the current state of the field and ending with a look forward to the future. Below is a brief summary of each chapter:

Chapter 1: I present the reader with a summary of why transparency is a fascinating, yet poorly studied, optical property in living organisms and how we have taken an interdisciplinary approach to address scientific questions using Lepidoptera as our study system. We show that within Lepidoptera, transparency has evolved multiple times independently and with these multiple gains comes a large diversity of morphological innovation.

Chapter 2: In the second chapter, I elucidate aspects of the developmental, cellular and biochemical basis of transparency. Recent studies have explored aspects of structural diversity, optical properties and phylogenetic distribution of transparency within butterflies and moths, which often possess scales of altered morphology and reduced size, and the presence of membrane surface nanostructures that dramatically reduce reflection. However, the developmental processes underlying transparency are currently unknown, let alone the dynamic formation of butterfly nanostructures in general. This presents a gap in our understanding of how wing scale microstructure morphologies and membrane surface nanostructures are generated within a living system. I therefore set out to explore the development of wing transparency in the so-called 'glasswing butterfly' *Greta oto*, which has thin, vertically oriented scales and irregularly arranged nanopillars on the wing membrane that enable omnidirectional anti-reflective properties. I demonstrated that this species could be easily reared in the lab, which allowed us to employ a multitude of techniques, including confocal and electron microscopy, GC-MS, optical spectroscopy and analytical simulations. I was then able to create a series of wing development stages for the first time, comparing transparent and non-transparent wing regions.

Chapter 3: In the third chapter, I explore the genetic basis of scale development and a form of wing transparency that occurs via suppression of scale development. I carried out stage-specific RNA sequencing of micro-dissected wing tissue in the buckeye butterfly *Junonia coenia* and the giant silkmoth *Antheraea polyphemus* to identify conserved differentially expressed genes. Additionally, I investigated differential expression between two regions within the wing of *A. polyphemus*: a region we refer to as a 'window' in which scale cells do not develop, and an adjacent region that undergoes canonical scale development. I then applied fluorescent in situ hybridization and CRISPR/Cas9 induced mutagenesis to characterize the spatiotemporal expression and function of several genes involved in wing scale development. I also identified that the 'window' scaleless region in *A. polyphemus* is associated with high expression

levels of Wnt ligands, including *wingless*, and the bHLH transcription factor *hairy*, a negative regulator of sensory bristles, revealing how co-option of neurogenesis regulatory factors could contribute to scale cell patterning in lepidoptera.

Chapter 4: Finally, in chapter four I lay out a protocol for DNA amplicon sequencing using miniaturized laboratory equipment for genetic biomonitoring and biodiversity exploration. I believe that recent advancements in miniaturized scientific tools (such as handheld thermocyclers and DNA sequencers) enable promising opportunities to "bring the lab into the field" and offer increased accessibility to equipment, which can promote local capacity building. Such advancements offer new and exciting opportunities to investigate and understand living organisms in the context of their natural environment. This is especially important today, as human-mediated environmental change is depleting biodiversity faster than it can be characterized, while invasive species cause agricultural damage, threaten human health, and disrupt native habitats. Consequently, the application of effective approaches for rapid surveillance and identification of biological samples is increasingly important to inform conservation efforts. Taxonomic assignments have been greatly advanced using sequence-based applications, such as DNA barcoding, a diagnostic technique that utilizes polymerase chain reaction (PCR) and DNA sequence analysis of standardized genetic regions. However, in many biodiversity hotspots, endeavors are often hindered by a lack of genomic infrastructure and funding for biodiversity research and restrictions on the transport of biological samples. A promising development is the advent of low-cost, miniaturized scientific equipment. Such tools can be assembled into functional laboratories to carry out genetic analyses *in situ*, at local institutions, field stations, or classrooms. Here, I outline all the steps required to perform amplicon sequencing applications outside of a conventional laboratory environment using miniaturized scientific equipment.

Chapter 1: Introduction and background

Why study wing transparency in Lepidoptera?

Lepidoptera (butterflies and moths) are a widespread and diverse order of insects, with over 180,000 described species. As the name of the order suggests, members of this group are characterized in part by wings that are covered with scales, after the ancient greek *lepís* = scale and *pterón* = wing. Adult scales are cuticular projections that derive from single cells during pupal development and serve as the unit of color for the wing. Each scale can generate color through pigmentation, which results from molecules that selectively absorb certain wavelengths of light, structural coloration, which results from light interacting with the physical nano-architecture of the scale, or a combination of both (Wasik et al., 2014; Zhang et al., 2017; Thayer et al., 2020). During pupal wing development, scale cell extensions grow rapidly, creating finely-tuned structures ~200 µm long and ~60 µm wide. Chitin synthesis and cytoskeletal dynamics, including highly organized F-actin filaments, during scale cell development, play essential roles in wing scale elongation and aspects of scale ultrastructure (Dinwiddie et al., 2014; Day et al., 2019; Liu et al., 2021). Once fully developed and matured, scales and most of the adult wing itself remain non-living tissues. The colorful patterns produced by wing scales can serve numerous functions, including antipredator strategies (e.g. camouflage, deflection, aposematism, mimicry), attracting mates, thermoregulation, and water repellency (Nijhout, 2001; Perez Goodwyn et al., 2009; Mazo-Vargas et al., 2017; Deshmukh et al., 2018; Tsai et al., 2020).

Yet, numerous species of butterflies and moths possess *transparent* wings that allow light to pass through, so that objects behind them can be distinctly seen (Fig 1.1, Yoshida et al., 1997; Siddique et al., 2015). This has typically been interpreted as an adaptation in the context of camouflage, in which some lineages evolved transparent wings as a form of crypsis to reduce predation (Arias et al., 2019; McClure et al., 2019). Indeed, one can easily imagine how an 'invisibility cloak' would serve as an effective camouflage strategy for an organism to blend in with any background and remain 'hidden in plain sight' (Cuthill, 2019). Controlled field studies using butterflies with varying degrees of wing transparency and wild birds have additionally supported the hypothesis that transparency is involved in camouflage and decreases detectability by predators (Arias et al., 2019, 2020). Perhaps counter-intuitively, wing transparency is widespread among many groups of aposematic Lepidoptera, whereby chemically-defended species display conspicuous wing patterns that advertise unpalatability to predators. This phenomenon did not go unnoticed by Henry Walter Bates over 150 years ago, who observed the striking similarities between distantly related clearwing butterflies and moths collected in nearby regions of the Brazilian Amazon, contributing to his theory of mimicry (Fig 1.2, Bates, 1862). Many different chemically-defended clearwing species locally converge on similar wing patterns and form mimicry 'rings', thus begging the question: why adapt to be both transparent and aposematic?

Despite the natural history and physical properties described in a handful of transparent species, the evolutionary history and potential mechanisms underlying transparency in terrestrial organisms has been lacking. In order to achieve efficient transparency, the wing scales must be modified in such a way that they are no longer

covering the wing membrane, and some species additionally form nanostructures on their wings that generate anti-glare properties. As such, clearwing Lepidoptera make up an excellent model group for studying terrestrial transparency, as it raises puzzling physical, physiological, developmental and evolutionary questions. For instance we can ask, what are the structural solutions for being transparent? Are there trade-offs between transparency and other important functions that are normally fulfilled by scales, such as hydrophobicity? What are the developmental and genetic pathways underlying wing transparency? How has transparency evolved in Lepidoptera? Understanding such questions around transparency requires working at the interface between physics, evolutionary biology and developmental biology. Thus, my lab (Patel) formed an intercontinental collaborative effort (with labs including Elias, Gomez and Berthier) that aimed to elucidate the adaptive functions of transparency in clearwing Lepidoptera and the generative processes leading to modified scales and nanostructures in transparent wings, bridging the gap between ontogeny, function and evolution.

Prevalence of transparent organisms in aquatic and terrestrial environments

When light interacts with the surface of an organism, different levels of reflected light will be produced at the interface between the air and the substrate of the organism depending on the refractive index, a dimensionless number that describes how fast light travels through the material (Johnsen and Widder, 1999). When light moves between materials with different refractive indices, such as a mix of cellular proteins or lipids in a living organism, the light scatters rather than passing through unaffected (Benedek, 1971; Chapman, 1976). Transparency is the transmission of light across the visible spectrum without significant absorption or scattering so that the material appears completely see-through, and a number of conditions must be fulfilled if incident light is to be transmitted freely through a living organism.

Until recently, research had primarily been devoted to the occurrence of transparency in water, which is much more frequent than on land. Transparency in water can serve as protection against visually hunting predators through concealment, in particular in open ocean habitats where there may be nowhere to hide or blend in against the surrounding environment (Chapman, 1976; Johnsen, 2001; Tuthill and Johnsen, 2006). A survey by Johnsen (2001) revealed a broad phylogenetic distribution of transparent organisms, including members belonging to Arthropoda, Mollusca, Annelida, Cnidaria, and Chordata. For instance many fish (e.g. the ghost catfish *Kryptopterus vitreolus*), crustaceans (e.g. hyperiid amphipods *Phronima* sp, anemone shrimp *Periclimenes yucatanicus*), and cephalopods (e.g. the glass octopus *Vitreledonella* sp.) have highly transparent bodies.

By contrast, transparent organisms are nearly absent on land. This disparity is likely explained by physical and biological factors. The ability to achieve efficient transparency requires low absorption and reflection, as well as low scattering of light, and these constraints are often difficult to fulfill for organisms in terrestrial environments. The contrast between refractive indices is ultimately dictated by the material's composition (e.g. the exoskeleton of arthropods composed of chitin) and that of the medium surrounding the material, such as air for butterflies, and water for aquatic species. For instance, let us consider the refractive indices of air (n ~ 1.0), water (n ~

1.33) and biological tissue (n ~1.56 for insect chitin). In this case, we can see there is a larger difference between refractive indices of air and the biological tissue (1.0 versus 1.56), relative to water and the tissue (1.33 versus 1.56), resulting in higher reflection of the incident light in the terrestrial environment, which compromises camouflage by transparency (Meyer, 1979; Johnsen, 2001). Challenges faced by organisms are also different in air than water, such as the need for self-supporting anatomical structures (e.g. bones and protective epidermis typically composed of opaque biological materials) and for protection against higher levels of ultraviolet (UV) radiation (e.g. light-absorbing pigments such as melanin) (Chapman, 1976).

While the majority of studies on transparent organisms to date have focused on marine organisms, a handful of transparency-inducing properties have been investigated in terrestrial arthropods. For instance, early studies using electron microscopy identified that the outer surface of many insect compound eyes were covered in three dimensional cone-shaped protuberances that were in the range of ~50-200 nanometers in height and ~200 nm distance between nanostructures (Bernhard and Miller, 1962; Bernhard et al., 1970). The first type of nanostructures observed were termed ‘corneal nipples’, which were found to be arranged in a more or less perfect hexagonal array and were shown to create an interface with a gradient refractive index between air and the corneal lens material, thereby significantly reducing the amount of light reflection (Bernhard and Miller, 1962; Miller, 1979; Stavenga et al., 2006). Corneal nipple nanostructures are presumed to function in reducing eye glare of the insect, so that they are less visible to predators, or could additionally function to increase the amount of light that reaches the photoreceptive region of the eye in nocturnal insects to improve vision (Miller, 1979). The latter function highlights a potential use for technology, as synthetic insect-inspired anti-reflective nanostructure coatings are being investigated for applications to improve solar cells by reducing reflection and improving the absorption of light within, which could boost optoelectronic conversion efficiency (Cai and Qi, 2015). A more recent comparative study applied atomic force microscopy, providing nanometer and subnanometer resolution, for a survey of 23 insect orders and several non-insect arthropods (Blagodatski et al., 2015). This study identified a high diversity of nanostructure morphologies and patterns, including nipple-like structures, maze-like nanocoatings, parallel strands/ridges formed by fusion of nipple-type protrusions, dimple-type nanocoating and various transitions between these major forms, such as nipples-to-maze transition, maze-to-strands transition, nipples-to-strands transition and dimples-to-maze transition.

In addition to insect cornea, nanostructures have also been characterized on the wing membrane of several insect species. For example, the transparent-winged hawkmoth *Cephonodes hylas* was found to contain nanostructures coating the wing membrane that morphologically resembled the aforementioned ‘corneal nipple arrays’, and were shown to reduce surface reflection on the wings (Yoshida et al., 1997). Perhaps the most notable example of a transparent butterfly is *Greta oto*, commonly known as the ‘glasswing’ butterfly (Nymphalidae: Ithomiinae). The thin bristle-like and forked scales in the transparent regions of the wing expose nanoprotuberances on the surface of the membrane (Binetti et al., 2009). The structures were charactered as randomly sized ‘nanopillars’ and were found to enable broadband and omnidirectional anti-reflection properties (Fig 1.3, Siddique et al., 2015). This is in contrast to the typical

wing surface of butterflies and moths, which contain densely packed, colorful scale outgrowths and smooth chitinous wing membrane that results in high surface reflection. More recently, another transparent-winged butterfly in the genus *Chorinea* (Riodinidae: Riodininae) was found to contain small, widely spaced scales and dome-shaped chitin nanostructures on the membrane that result in anti-reflective and angle-independent transmission (Narasimhan et al., 2018). Such anti-reflective nanostructures found in nature are proving to be quite rewarding for applications in biomimetics and biophotonics, such as solar cells, anti-glare glasses and optical implant devices (Narasimhan et al., 2018; Han et al., 2019). However, despite efforts to produce small-scale artificial nanofabrications, the developmental and genetic mechanisms involved in the formation of anti-reflective nanostructures in a living system are virtually unknown.

Taken together, we can see how transparency can be a powerful evolutionary innovation, but it is likely uncommon due to physical and biological constraints. As a result, transparency is rare for terrestrial organisms due to challenging optical requirements, but it is frequently found in aquatic life, in which organisms are better able to match the refractive index of their surrounding environment. Perhaps unsurprisingly however, *life finds a way*. Many organisms have evolved morphological innovations that overcome the challenges of terrestrial transparency, notably in the form of anti-reflective nanostructures, such as those located on the corneal surface of insect eyes and wings.

Phylogenetic distribution of transparency in Lepidoptera

Over the past several years, members of the Patel Lab have conducted preliminary analyses of clearwing Lepidoptera by investigating specimens within private and public collections, which revealed numerous strategies to achieve clearwing phenotypes. For example, the giant glasswing butterfly *Methona confusa* (Nymphalidae: Ithomiini) was observed to contain modified scales that resembled thin bristle-like morphologies and a smooth wing membrane surface that lacked nanostructures (Fig 1.2, Fig 1.3). As a result, the wing is somewhat transparent, but retains a high degree of reflectivity. Interestingly, numerous distantly related species of butterflies and moths closely resemble the wing patterns of *Methona*, implying convergent evolution, but upon closer inspections we noticed that the scale morphologies were dissimilar between each species (Fig 1.2). This small sample of mimics highlights that there is more than one way to make a wing clear, as scale morphologies can be modified in independent ways to generate a clearwing phenotype.

Searching the literature also revealed a handful of studies on transparent lepidopteran species. For instance, the longtail glasswing, *Chorinea faunus* (Riodinidae), was found to contain small, widely spaced scales and dome-shaped chitin nanoprotuberances on the membrane that generate anti-reflective properties (Narasimhan et al., 2018). As previously mentioned, the hawkmoth, *Cephonodes hylas* (Sphingidae) has nude wings due to its scales falling out upon eclosion, and was found to possess anti-reflective nanostructures on its wing surface that morphologically resemble insect corneal nipple arrays (Yoshida et al., 1997). Nipple array nanostructures have also been characterized in transparent wing regions of the tiger moth *Cacostatia ossa* (Erebidae) (Deparis et al., 2014). Finally, the glasswing butterfly *Greta oto* (Nymphalidae: Ithomiini) was found to contain thin, vertically oriented scales,

allowing the wing surface to be exposed, along with nanopillars that coat the surface. In contrast to the previously described nipple arrays, the structures in *G. oto* were characterized as irregularly arranged 'nanopillars' that featured a random height and width distribution, and were found to enable omnidirectional anti-reflection properties (Fig 1.3, Binetti et al., 2009; Siddique et al., 2015).

While this sporadic sampling of clearwing butterflies and moths implied independent evolutionary events, open questions remained with regard to the phylogenetic extent of transparency, the diversity and evolution of underlying structures, the existence, if any, of structural constraints on transparency, and the ecological relevance of transparency on land. As such, our lab and collaborators (Patel, Elias, Berthier, and Gomez) surveyed the Order Lepidoptera for the presence of clearwing species on the basis of: our own experience, on previous published literature, on knowledge of museum curators and researchers, on species names, and Lepidoptera collections in various natural history museums (including the Essig Museum of Entomology, French Museum of Natural History, McGuire Center for Lepidoptera & Biodiversity). In total, we identified species with at least partially transparent wings in 31 out of the 124 existing Lepidoptera families and gathered a total of 123 species (Gomez et al., 2021). This distribution of clearwing species represents approximately one quarter of the extant butterfly families, the remainder of which are primarily composed of opaque species. Therefore, we can infer that transparency has evolved multiple times independently and may present evolutionary benefits (McClure et al., 2019; Gomez et al., 2020; Pinna et al., 2020). With these multiple evolutionary gains comes a wide diversity of structural strategies, including highly modified scale morphologies and intricate membrane-surface nanostructures.

Diversity of scale morphologies in clearwing Lepidoptera

Utilizing our diverse dataset of transparent lepidopteran species, we explored aspects of wing scale microstructure diversity (for instance the wing scale type, dimensions, and density) by taking images of disarticulated wings using high magnification digital and scanning electron microscopy. We found that there are a variety of ways to achieve transparency through modifications of scale cell organization and morphology. For example, species exhibiting transparent wings may contain a low density of scales (i.e. the number of scale cells are reduced and spaced farther apart relative to scale cells in an opaque butterfly wing), have a reduction in the overall size of the scales (i.e. scales are modified to be shorter, forked or bristle-like, thereby exposing the wing membrane), the scales themselves can transparent, the scales can be vertically-oriented, or a combination of the above (Fig 1.4).

A common structural strategy involves flat scales with either decreased scale cell densities relative to opaque wing regions, or the flat scales are transparent and packed in high density, together representing 49 species from 22 families (Gomez et al., 2021). Examples of flat, low density scales include species such as the danaid *Parantica sita* and the papilionid *Parnassius glacialis* (Fig 1.4) (Perez Goodwyn et al., 2009). Gomez et al. (2021) also documented species that contained transparent scales including the hesperid *Oxynetra semihyalina* and the crambid *Diaphania unionalis.* Such 'glass scales' have been previously recorded on the dorsal wing surface of a colorful species,

the papilionid *Graphium sarpedon*, which were reported to act as thin films and also function as polarizing reflectors (Stavenga et al., 2012). Scales can also be erected such that they are protruding from the wing at a non-flat angle, with examples including the riodinid *Chorinea faunus*, the papilionid *Lamproptera meges*, the nymphalids *Parantica sita* and *Methona confusa* and the pierid *Patia orise*. (Fig 1.4).

Another structural strategy involves thin bristle-like, or 'piliform' scales, which can refer to a scale that morphologically resembles the macrocheate bristle on the notum of *Drosophila melanogaster*. Examples include the forked and bristle-like scales present in numerous 'glasswing' butterflies belonging to the Ithomiini tribe (Fig 1.4 Binetti et al., 2009; Siddique et al., 2015). Forked scales can also be found in members belonging to the genus *Acraea* (Fig 1.4). Species with only bristle-like scales in their transparent wing regions have appeared multiple times and are phylogenetically widespread, including members belonging to the Haeterini such as *Cithaerias* sp., saturniids such as *Rothschildia lebeau,* or clearwing moths belonging to the Arctiidae (Fig 1.4, Hernández-Chavarría et al., 2004; Berthier, 2007; Pomerantz et al., 2021). Such repeated cases of bristle-like scales in Lepidoptera are potentially interesting, as scales are homologous to insect sensory bristles (Galant et al., 1998). As such, bristles may represent the ancestral state of a scale, and these clearwing butterflies may have reverted wing scales back to their ancestral morphology.

Nude wing membrane was another structural strategy that was identified in clearwings, such as the erebid moth *Senecauxia coraliae* and the psychiid *Chalioides ferevitrea* (Gomez et al., 2021). Nude wing membrane was previously recorded in the sphingid *Cephonodes hylas* which contains spread out sockets, which are the remnants of fully developed scales that fall out shortly after eclosion (Fig 1.4) (Yoshida et al., 1997). We observed small regions of nude wing membrane in members belonging to the saturniidae, such as the giant silkmoth *Antheraea polyphemus*, in which neither sockets nor scale cells form around the crossvein (Fig 1.4). It remains unclear what potential benefits exist for a large moth to contain small transparent wing spots but previous hypotheses have been put forward, such as serving as 'false holes' so predators visually mistake the moth as a rotten leaf (discussed in Hernández-Chavarría et al., 2004).

This wide variation and broad dispersion among transparent-winged phenotypes indicates that the different scale projection types arose as result of independent evolution. Until recently, comparative studies of clearwing Lepidoptera were lacking, but our surveys highlight a myriad of morphological innovations and combinations of scale modifications (such as scale type, insertion, and density) that now expand the range of strategies reported in the literature (Gomez et al., 2021). We find additional support for the independent evolution of scale projection types from investigating a group of *Methona* (Ithomiini) mimics. Adult ithomiids sequester toxic compounds such as pyrrolizidine alkaloids (PA's) which makes them distasteful to predators such as spiders and birds (Brown, 1984). Bates (1862) suggested that palatable species evolved to resemble unpalatable ithomiids, in order to share in their survival advantage. While the *Methona* butterflies and distantly related mimics we investigated were remarkably similar in their macro wing pattern (with characteristic black stripes and clear wing regions), the scale morphologies themselves were quite different, ranging from thin bristle-like erected scales to flat translucent scales, suggesting that the scale

morphologies evolved in an independent manner and subsequently converged similarity in wing pattern phenotype (Fig 1.2).

Taken together, these results suggest that there is not one common pathway to make a wing transparent. However, it does seem that there are three main strategies and/or combinations that contribute to making a wing clear from a presumably opaque ancestral state. As a first step, a lepidopteran can begin to achieve wing transparency through a reduction in the density of wing scales, and in some cases leave only the socket cell behind or abolish scale precursor cell development altogether. The scale dimensions can also be made as narrow as possible and erected vertically, enabling more light to reach and pass through the wing membrane, while potentially still retaining other benefits of the scale, such as water repellency. Finally, the scale projections themselves can be made translucent and lack pigmentation. While these strategies are necessary to get the wing scales 'out of the way' so to speak, the chitinous material of the wing membrane itself remains highly reflective. As such, numerous species of clearwings have evolved wing surface nanostructures, which generate considerable anti-reflective optical properties.

Diversity of wing surface nanostructures in clearwing Lepidoptera

In addition to scale modifications, the presence of nanoprotuberances on the surface of the wing membrane can enhance transparency through the reduction of light reflection, by generating a gradient of refractive index between the chitin-made membrane and the air (Yoshida et al., 1997; Binetti et al., 2009; Siddique et al., 2015; Pomerantz et al., 2021). Using scanning electron microscopy at high levels of magnification, we investigated the wing membrane surface of our clearwing specimens in order to determine if they harbored surface nanostructures (Fig. 1.5, 1.6, 1.7).

We identified wing surface nanostructures that were previously documented, such as nipple arrays in the hawkmoth *Hemaris thysbe*, as well as numerous previously undescribed nanostructures in a wide diversity of butterflies and moths (Fig 1.5). For instance, we identified irregularly arranged nanopillar-like structures within members of the Haeterini, including *Cithaerias merolina*, *Dulcedo polita* and *Haetera pierra*. These structures appear to morphologically resemble the anti-reflective wax-based nanopillars previously characterized in ithomiines, such as *Greta oto* and *Godyris duilia*, suggesting that this class of nanostructure has arisen at least twice independently in nymphalid butterflies (Fig 1.5, Fig 1.6) (Pomerantz et al., 2021). The two-dimensional fast Fourier transformation (FFT) based on SEM images of the nanostructures also reveals a ring-shaped distribution of the squared Fourier components, a characteristic from the irregular arrangement of nanopillars (Fig 1.6, as reported in Siddique et al., 2015). Within nymphalids, we also observed irregular maze-like nanostructures in species such as *Parantica sita*, and similar maze-like nanocoatings have previously been documented covering the corneae insect eyes, such as Gyrinidae beetles (Fig 1.5, 1.6) (Blagodatski et al., 2015). In addition to ordered nipple arrays that have been previously characterized in the sphingid *Hemaris* and the riodinid *Chorinae*, we identified similar highly ordered to quasi-ordered dome-shaped nanostructures in the sessid moth *Synanthedon kathyae*, the arctiid wasp mimic *Cosmosoma myroadora*, the erebid *Cocytia durvillii* and the papilionid butterfly *Lamproptera meges* (Fig 1.5, 1.6, 1.7). The

Fourier power distribution based on SEM of the nipple-like nanostructures shows that the arrays are highly organized at intermediate spatial frequencies and equivalently periodic in all directions (Fig 1.6, 1.7). Such ordered nipple-like nanostructures appear to be chitin-based and have evolved numerous times in phylogenetically distant butterflies and moths, suggesting that this class of nanostructure is a common route to achieve transparency (Pomerantz et al., 2021). We also observed dome-shaped nanostructures that were spaced-out relative to the highly ordered nipple array class of structures. These spaced-out domes were present in species that mimic ithomiine butterflies, such as the pierid *Dismorphia fortunata* and the notodontid *Dioptis* sp. (Fig 1.5, 1.6, 1.7). Finally, we observed several cases where a clearwing species had modified scales that exposed the wing membrane, but did not harbor any type of surface nanostructure and consequently had glossier, more reflective wings. Clearwing species that lacked wing surface nanostructures included members of the *Acraea*, *Parnassius*, *Antheraea*, as well as *Methona* and species that mimic *Methona*, including members belonging to *Lycorea*, *Parides*, and *Notophyson* (Fig 1.6, 1.7).

In a recent study, (Pinna et al., 2020) conducted a comparative study of optics and structures in transparent wings of neotropical mimetic clearwing Lepidoptera. They identified an unexpected diversity of nanostructures that cover the wing membrane and report five types of nanostructures: absent, maze, nipple arrays, pillars, and sponge-like. They reveal that nanostructure density largely influences light transmission and while transparency was largely influenced by modification of scale microstructure features, nanostructures provide a means to additionally fine-tune the degree of transparency.

Taken together, we report a wide variety of wing surface nanostructures in clearwing butterflies and moths, just as there are a wide variety of corneal surface nanostructures previously documented on the surface of arthropod cornea (Blagodatski et al., 2015). Turing reaction–diffusion mechanisms have been proposed as a model for the formation of different corneal nanostructure morphologies (such as spacing, height, and spatial organization) during insect eye development (Blagodatski et al., 2015; Kryuchkov et al., 2017). While the formation of such biological structures have been mathematically modelled, the developmental processes underlying transparency were virtually unknown, let alone the dynamic formation of butterfly epithelial nanostructures in general (see Chapter 2).

Ecological relevance of transparency: vision, thermoregulation and hydrophobicity

Intuitively, one might perceive transparency in Lepidoptera as a means of ‘being invisible’ to go undetected by visually hunting predators and the trait therefore provides a selective advantage. However, wing scales represent a multifunctional evolutionary innovation that can function in camouflage, mimicry, attracting mates, thermoregulation, and water repellency (Nijhout, 2001; Perez Goodwyn et al., 2009; Mazo-Vargas et al., 2017; Deshmukh et al., 2018; Tsai et al., 2020). Because clearwings harbor modified and/or reduced wing scales, there is an open question as to whether there are trade-offs or constraints to evolving wing transparency.

Scales in Lepidoptera are known to play a role in hydrophobicity based on wing scale morphology, arrangement, and insertion angle. One study investigated whether or not hydrophobicity was affected based on scale morphology for two butterfly species with clear wings, *Parantica sita* (Nymphalidae: Danaeinae) and *Parnassius glacialis* (Papilionidae: Parnassiinae). *Parantica sita* has thin ovoid, erected scales and was found to have super-hydrophobic wing surfaces, whereas *Parnassius glacialis* has bristle and spade-like scales and was found to have wings that were less hydrophobic than most opaque Lepidoptera. (Perez Goodwyn et al., 2009). Another study reported low hydrophobicity values in the clearwing species *Greta oto*, which has a high proportion of transparent wing area (Wanasekara and Chalivendra, 2011). Transparency may therefore come at a cost, especially for Lepidoptera endemic to tropical environments that receive a high amount of rainfall. Perhaps this is why we find relatively few cases in which clearwings have abolished scale cell development altogether, but instead retain modified scales that coat the wing in order to retain water-proofing properties. The structural determinants of hydrophobic properties in clearwing Lepidoptera, such as surface nanostructure and roughness, remains unknown and potential trade-offs between optics and hydrophobicity are yet to be determined.

Species from colder environments, such as higher latitudes or altitudes, can gain thermal benefits from being more strongly pigmented, as radiation absorption aids in heating up (Bogert, 1949). Pigmented wing scales in Lepidoptera can help to fulfill thermoregulatory functions in opaque species and can play a thermostatic role by emitting or absorbing infrared radiations relative to temperature. For example, a large-scale comparative analysis in opaque butterflies showed that body and proximal wing coloration correlates to climate in the near-infrared (700-1100 nm) range but not below 700 nm, where both thermoregulation and vision spectrums coincide (Munro et al., 2019). As such, we could expect wavelength absorption to be higher at high latitudes compared to low latitudes, in particular in the infrared range, and that transparency could entail thermal costs. In a recent study, Gomez et al. (2021) tested the ecological relevance of transparency for thermoregulation and found that mean transmittance in lepidopteran wings in the near infrared range and in the human-visible range decreased with increasing latitude. On the other hand, mean transmittance in the ultraviolet (UV) range did not strongly vary with latitude (Gomez et al., 2021). Overall, results with transparency in Lepidoptera and their links with ecology (e.g. latitude, activity rhythm) are in line with trends identified in opaque Lepidoptera and suggest that transparency may be more complex than just enhancing concealment, as it is likely a multifunctional compromise (Gomez et al., 2021).

Wing pattern development in glasswing butterflies

Recent studies have explored aspects of structural diversity, optical properties and phylogenetic distribution of transparency within Lepidoptera. However, such studies involved the use of dead adult specimens deposited in museums, which represents the 'finished' form the wing and associated structures, whereas the developmental processes underlying transparency remained unknown.

From 2016-2017, we undertook expeditions to Tarapoto, Peru, where our collaborators were rearing clearwing butterflies at a local field site. During these trips,

we performed pupal wing dissections of several species belonging to the Ithomiini, including *Methona*, *Ithomia* and *Oleria*. The aim of these experiments was to gather preliminary data and characterize scale development, which involved wing tissue fixation and staining with compounds to mark cellular components such as actin, cell membrane and nuclei (see Chapter 2). Near the end of the trip, I performed pharmacological treatments on a subset of *Oleria onega*, which involved injecting doses of heparin directly into pupae. Heparin is a highly sulfated form of heparan sulfate glycosaminoglycan, which is known to modulate wing patterns in Lepidoptera. Heparin is thought to bind to Wnt ligands (among other signaling molecules) and facilitate their transport, inducing what appears to be Wnt gain-of-function effects in butterfly wing patterns, in particular with extension of basal, discal, and external wing symmetry systems (Serfas and Carroll, 2005).

In wild-type individuals, the dorsal surface of *O. onega* forewings have clear regions and two black striped patterns that appear to be associated with the Nymphalid central symmetry system, as well as black scales around the wing margin (Fig 1.8). The central region of the hindwing is primarily clear, while the margin contains black scales. The ventral wing surface mirrors the same pattern as the dorsal surface, with the addition of a stripe of orange scales and white spots near the distal margin. In heparin treated individuals, it was apparent that wing patterns had been highly modified (Fig 1.8). One individual contained mostly black wings with severely reduced transparent zones. The shift in patterns corresponded with a shift in scale morphology, resulting in fewer bristle-like and forked scales that are normally present in the clear wing regions. On the ventral side of the wing, the orange marginal band had decreased in size and several white spot patterns had merged. Another heparin-treated individual could hardly be called a glasswing butterfly anymore, as both the forewing and hindwing appeared completely black and covered in opaque scales (Fig 1.8). These results are in line with previous studies that treated opaque butterfly species with heparin, noting that such pharmacological treatments potentially affected Wnt signaling ligands and shifted wing pattern boundaries. As such, we can hypothesize that Wnts, such as WntA are possible candidate wing patterning factors in glasswing butterflies (Nymphalidae: Ithomiini) that demarcate the boundaries between opaque and transparent zones. It would be interesting for future studies to investigate these signaling ligands in wing discs and probe how the regulation of wing patterning factors may have been modified to enable wing transparency in different lineages of Lepidoptera.

Conclusion and future directions for transparent Lepidoptera

Transparency is a fascinating, yet poorly studied, optical property in living organisms. We have shown that among the Lepidoptera, transparency has evolved multiple times independently in an insect order characterized by wing opacity, and with these multiple gains comes a large diversity of morphological innovation (Pinna et al., 2020; Gomez et al., 2021; Pomerantz et al., 2021). Optical transparency is determined by macrostructure (such as the proportion of clearwing area), scale microstructure (such as scale morphology and density) as well as wing surface nanostructures (such as nipple arrays and nanopillars). Microstructural traits are tightly linked in their evolution, perhaps resulting in differential investment of chitin and pigment synthesis based on the

structural strategy. The links between transparency and latitude are also consistent with thermal benefits, and less with UV protection (Gomez et al., 2021). Comparative analyses in clearwing Lepidoptera also mirror findings from opaque Lepidoptera, demonstrating that transparency is more complex than just enhancing concealment and is likely a multifunctional compromise.

Research on transparent aquatic organisms has left many questions unanswered with regard to the structural basis of transparency, its genetic and developmental pathways, its functional roles and potential trade-offs with optics, and the selective pressures driving its evolution and form (Johnsen, 2014; Bagge, 2019). The same can be said about studies of transparent species on land, which have historically consisted of physicists applying imaging and modeling approaches on museum specimens with a bioinspiration goal of replicating anti-reflective, hydrophobic or anti-fouling properties (e.g. Narasimhan et al., 2018; Han et al., 2019). Future studies should utilize an evolutionary framework and comparative approaches with transparent organisms to expand optical measurements (taking into consideration both micro- and nano-structures within each individual), explore developmental and molecular pathways, and experimentally test adaptive functions and potential trade-offs with other vital functions such as water repellency.

The fact that many transparent Lepidoptera are chemically protected and display warning patterns in their wings (aposematism) and form sympatric mimicry complexes is also a remaining puzzle. Many adult ithomiine butterflies chemically protect themselves from predators by sequestering pyrrolizidine alkaloids, and are involved in mimicry rings, in which co-occurring species converge in conspicuous wing patterns that advertise toxicity to would-be predators. Based on their local abundance, ithomiine species are considered to act as mimetic models for many other distantly related species of Lepidoptera. Transparency as a means of concealment in this group is therefore puzzling, as conspicuous wing patterns are expected to be selected for maximal detection by predators. Nevertheless, recent studies suggest that transparent patches might also participate in the aposematic signal and that selection acts on the transmission properties of these patches, such as the degree of transparency (Pinna et al., 2020). Therefore, transparent aposematic Lepidoptera may benefit from a double protection via a ‘dual signal’ from predation. For instance, a transparent species may be less detectable than an opaque species, but if detected by a predator may be recognized as unpalatable due to the aposematic wing pattern (Tullberg et al., 2005; Willmott et al., 2017; Cuthill, 2019). As such, Pinna et al. (2020) call for a change of paradigm in transparent mimetic lepidoptera: transparency not only enables camouflage but can also be part of aposematic signals. While there is clearly a high amount of structural diversity, future studies should also thoroughly quantify the relative contributions of micro- and nanostructures on associated optical effects, and how such structures develop and evolve, which may contribute to our ability to truly harness the power of transparency in nature and generate bio-inspired applications for anti-reflective materials.

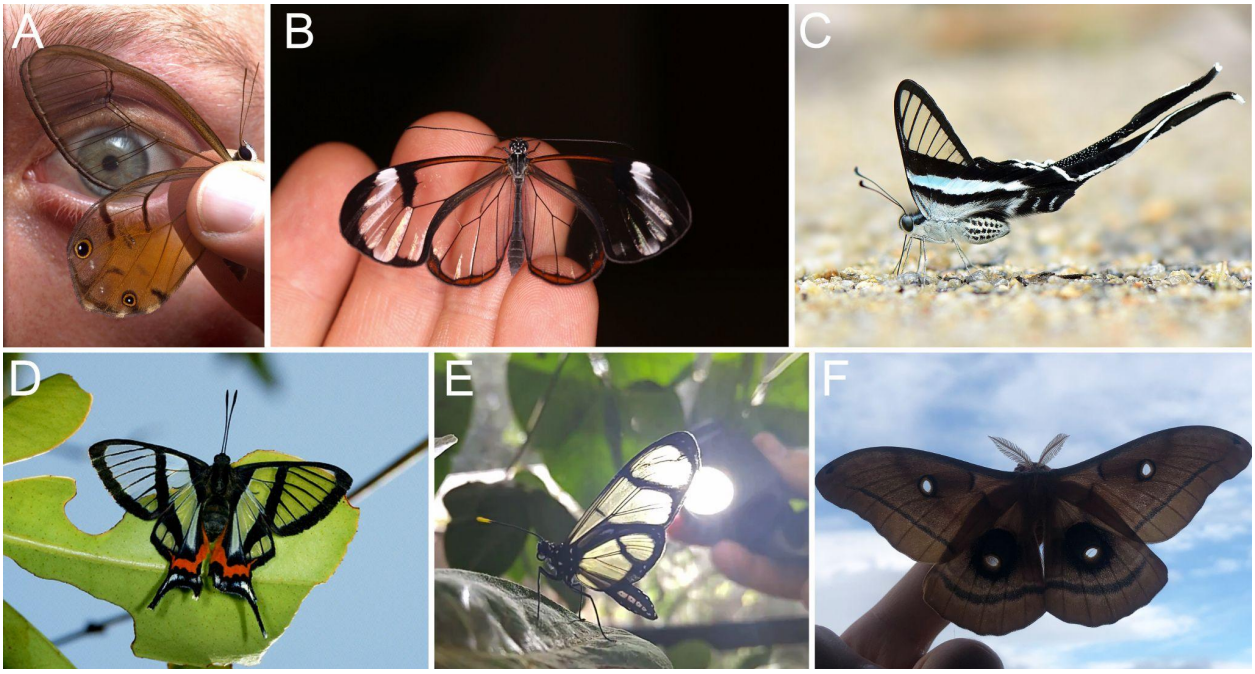


Figure 1.1. Examples of clearwing butterfly and moth species *in situ*.

A) The 'Amber phantom' *Haetera piera* (Nymphalidae: Haeterini) held up against my eye to illustrate its remarkably transparent wings (Image credit: Aaron Pomerantz, location: Yasuni, Ecuador). B) The 'glasswing butterfly' *Greta oto* (Nymphalidae: Ithomiini) freshly eclosed and perched on my hand (Image credit: Aaron Pomerantz, location: Gamboa, Panama). C) The 'longtail glasswing' *Lamproptera meges* (Papilionidae: Leptocircini) (Image credit: Yi-Kai Tea, shared with permission, location: northern Thailand). D) The 'Angel' *Chorinea sylphina* (Riodinidae: Riodinini) (Image credit: Adrian Hoskins, shared with permission, location: Machu Picchu, Peru). E) The 'giant glasswing' *Methona confusa* (Nymphalidae: Ithomiini) (Image credit: Aaron Pomerantz, location: Mindo, Ecuador). F) The 'giant silkmoth' *Antheraea polyphemus* (Saturniidae: Saturniini) (Image credit: Aaron Pomerantz, location: Berkeley, California, USA).

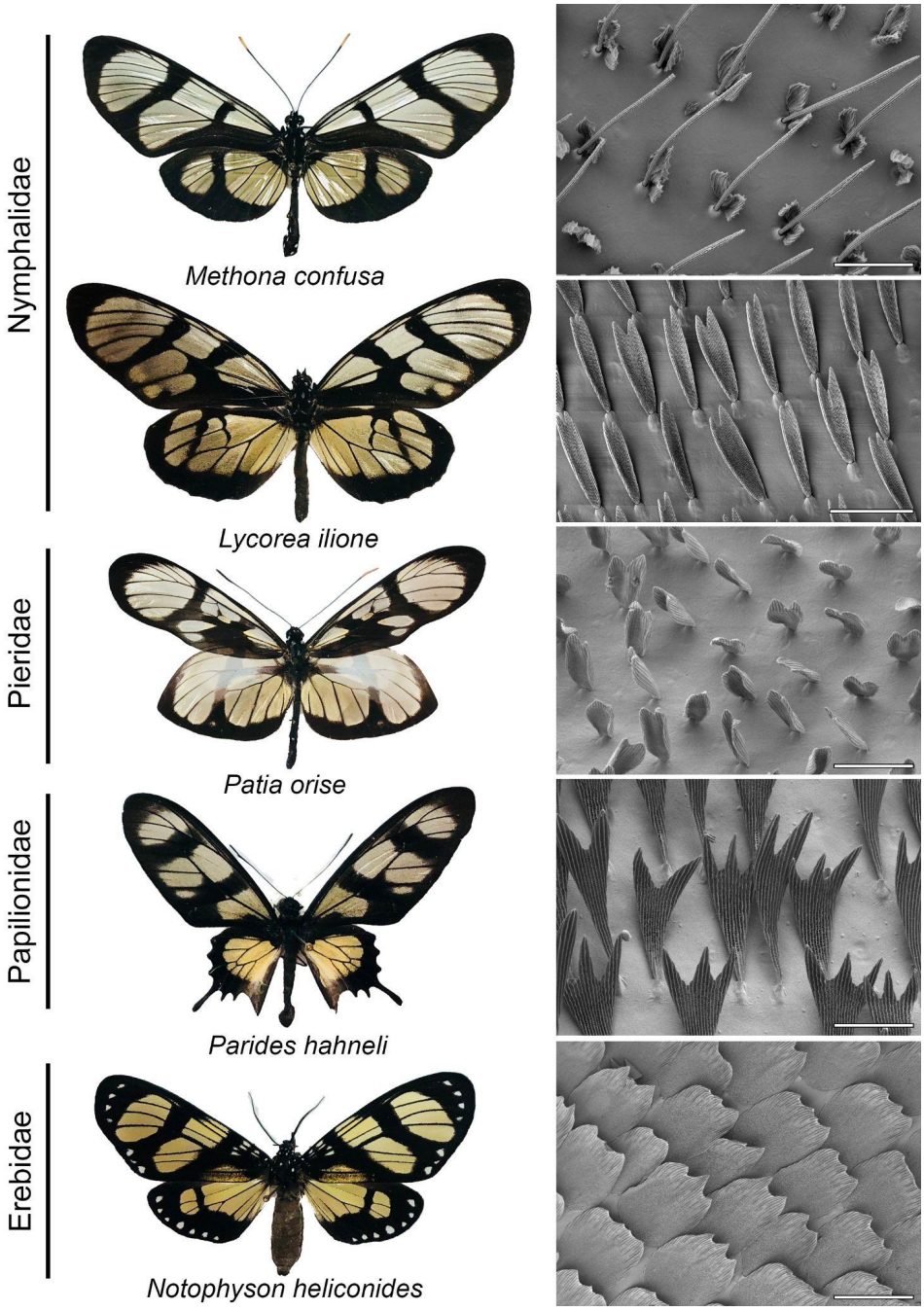


Figure 1.2. The *Methona* mimicry complex and associated scale morphologies.

The *Methona* mimicry complex is composed of distantly related butterfly and moth species whose wing patterns morphologically resemble the giant glasswing butterfly in the genus *Methona* (Nymphalidae: Ithomiinae). Scanning electron microscopy (SEM) reveals that the scale morphologies amongst co-mimics differ dramatically, yet the overall wing patterns bear striking similarity. Scale bars = 100 μm.

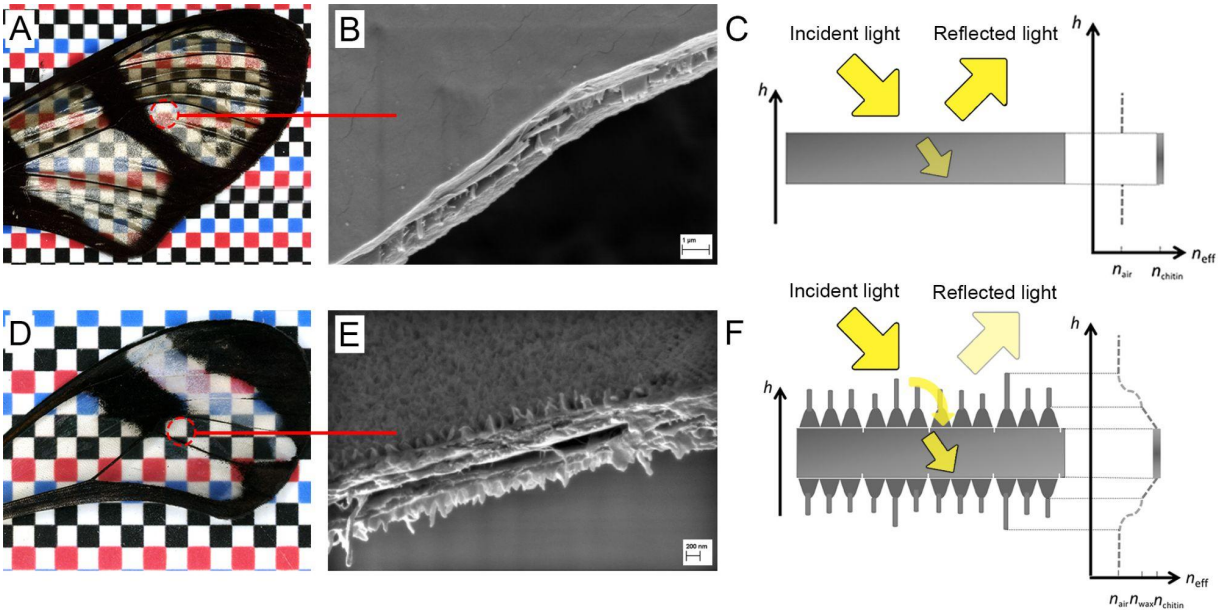


Figure 1.3 Comparison of wing surface ultrastructure and mechanism of transparency via anti-reflective nanostructures.

A) The giant glasswing *Methona confusa* has reflective wings. B) Scanning electron microscopy of the wing surface reveals that it is smooth and devoid of nanostructures. C) Schematic of wing surface reflection. Light travels through air and interacts with the high refractive index chitin, producing surface reflection. D) The glasswing butterfly *Greta oto* has highly transparent wings. E) Scanning electron microscopy of the wing surface reveals that it is coated with nanopillars. F) Schematic of wing surface reflection when nanostructures are present on the wing surface, which function to create a refractive index gradient to reduce surface reflection.

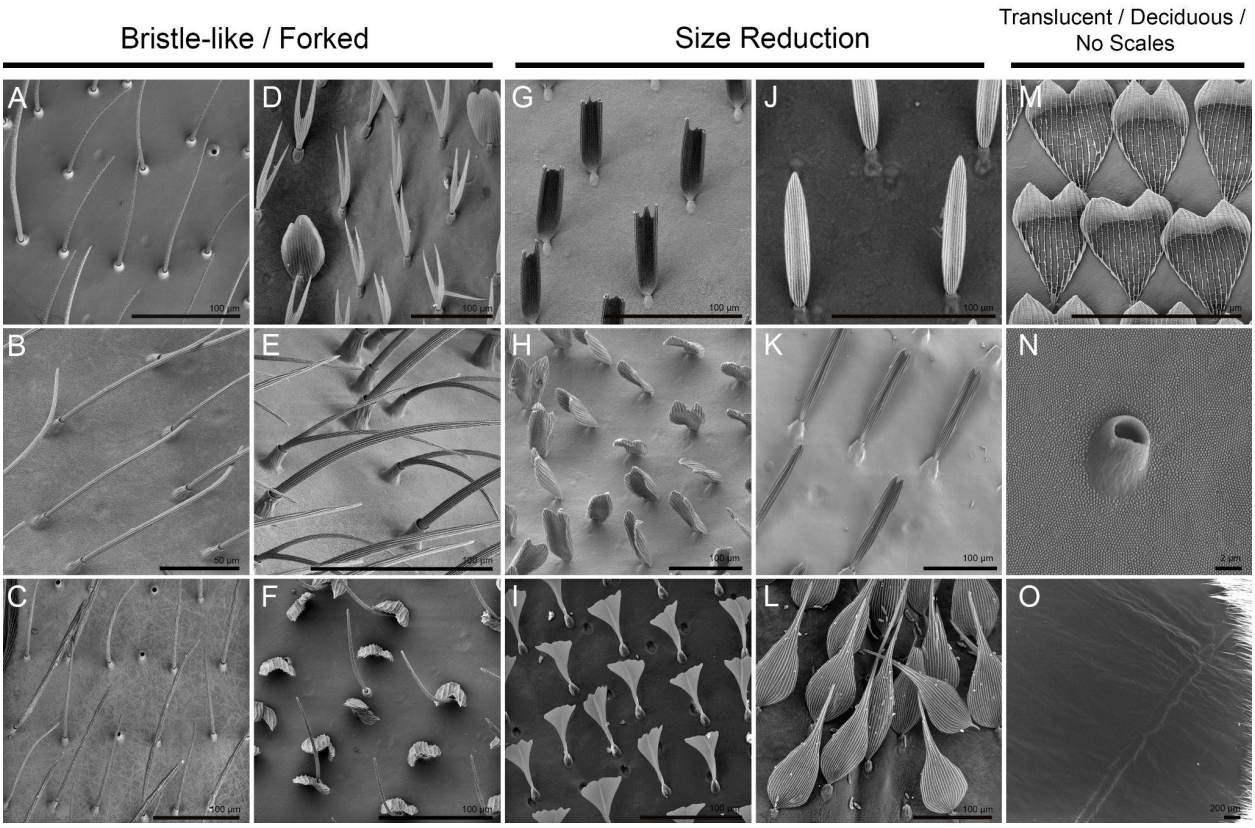


Figure 1.4. Compilation of scanning electron microscopy showcasing the diversity of scale morphologies in clearwing Lepidoptera.

Clearwing butterflies and moths produce a wide range of strategies to modify wing scale morphology. Scales can be modified so that they have thin, bristle-like or forked morphologies, for example in (A) *Olyras insignis* (Nymphalidae: Ithomiini), (B) *Cosmosoma myrodora* (Erebidae: Arctiinae), (C) *Dulcedo polita* (Nymphalidae: Haeterini), (D) *Acraea andromacha* (Nymphalidae: Acraeini), (E) *Godyris duilia* (Nymphalidae: Ithomiini), and (F) *Thyridia psidii* (Nymphalidae: Ithomiini). Scales can also be modified such that they are reduced in size and vertically oriented, for instance in species such as (G) *Chorinaea faunus* (Rionidae: Riodinini), (H) *Patia orise* (Pieridae: Dismorphiinae), (I) *Lamproptera meges* (Papilionidae: Leptocircini), (J) *Parantica sita* (Nymphalidae: Danaina), (K) *Eurytides aguari* (Papilionidae: Leptocircini). (L) *Parnassius glacialis* (Papilionidae: Parnassiini). The scales themselves can be translucent, for example in (M) *Ideopsis vitraea* (Nymphalidae: Danaina), scales can be deciduous and fall out upon eclosion, leaving only the socket cell, for example in (N) *Hemaris thysbe* (Sphingidae: Hemaris) and scale may lack entirely because they do not form in parts of the wing during pupal development, as in (O) *Antheraea polyphemus* (Saturniidae: Saturniini).

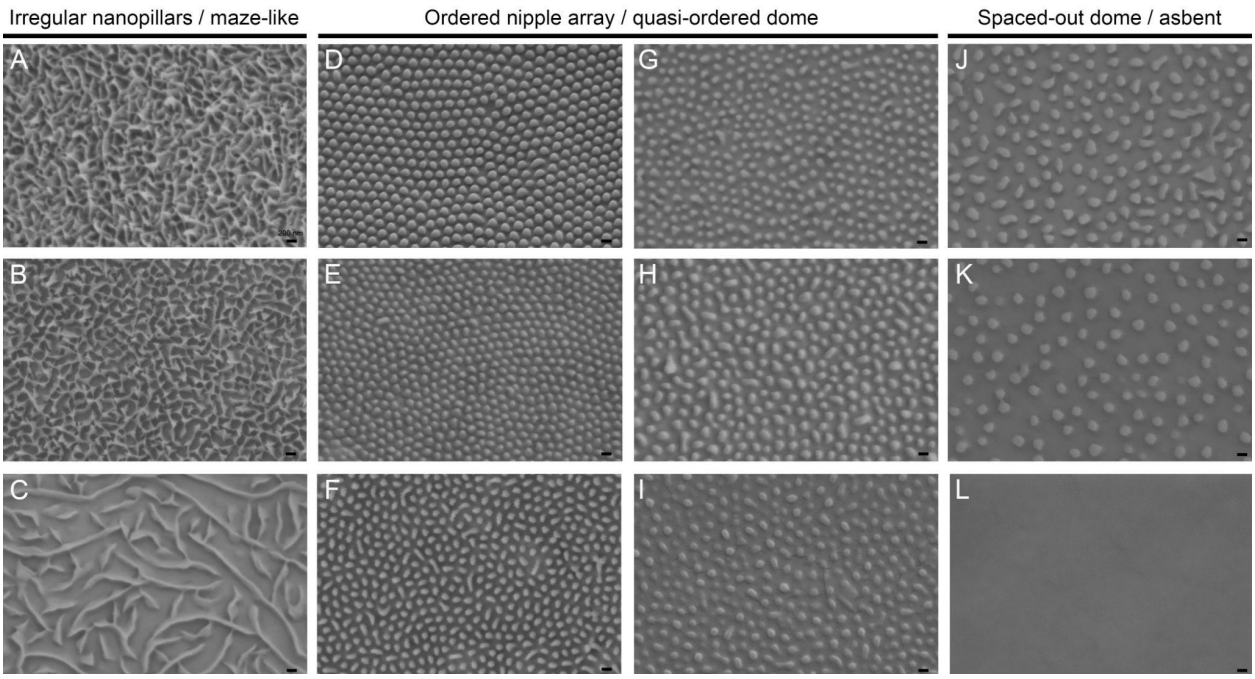


Figure 1.5. Scanning electron microscopy reveals several distinct classes of wing surface nanostructures in transparent butterflies and moths.

Clearwing butterflies and moths have evolved a variety of nanostructures on the surface of the wing membrane, many of which have been shown to generate anti-reflective properties. For instance, irregularly arranged nanopillars can be found in (A) *Greta oto* (Nymphalidae: Ithomiini), (B) *Cithaerias merolina* (Nymphalidae: Haeterini), and maze-like nanostructures can be found in (C) *Parantica sita* (Nymphalidae: Danaina). Ordered nipple arrays can be found in species such as (D) *Hemaris thysbe* (Sphingidae: Hemaris), (E) *Synanthedon kathyae* (Sesiidae: Sesiinae), (F) *Chorinaea faunus* (Rionidae: Riodinini), and quasi-ordered dome structures can be found in species such as (G) *Cosmosoma myrodora* (Erebidae: Arctiinae), (H) *Cocytia durvillii* (Erebidae), and (I) *Lamproptera meges* (Papilionidae: Leptocircini). Spaced-out dome nanostructures can be found in (J) *Patia orise* (Pieridae: Dismorphiinae), and (K) *Dioptis* sp. (Notodontidae: Dioptini) or the wing membrane can be smooth and absent of any nanostructure, such as (L) the scaleless window region of *Antheraea polyphemus*. Scale bars = 20 μm.

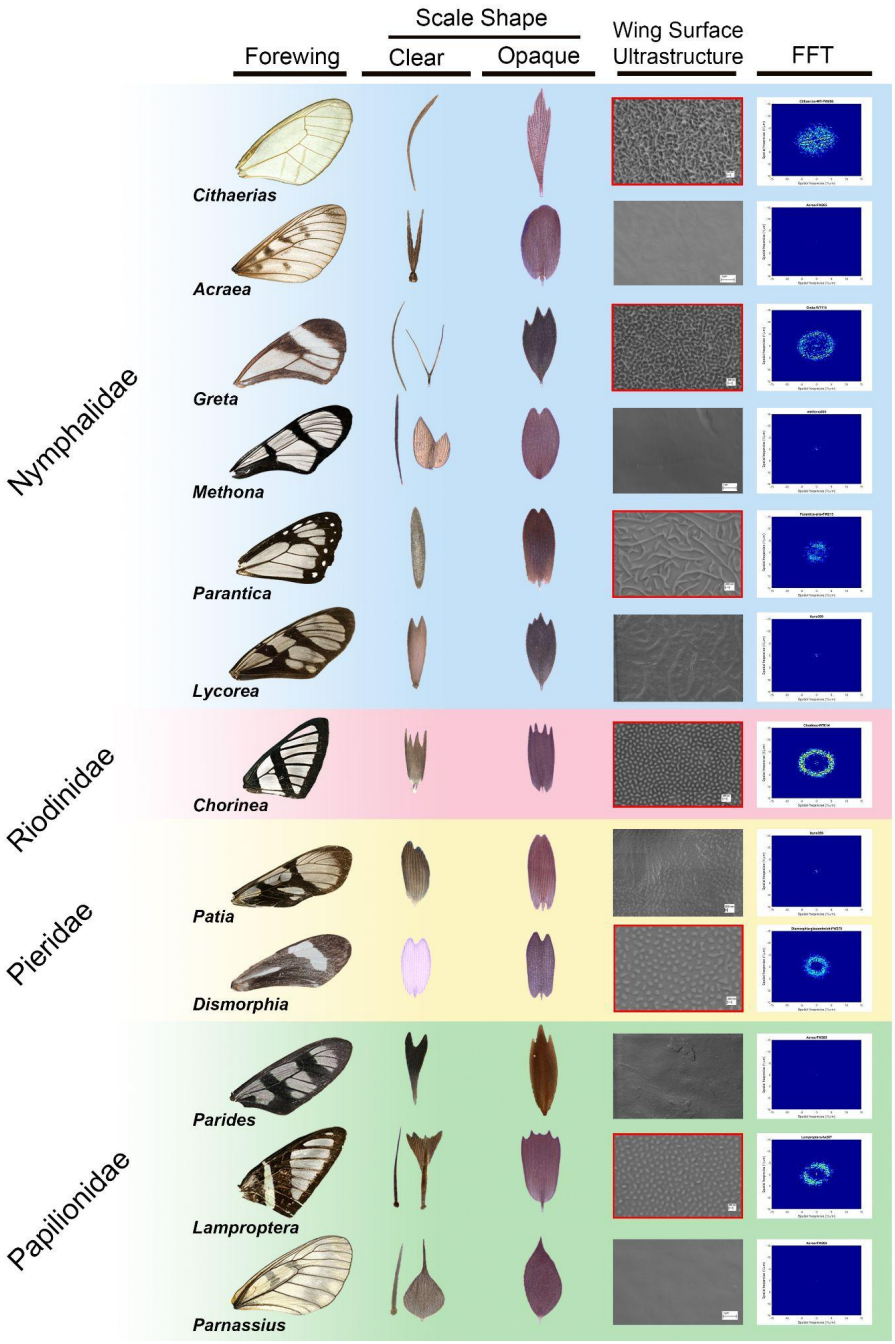


Figure 1.6. Compilation of clearwing butterfly species and their associated scale and wing surface features.

Representative images of clearwing species showcasing the whole forewing, a representative scale from a clear wing region and an opaque wing region, scanning electron micrograph (SEM) of the wing surface ultrastructure (red outline indicating nanostructures are present) and a two-dimensional fast Fourier transformation (FFT) obtained from the SEM.

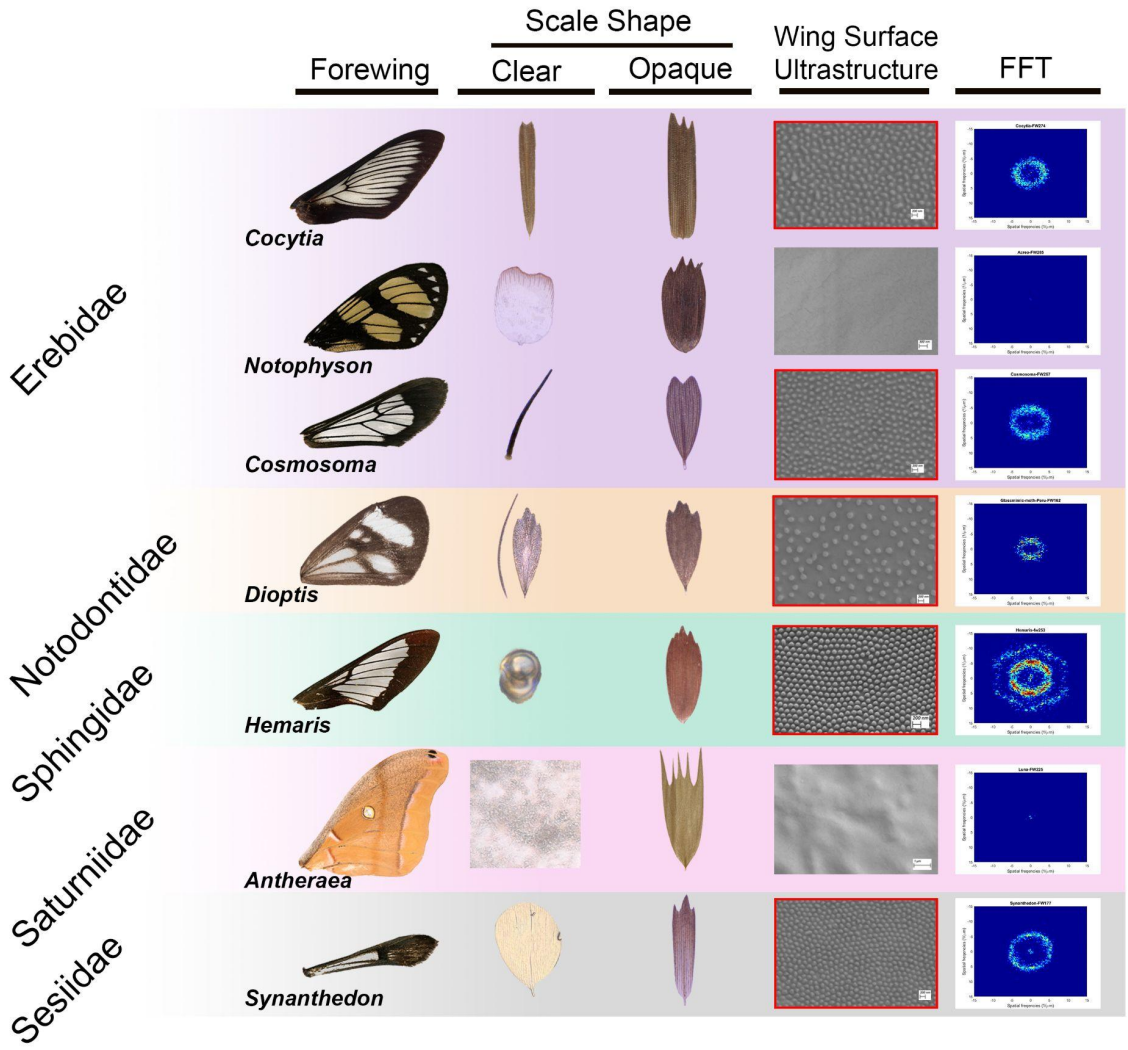


Figure 1.7. Compilation of clearwing moth species and their associated scale and wing surface features.

Representative images of clearwing species showcasing the whole forewing, a representative scale from a clear wing region and an opaque wing region, scanning electron micrograph (SEM) of the wing surface ultrastructure (red outline indicating nanostructures are present) and a two-dimensional fast Fourier transformation (FFT) obtained from the SEM.

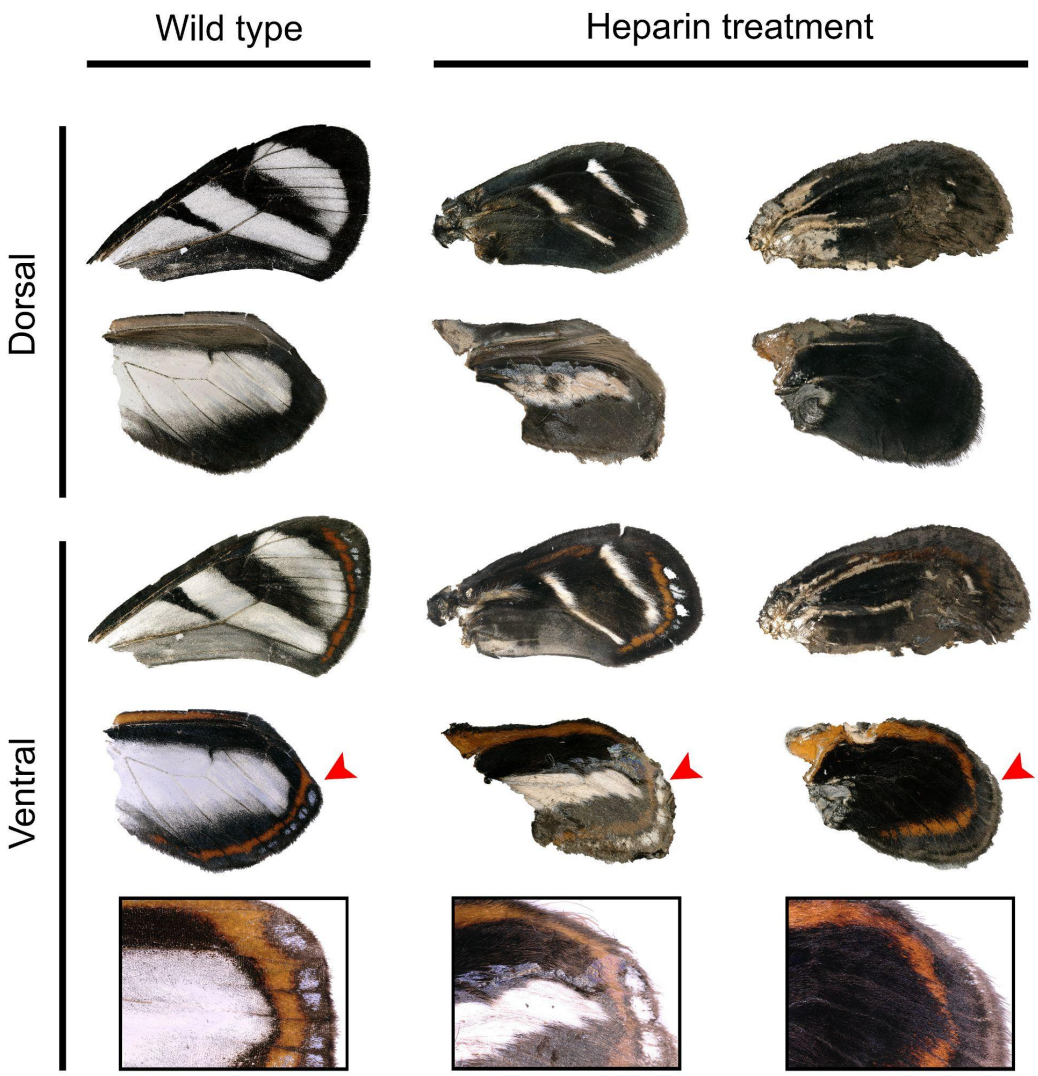


Figure 1.8. Effects of heparin treatments on the wing pattern of *Oleria onega* (Nymphalidae: Ithomiini).

Heparin treatments modify the boundaries of transparent and opaque regions of the wing, such as an expansion of melanic central symmetry system patterns, suggesting a role of Wnt signaling ligands in demarcating wing patterns in glasswing butterflies.

References

1. Arias, M., Elias, M., Andraud, C., Berthier, S., and Gomez, D. (2020). Transparency improves concealment in cryptically coloured moths. *J Evol Biol* 33, 247–252. doi:10.1111/jeb.13560.
2. Arias, M., Mappes, J., Desbois, C., Gordon, S., McClure, M., Elias, M., et al. (2019). Transparency reduces predator detection in mimetic clearwing butterflies. *Funct Ecol* 33, 1110–1119. doi:10.1111/1365-2435.13315.
3. Bagge, L. E. (2019). Not As Clear As It May Appear: Challenges Associated with Transparent Camouflage in the Ocean. *Integrative and Comparative Biology* 59, 1653–1663. doi:10.1093/icb/icz066.
4. Bates, H. W. (1862). XXXII. Contributions to an Insect Fauna of the Amazon Valley. Lepidoptera: Heliconidae. *Transactions of the Linnean Society of London* 23, 495–566. doi:10.1111/j.1096-3642.1860.tb00146.x.
5. Benedek, G. B. (1971). Theory of Transparency of the Eye. *Appl. Opt.* 10, 459. doi:10.1364/AO.10.000459.
6. Bernhard, C. G., Gemne, G., and Sallstrom, J. (1970). Comparative ultrastructure of corneal surface topography in insects with aspects on phylogenesis and function. *Z. Vergl. Physiol.* 67, 1–25. doi:10.1007/BF00298117.
7. Bernhard, C. G., and Miller, W. H. (1962). A Corneal Nipple Pattern in Insect Compound Eyes. *Acta Physiologica Scandinavica* 56, 385–386. doi:10.1111/j.1748-1716.1962.tb02515.x.
8. Berthier, S. (2007). *Iridescences*. New York, NY: Springer New York doi:10.1007/b106784.
9. Binetti, V. R., Schiffman, J. D., Leaffer, O. D., Spanier, J. E., and Schauer, C. L. (2009). The natural transparency and piezoelectric response of the *Greta oto* butterfly wing. *Integr. Biol.* 1, 324. doi:10.1039/b820205b.
10. Blagodatski, A., Sergeev, A., Kryuchkov, M., Lopatina, Y., and Katanaev, V. L. (2015). Diverse set of Turing nanopatterns coat corneae across insect lineages. *Proc Natl Acad Sci USA* 112, 10750–10755. doi:10.1073/pnas.1505748112.
11. Bogert, C. M. (1949). Thermoregulation in Reptiles, A Factor in Evolution. *Evolution* 3, 195. doi:10.2307/2405558.
12. Brown, K. S. (1984). Adult-obtained pyrrolizidine alkaloids defend ithomiine butterflies against a spider predator. *Nature* 309, 707–709. doi:10.1038/309707a0.
13. Cai, J., and Qi, L. (2015). Recent advances in antireflective surfaces based on nanostructure arrays. *Mater. Horiz.* 2, 37–53. doi:10.1039/C4MH00140K.
14. Chapman, G. (1976). Transparency in organisms. *Experientia* 32, 123–125. doi:10.1007/BF01932661.
15. Cuthill, I. C. (2019). Camouflage. *J Zool* 308, 75–92. doi:10.1111/jzo.12682.
16. Day, C. R., Hanly, J. J., Ren, A., and Martin, A. (2019). Sub-micrometer insights into the cytoskeletal dynamics and ultrastructural diversity of butterfly wing scales. *Developmental Dynamics* 248, 657–670. doi:10.1002/dvdy.63.
17. Deparis, O., Mouchet, S., Dellieu, L., Colomer, J.-F., and Sarrazin, M. (2014). Nanostructured Surfaces: Bioinspiration for Transparency, Coloration and

Wettability. *Materials Today: Proceedings* 1, 122–129. doi:10.1016/j.matpr.2014.09.008.
18. Deshmukh, R., Baral, S., Gandhimathi, A., Kuwalekar, M., and Kunte, K. (2018). Mimicry in butterflies: co-option and a bag of magnificent developmental genetic tricks: Mimicry in butterflies. *WIREs Dev Biol* 7, e291. doi:10.1002/wdev.291.
19. Dinwiddie, A., Null, R., Pizzano, M., Chuong, L., Leigh Krup, A., Ee Tan, H., et al. (2014). Dynamics of F-actin prefigure the structure of butterfly wing scales. *Developmental Biology* 392, 404–418. doi:10.1016/j.ydbio.2014.06.005.
20. Galant, R., Skeath, J. B., Paddock, S., Lewis, D. L., and Carroll, S. B. (1998). Expression pattern of a butterfly achaete-scute homolog reveals the homology of butterfly wing scales and insect sensory bristles. *Current Biology* 8, 807–813. doi:10.1016/S0960-9822(98)70322-7.
21. Gomez, D., Pinna, C., Pairraire, J., Arias, M., Barbut, J., Pomerantz, A., et al. (2020). Transparency in butterflies and moths: structural diversity, optical properties and ecological relevance. Evolutionary Biology doi:10.1101/2020.05.14.093450.
22. Gomez, D., Pinna, C., Pairraire, J., Arias, M., Barbut, J., Pomerantz, A., et al. (2021). Wing transparency in butterflies and moths: structural diversity, optical properties, and ecological relevance. *Ecol Monogr*. doi:10.1002/ecm.1475.
23. Han, Z., Jiao, Z., Niu, S., and Ren, L. (2019). Ascendant bioinspired antireflective materials: Opportunities and challenges coexist. *Progress in Materials Science* 103, 1–68. doi:10.1016/j.pmatsci.2019.01.004.
24. Hernández-Chavarría, F., Hernández, A., and Sittenfeld, A. (2004). The “windows”, scales, and bristles of the tropical moth Rothschildia lebeau (Lepidoptera: Saturniidae). *Rev Biol Trop* 52, 919–926.
25. Johnsen, S. (2001). Hidden in Plain Sight: The Ecology and Physiology of Organismal Transparency. *The Biological Bulletin* 201, 301–318. doi:10.2307/1543609.
26. Johnsen, S. (2014). Hide and Seek in the Open Sea: Pelagic Camouflage and Visual Countermeasures. *Annu. Rev. Mar. Sci.* 6, 369–392. doi:10.1146/annurev-marine-010213-135018.
27. Johnsen, S., and Widder, E. A. (1999). The Physical Basis of Transparency in Biological Tissue: Ultrastructure and the Minimization of Light Scattering. *Journal of Theoretical Biology* 199, 181–198. doi:10.1006/jtbi.1999.0948.
28. Kryuchkov, M., Blagodatski, A., Cherepanov, V., and Katanaev, V. L. (2017). “Arthropod Corneal Nanocoatings: Diversity, Mechanisms, and Functions,” in *Functional Surfaces in Biology III* Biologically-Inspired Systems., eds. S. N. Gorb and E. V. Gorb (Cham: Springer International Publishing), 29–52. doi:10.1007/978-3-319-74144-4_2.
29. Liu, J., Chen, Z., Xiao, Y., Asano, T., Li, S., Peng, L., et al. (2021). Lepidopteran wing scales contain abundant cross-linked film-forming histidine-rich cuticular proteins. *Commun Biol* 4, 491. doi:10.1038/s42003-021-01996-4.
30. Mazo-Vargas, A., Concha, C., Livraghi, L., Massardo, D., Wallbank, R. W. R., Zhang, L., et al. (2017). Macroevolutionary shifts of *WntA* function potentiate butterfly wing-pattern diversity. *Proc Natl Acad Sci USA* 114, 10701–10706. doi:10.1073/pnas.1708149114.

31. McClure, M., Clerc, C., Desbois, C., Meichanetzoglou, A., Cau, M., Bastin-Héline, L., et al. (2019). Why has transparency evolved in aposematic butterflies? Insights from the largest radiation of aposematic butterflies, the Ithomiini. *Proc. R. Soc. B.* 286, 20182769. doi:10.1098/rspb.2018.2769.
32. Meyer, R. A. (1979). Light scattering from biological cells: dependence of backscatter radiation on membrane thickness and refractive index. *Appl. Opt.* 18, 585. doi:10.1364/AO.18.000585.
33. Miller, W. H. (1979). “Ocular Optical Filtering,” in *Comparative Physiology and Evolution of Vision in Invertebrates* Handbook of Sensory Physiology., ed. H. Autrum (Berlin, Heidelberg: Springer Berlin Heidelberg), 69–143. doi:10.1007/978-3-642-66999-6_3.
34. Munro, J. T., Medina, I., Walker, K., Moussalli, A., Kearney, M. R., Dyer, A. G., et al. (2019). Climate is a strong predictor of near-infrared reflectance but a poor predictor of colour in butterflies. *Proc. R. Soc. B.* 286, 20190234. doi:10.1098/rspb.2019.0234.
35. Narasimhan, V., Siddique, R. H., Lee, J. O., Kumar, S., Ndjamen, B., Du, J., et al. (2018). Multifunctional biophotonic nanostructures inspired by the longtail glasswing butterfly for medical devices. *Nature Nanotech* 13, 512–519. doi:10.1038/s41565-018-0111-5.
36. Nijhout, H. F. (2001). Elements of butterfly wing patterns. *J. Exp. Zool.* 291, 213–225. doi:10.1002/jez.1099.
37. Perez Goodwyn, P., Maezono, Y., Hosoda, N., and Fujisaki, K. (2009). Waterproof and translucent wings at the same time: problems and solutions in butterflies. *Naturwissenschaften* 96, 781–787. doi:10.1007/s00114-009-0531-z.
38. Pinna, C., Vilbert, M., Borensztajn, S., de Marcillac, W. D., Piron-Prunier, F., Pomerantz, A. F., et al. (2020). Mimicry drives convergence in structural and light transmission features of transparent wings in Lepidoptera. Evolutionary Biology doi:10.1101/2020.06.30.180612.
39. Pomerantz, A. F., Siddique, R. H., Cash, E. I., Kishi, Y., Pinna, C., Hammar, K., et al. (2021). Developmental, cellular, and biochemical basis of transparency in clearwing butterflies. *Journal of Experimental Biology*, jeb.237917. doi:10.1242/jeb.237917.
40. Serfas, M. S., and Carroll, S. B. (2005). Pharmacologic approaches to butterfly wing patterning: Sulfated polysaccharides mimic or antagonize cold shock and alter the interpretation of gradients of positional information. *Developmental Biology* 287, 416–424. doi:10.1016/j.ydbio.2005.09.014.
41. Siddique, R. H., Gomard, G., and Hölscher, H. (2015). The role of random nanostructures for the omnidirectional anti-reflection properties of the glasswing butterfly. *Nat Commun* 6, 6909. doi:10.1038/ncomms7909.
42. Stavenga, D. G., Foletti, S., Palasantzas, G., and Arikawa, K. (2006). Light on the moth-eye corneal nipple array of butterflies. *Proc. R. Soc. B.* 273, 661–667. doi:10.1098/rspb.2005.3369.
43. Stavenga, D. G., Matsushita, A., Arikawa, K., Leertouwer, H. L., and Wilts, B. D. (2012). Glass scales on the wing of the swordtail butterfly *Graphium sarpedon* act as thin film polarizing reflectors. *Journal of Experimental Biology* 215, 657–662. doi:10.1242/jeb.066902.

44. 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. doi:10.7554/eLife.52187.
45. Tsai, C.-C., Childers, R. A., Nan Shi, N., Ren, C., Pelaez, J. N., Bernard, G. D., et al. (2020). Physical and behavioral adaptations to prevent overheating of the living wings of butterflies. *Nat Commun* 11, 551. doi:10.1038/s41467-020-14408-8.
46. Tullberg, B. S., Merilaita, S., and Wiklund, C. (2005). Aposematism and crypsis combined as a result of distance dependence: functional versatility of the colour pattern in the swallowtail butterfly larva. *Proc. R. Soc. B.* 272, 1315–1321. doi:10.1098/rspb.2005.3079.
47. Tuthill, J. C., and Johnsen, S. (2006). Polarization sensitivity in the red swamp crayfish *Procambarus clarkii* enhances the detection of moving transparent objects. *Journal of Experimental Biology* 209, 1612–1616. doi:10.1242/jeb.02196.
48. Wanasekara, N. D., and Chalivendra, V. B. (2011). Role of surface roughness on wettability and coefficient of restitution in butterfly wings. *Soft Matter* 7, 373–379. doi:10.1039/C0SM00548G.
49. Wasik, B. R., Liew, S. F., Lilien, D. A., Dinwiddie, A. J., Noh, H., Cao, H., et al. (2014). Artificial selection for structural color on butterfly wings and comparison with natural evolution. *Proceedings of the National Academy of Sciences* 111, 12109–12114. doi:10.1073/pnas.1402770111.
50. Willmott, K. R., Robinson Willmott, J. C., Elias, M., and Jiggins, C. D. (2017). Maintaining mimicry diversity: optimal warning colour patterns differ among microhabitats in Amazonian clearwing butterflies. *Proc. R. Soc. B.* 284, 20170744. doi:10.1098/rspb.2017.0744.
51. Yoshida, A., Motoyama, M., Kosaku, A., and Miyamoto, K. (1997). Antireflective Nanoprotuberance Array in the Transparent Wing of a Hawkmoth, Cephonodes hylas. *Zoological Science* 14, 737–741. doi:10.2108/zsj.14.737.
52. Zhang, L., Martin, A., Perry, M. W., van der Burg, K. R. L., Matsuoka, Y., Monteiro, A., et al. (2017). Genetic Basis of Melanin Pigmentation in Butterfly Wings. *Genetics* 205, 1537–1550. doi:10.1534/genetics.116.196451.

Preface to Chapter 2

Until recently, research attention has been primarily devoted to transparent organisms in aquatic environments, which are much more frequent compared to cases of transparency on land. Recent studies have explored aspects of structural diversity, optical properties and phylogenetic distribution of transparency within butterflies and moths, which often possess scales of altered morphology and reduced size, and the presence of membrane surface nanostructures that dramatically reduce reflection. However, the developmental processes underlying transparency are currently unknown, let alone the dynamic formation of butterfly nanostructures in general. This presents a gap in our understanding of how wing scale microstructure morphologies and membrane surface nanostructures are generated within a living system. We therefore set out to explore the development of wing transparency in *Greta oto*, which has thin, vertically oriented scales and irregularly arranged nanopillars on the wing membrane that enable omnidirectional anti-reflective properties. We showed that this species could be easily reared in the lab, which allowed us to employ a multitude of techniques, including confocal and electron microscopy, GC-MS, optical spectroscopy and analytical simulations. We were then able to create a time-series of wing development for the first time, comparing transparent and non-transparent wing regions. In this study, we highlight several novel features related to the development of wing transparency. (1) We found that early into wing development, scale precursor cell density is reduced in transparent regions, and cytoskeletal organization (such as actin bundle and microtubule distribution) during scale growth differs between thin, bristle-like scale morphologies within transparent regions and flat, round scale morphologies within opaque regions. These features of modified scale development aid in exposing the wing surface in transparent regions, which contains anti-reflective nanopillars. (2) We showed that these sub-wavelength nanopillars on the wing membrane are an epicuticular wax-based layer, which derives from wing epithelial cells and their associated microvillar projections, and nanopillars are composed predominantly of long-chain n-alkanes. (3) By chemically and physically removing wax-based nanopillars, along with analytical simulations, we further demonstrated their role in generating anti-reflective properties. To the best of our knowledge, the biochemical composition and ontogeny of wax-based anti-reflective nanostructures in Lepidoptera wings have not been characterized until now. (4) Finally, we tested the solubility of wing surface nanostructures in several additional species of clearwing Lepidoptera. We found that the wings of ‘glasswing’ butterflies (Nymphalidae: Ithomiini) and ‘phantom’ butterflies (Nymphalidae: Satyrinae) harbor wax-based nanopillars, suggesting that this class of anti-reflective nanostructure evolved multiple times independently.

This chapter has been published in the Journal of Experimental Biology, https://doi.org/10.1242/jeb.237917

Chapter 2: Developmental, cellular, and biochemical basis of transparency in clearwing butterflies

Authors

Aaron F. Pomerantz[1,2*], Radwanul H. Siddique[3,4], Elizabeth I. Cash[5], Yuriko Kishi[6,7], Charline Pinna[8], Kasia Hammar[2], Doris Gomez[9], Marianne Elias[8], Nipam H. Patel[1,2,6*]

Affiliations

[1.] Department Integrative Biology, University of California Berkeley, Berkeley, CA 94720.
[2.] Marine Biological Laboratory, Woods Hole, MA 02543.
[3.] Image Sensor Lab, Samsung Semiconductor, Inc., 2 N Lake Ave. Ste. 240, Pasadena, CA 91101.
[4.] Department of Medical Engineering, California Institute of Technology, Pasadena, CA 91125.
[5.] Department of Environmental Science, Policy, & Management, University of California Berkeley, Berkeley, CA 94720.
[6.] Department Molecular Cell Biology, University of California Berkeley, Berkeley, CA 94720.
[7.] Department of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125.
[8.] ISYEB, 45 rue Buffon, CP50, Paris, CNRS, MNHN, Sorbonne Université, EPHE, Université des Antilles, France.
[9.] CEFE, 1919 route de Mende, Montpellier, CNRS, Univ Montpellier, Univ Paul Valéry Montpellier 3, EPHE, IRD, France.

*** Corresponding author**. Email: pomerantz_aaron@berkeley.edu, npatel@mbl.edu

Abstract

The wings of butterflies and moths (Lepidoptera) are typically covered with thousands of flat, overlapping scales that endow the wings with colorful patterns. Yet, numerous species of Lepidoptera have evolved highly transparent wings, which often possess scales of altered morphology and reduced size, and the presence of membrane surface nanostructures that dramatically reduce reflection. Optical properties and anti-reflective nanostructures have been characterized for several 'clearwing' Lepidoptera, but the developmental processes underlying wing transparency are unknown. Here, we applied confocal and electron microscopy to create a developmental time series in the glasswing butterfly, *Greta oto*, comparing transparent and non-transparent wing regions. We found that during early wing development, scale precursor cell density was reduced in transparent regions, and cytoskeletal organization during scale growth differed between thin, bristle-like scale morphologies within transparent regions and flat, round scale morphologies within opaque regions. We also show that nanostructures on the wing membrane surface are composed of two layers: a lower layer of regularly arranged nipple-like nanostructures, and an upper layer of irregularly arranged wax-based

nanopillars composed predominantly of long-chain n-alkanes. By chemically removing wax-based nanopillars, along with optical spectroscopy and analytical simulations, we demonstrate their role in generating anti-reflective properties. These findings provide insight into morphogenesis and composition of naturally organized microstructures and nanostructures, and may provide bioinspiration for new anti-reflective materials.

Key Words: Anti-reflection, nanostructures, glasswing, Lepidoptera, cytoskeleton, morphogenesis

Introduction

The wings of butterflies and moths (Lepidoptera) have inspired studies across a variety of scientific fields, including evolutionary biology, ecology and biophysics (Beldade and Brakefield, 2002; Prum et al., 2006; Gilbert and Singer, 1975). Lepidopteran wings are generally covered with rows of flat, partially overlapping scales that endow the wings with colorful patterns. Adult scales are chitin-covered projections that serve as the unit of color for the wing. Each scale can generate color through pigmentation via molecules that selectively absorb certain wavelengths of light, structural coloration, which results from light interacting with the physical nanoarchitecture of the scale; or a combination of both pigmentary and structural coloration (Stavenga et al., 2014; Thayer et al., 2020). Cytoskeletal dynamics, including highly organized F-actin filaments during scale cell development, play essential roles in wing scale elongation and prefigure aspects of scale ultrastructure (Dinwiddie et al., 2014; Day et al., 2019).

In contrast to typical colorful wings, numerous species of butterflies and moths possess transparent wings that allow light to pass through, so that objects behind them can be distinctly seen (Fig. 1A–H) (Goodwyn et al., 2009; Yoshida et al., 1997; Siddique et al., 2015). This trait has been interpreted as an adaptation in the context of camouflage, in which some lineages evolved transparent wings as crypsis to reduce predation (Arias et al., 2019; 2020; Mcclure et al., 2019). Transparency results from the transmission of light across the visible spectrum through a material, in this case the chitin membrane, without significant absorption or reflection. Levels of reflection are largely determined by the differences in refractive indices between biological tissues and the medium, and a larger difference results in higher surface reflection. Previous studies on transparency in nature have primarily focused on aquatic organisms, which are frequently transparent, aided by the close match between the refractive indices of their aqueous tissue and the surrounding medium – water (e.g. Johnsen, 2001). By contrast, transparency is rare and more challenging to achieve on land, primarily owing to the large difference between the refractive indices of terrestrial organism's tissue (n=∼1.3–1.5) and air (n=1), which results in significant surface reflection (Yoshida et al., 1997; Johnsen, 2014; Bagge, 2019).

Nevertheless, some organisms have evolved morphological innovations that overcome the challenges of terrestrial transparency, notably in the form of anti-reflective nanostructures. Early experiments elucidated highly ordered sub-wavelength nanostructures (termed 'nipple arrays') on the corneal surface of insect eyes (Bernhard, 1962). These structures were found to generally be ∼150–250 nm in height and spaced ∼200 nm apart, which reduces reflection across a broad range of wavelengths by

creating a smoother gradient of refractive indices between air and chitin (Stavenga et al., 2006). Nanostructure arrays have also been identified on the wings of cicadas, which help to reduce surface reflection over the visible spectrum (Huang et al., 2015).

Some lepidopterans possess 'clear wings' in which scales have undergone modifications that enable light to reach the wing membrane surface. The wing itself is composed of chitin and has some inherent transparency, but owing to the high refractive index of chitin, n=1.56, the wing surface reflects light (Vukusic et al., 1999). For example, the butterfly *Methona confusa* (Nymphalidae: Ithomiini) has an exposed wing membrane that lacks nanostructures on the surface, and as a result, the wing is somewhat transparent, but retains a high degree of reflectivity (Fig. 1A–E). Conversely, the longtail glasswing, *Chorinea faunus* (Riodinidae), contains small, widely spaced scales and dome-shaped chitin nanoprotuberances on the membrane that generate anti-reflective properties (Narasimhan et al., 2018). The hawkmoth, *Cephonodes hylas* (Sphingidae), has nude wings owing to deciduous scales that fall out upon eclosion, and possesses anti-reflective nanostructures on its wing surface that morphologically resemble insect corneal nipple arrays (Yoshida et al., 1997). Nipple array nanostructures have also been characterized in transparent wing regions of the tiger moth *Cacostatia ossa* (Erebidae) (Deparis et al., 2014). Finally, the glasswing butterfly, *Greta oto* (Nymphalidae: Ithomiini), contains thin, vertically oriented scales, allowing the wing surface to be exposed, along with nanopillars that coat the surface (Fig. 1F–J). These irregularly arranged nanopillars feature a random height distribution and enable omnidirectional anti-reflective properties (Fig. 1I,J) (Siddique et al., 2015; Binetti et al., 2009). More recent studies have explored aspects of structural diversity, optical properties, phylogenetic distribution and ecological relevance of transparency within a wide range of butterflies and moths, highlighting that transparency has evolved multiple times independently and may present evolutionary benefits (Mcclure et al., 2019; Gomez et al., 2020 preprint; Pinna et al., 2020 preprint).

Lepidoptera are proving to be an excellent group to investigate transparency on land, but the developmental processes underlying wing transparency are currently unknown. This presents a gap in our understanding of lepidopteran wing evolution and diversification, as transparent butterflies and moths contain multitudes of intriguing scale modifications and sub-wavelength cuticular nanostructures (Gomez et al., 2020 preprint; Pinna et al., 2020 preprint). Therefore, we set out to explore the development of wing transparency in the glasswing butterfly, *G. oto*, which belongs to a diverse tribe (~393 species) of predominantly transparent neotropical butterflies (Elias et al., 2008). We applied confocal and transmission electron microscopy (TEM) to compare wing development, scale cytoskeletal organization and membrane surface nanostructures between clear and opaque wing regions. Using chemical treatments, scanning electron microscopy and gas chromatography–mass spectrometry, we found that nanostructures on the wing membrane surface are composed of two layers: a lower layer of chitin-based nipple-like nanostructures, and an upper layer of wax-based nanopillars composed predominantly of long-chain n- alkanes. Finally, by removing wax-based nanopillars, we demonstrate their role in dramatically reducing reflection on the wing surface via optical spectroscopy and analytical simulations.

Materials and Methods

Samples

Glasswing butterfly [*Greta oto* (Hewitson 1854)] pre-pupae were purchased from Magic Wings Butterfly House (Deerfield, MA, USA) and reared on Cestrum nocturnum (Solanaceae) leaves at 27°C and 60% humidity on a 16 h:8 h light:dark cycle at the Marine Biological Laboratory (Woods Hole, MA, USA) under United States Department of Agriculture permit number P526P-19-02269. At the appropriate time of development, pupal wings were dissected and age was recorded as hours after pupal case formation (h APF) Dinwiddie et al. (2014). The average timeline from pupation to eclosion (adult emergence) for *G. oto* at 27°C is approximately 7 days, and we report our time series here which covers early aspects of wing scale development.

Optical imaging and scale measurements

Images of whole-mounted specimens were taken with a Canon EOS 70D digital camera with an EF 100 mm f/2.8 L macro lens. High- magnification images of disarticulated wings were taken with a Keyence VHX-5000 digital microscope. Scale density was determined by counting the numbers of scales in a 1 mm^2 area. Scales were also removed from the wings and laid flat onto a slide, and Keyence software was used to measure the surface area of individual scales. Images of clear and opaque regions were processed with Keyence software to measure the percentage of area covered by scales. We took measurements from three individual males and three individual females that were reared in the same cohort. All measurements were taken on the dorsal surface of the forewing (indicated by the red box in Fig. 1F) and each measurement was replicated three times per individual. For statistics, we used N=3, where measurements for each individual were averaged and the difference between each wing measurement group (scale density in clear versus opaque regions and percent wing membrane exposed in clear versus opaque regions) was analyzed using t-tests for two independent samples with unequal variance estimates. An ANOVA test was used to analyze scale area measurements between different scale morphologies (bristle, forked and opaque).

Confocal microscopy

For confocal microscopy of fixed tissue, pupal wings were dissected and fixed in PEM buffer (0.1 mol l^{-1} PIPES, 2 mmol l^{-1} EGTA, 1 mmol l^{-1} MgSO4, pH 6.95) with 3.7% paraformaldehyde for 20– 30 min at room temperature, as described previously (Dinwiddie et al., 2014). Fixed wings were incubated in 1X PBS+0.1% Triton- X 100 (PT) with 1:200 dilution of phalloidin, Alexa 555 conjugated (Invitrogen A34055), and wheat germ agglutinin, Alexa 647 conjugated (Invitrogen W32466) at a dilution of 1:200 overnight at 4°C. Wings were washed in PT and then placed in 50% glycerol: PBS with 1 µg ml^{-1} DAPI overnight at 4°C. Wing samples were placed on microscope slides and mounted in 70% glycerol:PBS. A coverslip (#1.5 thickness) was applied, and each preparation was sealed around the edges with nail polish. Slides of fixed tissue were examined with an LSM 880 confocal microscope (Carl Zeiss, Germany) with 40× and 63× objectives. Confocal images and movies were generated using Imaris Image Analysis Software (Bitplane, Oxford Instruments, UK).

Scanning electron microscopy
We cut 2 mm square pieces from dry wings, coated them with a 10 nm layer of gold using the Bio-Rad E5400 Sputter Coater, and imaged with a Hitachi TM-1000 SEM at 5 kV. Top-view and cross- section SEM images were analysed with ImageJ 1.52 to measure membrane thickness and nanostructure dimensions (n=6 individuals).

Transmission electron microscopy
For TEM, wings of *G. oto* pupae were dissected and fixed in 2% glutaraldehyde, 2% paraformaldehyde in 0.1 mol l^{-1} sodium cacodylate buffer overnight at 4°C (pH 7.4). Samples were then rinsed in 0.1 mol l^{-1} cacodylate buffer (pH 7.4) and post-fixed in 1% aqueous osmium tetroxide in 0.1 mol l^{-1} cacodylic buffer overnight at 4°C, then rinsed in water. Samples were en bloc stained with 1% uranyl acetate in water and then rinsed in water. Samples were dehydrated through a graded ethanol series (50–100% in 10% steps), rinsed in propylene oxide, and then infiltrated in 50% resin and propylene oxide overnight. Samples were infiltrated with Epon/ Alardite embedding medium (70%, 80%, 95% to 100% steps) and polymerized at 60°C for 2 days. Thin sections (~70 nm) were cut on an Ultramicrotome RMC PowerTome XL using a Diatome diamond knife. Digital images were taken using a JEOL 200 transmission electron microscope (JEOL, USA).

Wing surface wax extraction and analysis
To identify the molecular composition of the transparent wing surface, we pooled forewing dissections from three individual adults and performed two replicates for chloroform-based extractions and two replicates for hexane-based extractions (after Futahashi et al., 2019). First, the samples were soaked with 100 µl of either hexane or chloroform and gently mixed for 15 min on a Thermolyne RotoMix 51300. The liquid solutions containing dissolved wing surface compounds were then transferred to glass vials with fixed microvolume inserts, and the solvent was evaporated under a stream of high-purity nitrogen gas (99.99%). Dried extracts were re- dissolved in fixed volumes of hexane (10 µl), and half of the extract (5µl) was injected by automatic liquid sampler into a gas chromatograph coupled with a mass selective detector (GC: 7890A; MS: 5975C; Agilent Technologies, USA) operating in electron impact mode. The injection was performed in a split/ splitless injector in the splitless mode. Separation of compounds was performed on a fused silica capillary column (DB-5MS, 30 m×0.32 mm×0.25 µm, Agilent J&W GC columns, USA) with a temperature program starting from 80°C for 5 min and increasing by 80°C min^{-1} to 200°C, followed by an increase of 5°C min^{-1} to 325°C, which was held for 3 min, with helium used as the carrier gas, positive electron ionization (70 eV), analog to digital (A/D) sampling rate was set at 4, and the scan range was m/z 40.0 to 650.0. Chemical data processing was carried out using the software Enhanced Chemstation (Agilent Technologies). We retained peaks with abundances greater than 0.25% of the total and compounds were identified according to their retention indices, diagnostic ions and mass spectra, which are provided in Table S1. For some peaks, it was not possible to narrow the identity to a single specific compound because (1) some low abundance substances produced poor quality mass spectra, (2) multiple compounds could have produced the observed fragmentation patterns and/or (3) multiple compounds may have co-eluted at the same retention time.

Optical measurements

The wing reflection measurements were performed on a Cary 5000 UV-Vis-NIR spectrophotometer, equipped with a light source of tungsten halogen and an integrating sphere diffuse reflectance accessory (Internal DRA 1800). Wing measurements from the dorsal wing surface were recorded using three different individuals for control treatments (untreated) and three different individuals for hexane treatments with unpolarized light with a spot size of 100 μm for an incident angle of 8 deg to avoid the loss of direct specular reflectance component through the aperture. All measurements were taken in the dark to avoid possible stray illumination from the surrounding environment and we performed two technical replicates for each individual wing. A reference measurement was done with a calibrated commercial white spectralon standard to calculate the relative diffuse reflectance. The reflectance measurements and mean data are available from Dryad (https:// doi.org/10.6078/D1TD7H).

Optical simulations

The total volume fraction of the untreated wing along the height h can be given by:

$$\frac{\pi r_{np}^2}{\sqrt{3}d^2}\,\mathrm{erfc}\left(\frac{h-h_{np}}{\sigma_{np}\sqrt{2}}\right); \quad \text{zone: dorsal wax based nanopillar (1)}$$

$$\frac{\pi}{\sqrt{3}d^2}\left(r_{np}+\left(\frac{d}{2}-r_{np}\right)\left(1-\frac{h}{h_p}\right)\right)^2; \quad \text{zone: dorsal chitin-based nipple array (2)}$$

$$f_{untreated}(h)=1; \quad \text{zone: chitin membrane (3)}$$

$$\frac{\pi}{\sqrt{3}d^2}\left(r_{np}+\left(\frac{d}{2}-r_{np}\right)\left(1-\frac{h}{h_p}\right)\right)^2; \quad \text{zone: ventral chitin-based nipple array (4)}$$

$$\frac{\pi r_{np}^2}{\sqrt{3}d^2}\,\mathrm{erfc}\left(\frac{h-h_{np}}{\sigma_{np}\sqrt{2}}\right); \quad \text{zone: ventral wax based nanopillar (5)}$$

The average distance between two nanostructures is represented as d, conical shaped cuticular nipple nanostructure height as h_p, wax-based irregular nanopillar radius as r_{np}, mean height of the irregular nanopillar distribution as h_{np} and their corresponding variance as σ_{np}. The volume fraction of the treated wing without the irregular nanopillars will be:

$$\frac{\pi}{2\sqrt{3}}\left(1-\frac{h}{h_p}\right); \quad \text{zone: dorsal chitin-based nipple array (6)}$$

$$f_{untreated}(h)=1; \quad \text{zone: chitin membrane (7)}$$

$$\frac{\pi}{2\sqrt{3}}\left(1-\frac{h}{h_p}\right). \quad \text{zone: ventral chitin-based nipple array (8)}$$

After determining the volume fraction, the corresponding refractive index changes along the wing at any height h were calculated using the effective medium theory (EMT) with the Maxwell–Garnett approximation as shown in Fig. 6E (see Fig. S2). EMT pertains to analytical or theoretical modeling that describes the macroscopic properties of subwavelength nanostructured materials, when the nanostructures collectively affect the optical properties. EMT is developed from averaging the multiple values of the constituents that directly make up the nanostructured material including the surrounding media, in this case, chitin, wax and air. The refractive indices of the different materials were considered as n_{air}=1, n_{chitin}=1.56+i0.008 (Vukusic et al., 1999; Narasimhan et al., 2018), and we considered n_{wax}=1.39 (based on Hooper et al., 2006). Therefore, the effective refractive index neff can be calculated for any h using the equations below with the calculated volume fractions, where air volume fraction can be calculated by corresponding $f_{air}=1-f_{wax/chitin}$:

$$n_{wax}^2 \frac{2(1-f_{air})n_{wax}^2 + (1+2f_{air})n_{air}^2}{(2+f_{air})n_{wax}^2 + (1-f_{air})n_{air}^2};$$ zone: dorsal wax based nanopillar (9)

$$n_{chitin}^2 \frac{2(1-f_{air})n_{chitin}^2 + (1+2f_{air})n_{air}^2}{(2+f_{air})n_{chitin}^2 + (1-f_{air})n_{air}^2};$$ zone: dorsal chitin-based nipple array (10)

$$n_{eff}^2 = n_{chitin}^2;$$ zone: chitin membrane (11)

$$n_{chitin}^2 \frac{2(1-f_{air})n_{chitin}^2 + (1+2f_{air})n_{air}^2}{(2+f_{air})n_{chitin}^2 + (1-f_{air})n_{air}^2};$$ zone: ventral chitin-based nipple array (12)

$$n_{wax}^2 \frac{2(1-f_{air})n_{wax}^2 + (1+2f_{air})n_{air}^2}{(2+f_{air})n_{wax}^2 + (1-f_{air})n_{air}^2}.$$ zone: ventral wax based nanopillar (13)

Afterwards, the transfer matrix method (TMM) computed the reflectance from the stratified medium with calculated refractive index profiles as shown in Fig. 6E for the unpolarized condition (taking the average of both s- and p-polarization) at an incident angle of 8 deg (to replicate the experimental condition). The basic formalism of TMM relies on the calculation of thin film reflection and transmission from Maxwell's electromagnetic equations using the boundary conditions. Because of the stack of thin films, the reflectance and transmittance is calculated with a transfer matrix formalism describing the propagation of light from layer to layer. The membrane-only reflection at normal incident light can be directly calculated from Siddique et al. (2016):

$$R(\lambda) = \int_0^\infty \left| \frac{r(1-e^{-2i\delta})}{1-r^2e^{-2i\delta}} \right|^2 \frac{1}{\sigma_m\sqrt{2\pi}} e^{-(h-h_m)^2/2\sigma_m^2} dh, \qquad (14)$$

where membrane thickness is h_m and modulation is σ_m, $\delta=(2\pi n_{chitin}h)/\lambda$ is the phase delay introduced by the membrane thickness of h, and r is the reflection coefficient at the air–chitin boundary governed by Fresnel's equation for a normal incident light, i.e. $r=(1-n_{chitin})/(1+n_{chitin})$.

Results

Scale measurements in clear and opaque wing regions of adult *Greta oto*

We investigated features of scale density, scale morphology and the amount of wing surface exposed in adult *G. oto*. We focused on two adjacent regions on the dorsal surface of the forewing for consistency: a clear region within the discal cell and an opaque region that consists mainly of black scales near the cross-vein (indicated by the red box in Fig. 1F). The clear wing region contained two types of alternating scale morphologies – bristle-like scales and narrow, forked scales – while within the opaque wing region, scale morphologies resembled 'typical' butterfly pigmented scales – flat and ovoid with serrations at the tips (Fig. 1K,L). The mean (±s.d.) density of scales in the adult wing were significantly lower within the clear region, with 98.2±18.1 scales per mm2 in males and 102.3±17.2 in females, compared with the opaque region with 374.3±22.2 scales per mm2 in males and 358.1+19.6 in females (t=−30.9, d.f.=4, P<0.0001 for male sample comparison, t=−21.9, d.f.=4, P<0.0001 for female sample comparison; Fig. 1N). In the clear region, forked scales were significantly smaller in size (498±39 μm2) compared with the bristle-like scales (831±183 μm2), while in the opaque region, scales were the largest (3467±382 μm2) (Fig. 1O). Finally, the amount of exposed wing membrane was significantly different between wing regions, with an average of 81.6±2.7 and 82.2±4.3% of exposed membrane in the clear wing regions of males and females, respectively, compared with 2.6±1.1 and 1.4±0.7% membrane exposed in opaque regions of males and females, respectively (t=78.9423, d.f.=4, P<0.0001 for male sample comparison, t=48.3854, d.f.=4, P<0.0001 for female sample comparison, Fig. 1P).

Morphogenesis and cytoskeletal organization of developing scale cells

To investigate developmental processes of wing and scale development, we performed dissections of *G. oto* pupae at different time points (Fig. 2). As in other species of Lepidoptera, the early pupal wing consisted of a thin bilayer of uniform epithelial tissue and by 16 h APF, numerous epidermal cells had differentiated to produce parallel rows of sensory organ precursor (SOP) cells (the precursors to the scale and socket cells) (Fig. 2B,C). At this early stage of wing development, we observed that the clear wing region harbored a lower density of SOP cells relative to the opaque wing region (Fig. 2B,C). In a 400 μm2 area, the density of SOP cells in the clear region was 65.2±7.0, compared with the density of SOP cells in the opaque region of 169.2±15.7 (t=−10.4629 d.f.=4, P=0.0003, N=3 pupae). We can therefore infer that early into wing development, SOP cell patterning is differentially regulated between clear and opaque regions, which impacts the adult wing scale density and the amount of wing membrane surface exposed in different parts of the wing.

Next, we investigated cellular and cytoskeletal organization during scale growth in clear and opaque wing regions (Fig. 2D–I). We found that general aspects of scale development in *G. oto* follow those previously reported in several butterfly and moth species by Dinwiddie et al. (2014), with some notable distinctions for modified scale growth in the clear wing regions of *G. oto*. By 30 h APF, the SOP cells have divided to produce the scale and socket cells (Fig. 2D,E). The scale cell body lies internally within the wing, while the socket cell associated with each scale cell lies in a more superficial position. Phalloidin staining showed the appearance of small cylindrical scale outgrowths containing F-actin filaments, and *WG*A staining showed outlines of the membrane as the scale outgrowths begin to project and elongate beyond the wing surface. At this stage, budding scales in the clear wing region appeared morphologically similar to the unspecialized opaque scales: roughly elongated balloon-shaped with numerous small actin rods fanning out from the pedicel to the apical tip of the scale (Fig. 2D,E). By 48 h APF, scale cell extensions have grown and elongated (Fig. 2F,G). The actin filaments have reorganized into smaller numbers of thick, regularly spaced bundles along the proximal– distal axis of the scale just under the surface of the cell membrane. Fluorescent staining revealed larger bundles of F-actin in the adwing (facing the wing membrane) side of the scales relative to the abwing side (Movie 1). At this stage, scales in different regions of the wing had taken on dramatically different morphologies. Scales in the clear region had elongated in a vertical orientation and obtained two types of alternating morphologies: short and triangular, or long and bristle-like outgrowths (Fig. 2F). In the opaque region, scales had taken on a round and flattened morphology, with ground scales shorter than the cover scales (Fig. 2G). By 60 h APF, scale projections were even more elongated (Fig. 2H,I). The triangular scales in the clear wing region had proceeded to generate two new branches, which forked and elongated at the tips bidirectionally, while bristle-like scales had elongated and curved (Fig. 2H). In the opaque region, scales were longer, wider and flatter, and had developed serrations at the tips (Fig. 2I).

Ultrastructure analysis of developing bristle, forked and opaque scales

To reveal ultrastructural detail of developing wing scale morphology, we performed TEM on pupal wing tissue of *G. oto* at 48 h APF (Fig. 3). In transverse sections, we could resolve distinct scale morphologies (bristle, forked and opaque) and their associated cytoskeletal elements.

Bristle-like scales in the clear wing regions were circular in cross- sections (Fig. 3A–C). We could also distinguish between distal and basal regions of bristle-like scales, the latter of which had the presence of a surrounding socket cell in the cross-section (Fig. 3B, C). TEM revealed that these bristle-like scales were ringed by peripheral bundles of actin filaments, which lay spaced just under the cell membrane (Fig. 3B,C′). In distal regions of the bristle-like scale, actin bundles were larger on the adwing side relative to the abwing (Fig. 3B), while near the base of the bristle-like scale (indicated by the presence of a surrounding socket cell), actin bundles were more evenly distributed around the periphery (Fig. 3C).

We also observed large populations of microtubules distributed throughout developing scales, which were internal relative to the actin bundles. Interestingly, we

observed distinct patterns of microtubule distribution within different developing scale morphologies. The cross-section of bristle-like scales revealed large populations of internal microtubules, which we identified owing to their characteristic ring shape and diameter of ∼25 nm (Fig. 3B′,C′). The circular ring shape of microtubules in cross-sections of both the basal and distal parts of the bristle-like scale suggested that microtubules are all longitudinally oriented, running in the same direction as the actin filaments, parallel to growth. We also observed that populations of microtubules were localized primarily away from the surface of the scale in its interior, and microtubules were fewer distally than basally (Fig. 3B′,C′).

In our TEM cross-sections, we also observed scale types that appeared more triangular in shape, suggesting that they correspond to developing forked scales within the clear wing region (Fig. 3D,E). These scales were ringed by peripheral bundles of actin filaments, with larger actin bundles on the adwing side of the scale. Interestingly, we observed two internal bundles of actin filaments that were not observed in bristle-like scale morphologies, although we note that these could also be internal actin bundles previously referred in other butterfly species as ‘rods’, which only extend approximately two-thirds of the way along the proximal–distal axis and are only on the lower surface of the scale (Fig. 3E′) (Dinwiddie et al., 2014). We also note that there was variability in microtubule orientation, rather than the ubiquitous longitudinal orientations observed in bristle-like scales.

Finally, developing opaque scales were easily identified in cross-sections owing to their large size and flattened morphology (Fig. 3F,G). We observed peripheral bundles of actin filaments that were widely spaced and smaller in size in distal parts of the scale (Fig. 3G,G′). We observed a clear asymmetry in actin bundle size, which were larger on the adwing side of the scale relative to the abwing surface. In opaque wing regions, TEM micrographs revealed what appeared to be concentrated parallel-running populations of microtubules near the narrow base of the scales, and then a more mesh-like network of microtubules in more distal flattened regions, indicating that microtubules have varying orientations within different regions of the scale (Fig. 3G,G′, Fig. S1). In contrast to the bristle-like scales, large, flattened opaque scales appeared to contain populations of microtubules that were more widely distributed and less dense. In all scale types, we observed the presence of hexagonally packed F-actin filaments and numerous internal organelles and vesicles, including mitochondria, electron-dense vesicles and free ribosomes (Fig. 3, Fig. S1).

Ontogeny of wing membrane nanostructures

The clear wing regions of *G. oto* contain nanopillars that cover the surface of the membrane (Fig. 1I). These nanopillars were previously characterized based on SEM in adult wings, which feature an irregular height distribution and help to generate omnidirectional anti-reflective properties (Siddique et al., 2015). To gain insight into the development of these nanostructures, we examined the surface of the wing membrane epithelial cells with TEM (Fig. 4B–F). At 60 h APF, a perpendicular section through the wing epithelia showed a continuous epithelial lamina (Fig. 4B,C). We observed that the epithelial cells contained microvilli, which appeared as slender linear extensions from the inner margins of the developing cells that insert into electron-dense material (Fig.

4B,C). The surface layer of the epithelia appeared as an extracellular lamellar system, and lamina evaginations appeared in the section as domes distal to the microvillar extensions (Fig. 4C). By 72 h APF, we observed a thin outer layer of the epicuticle that rose above the epidermal cells, and by 120 h APF, we found that this layer above the microvilli contained what appear to be dome-shaped protrusions and thickened cuticle, possibly secreted from regularly spaced microvilli (Fig. 4D,E). Finally, in our TEM cross-section of a fully developed adult wing of *G. oto*, we observed that the membrane surface harbors dome-shaped nanoprotrusions with morphologies similar to those of insect corneal surface nipple arrays (Yoshida et al., 1997; Bernhard, 1962), which we refer to throughout the text now as 'nipple nanostructures', and an upper layer containing pillar-like protrusions, which we refer to as 'nanopillars', that featured a more irregular height distribution (Fig. 4F). These results show early subcellular processes of developing nanopillars within the clear wing region, which arise distal to microvillar extension in epithelial cells.

Topographical organization and biochemical composition of wing surface nanostructures

Based on our electron microscopy results of membrane nanostructures, we investigated the topographical organization and biochemical composition of the adult wing surface. To do so, we treated individual, disarticulated adult *G. oto* wings in two ways: (1) by physically removing wing surface nanostructures by gently pressing and rubbing a wing in between paper and Styrofoam (Yoshida et al., 1997) and (2) by testing the wing surface structures for solubility in organic solvents, including hexane and chloroform to extract lipids (Futahashi et al., 2019). We then performed SEM to compare wing surface topography of untreated and treated wing samples (Fig. 5A–C′). SEM confirmed that the first treatment partially or completely removed nanostructures across the wing membrane surface (Fig. 5B). In a region of partial removal, we could identify smaller, dome-shaped nipple nanostructures underneath the top layer of nanopillars (Fig. 5B′). SEM of the chemically treated wing surface revealed that the upper layer of irregularly sized nanopillars was completely removed, revealing a layer of regularly arranged dome-shaped nipple nanostructures that did not dissolve through chloroform or hexane exposure (Fig. 5C,C′). Therefore, we hypothesized that the upper layer of irregularly sized nanopillars consisted of a secreted wax-based material, which sits above smaller chitin-based nipple nanostructures.

To test this hypothesis, we extracted the surface layer of *G. oto* clear wing regions with either hexane or chloroform and analyzed the chemical composition by gas chromatography–mass spectrometry (GC-MS). We found that the chemical profile generated by both hexane and chloroform extracts yielded similar results (Fig. 5D). In all extracts, we identified two straight-chain alkanes that made up approximately two-thirds of the compounds detected: 41.64±5.75% pentacosane ($C_{25}H_{52}$) and 23.32±5.35% heptacosane ($C_{27}H_{56}$) (Table S1). The remaining compounds were primarily composed of slightly larger methyl-branched alkanes (monomethyl and dimethyl C27, C29 and C31) and esters. Therefore, our results suggest that in *G. oto*, there are two components to wing surface ultrastructure: procuticle-based nipple

nanostructures, and an upper epicuticular layer of irregularly sized nanopillars, composed mainly of straight- chain alkanes (Fig. 5D,E).

Anti-reflective properties of wax-based nanopillars

To address whether the wax-based nanopillars play a role in wing reflection, we measured the reflectance spectra of untreated and hexane-treated wings (Fig. 6). Additionally, we measured nanostructure geometries and membrane thickness from wing SEM cross-sections and determined the average distance between two nanostructures as d=174 nm, conical-shaped cuticular nipple nanostructure height as hp=77 nm, wax-based irregular nanopillar radius as rnp=53 nm, mean height as hnp=224 nm and variance as σnp=49.3 nm, and membrane thickness as hm=746 nm and variance as σm=43nm (Fig. 6B,D, Fig. S2). On the basis of SEM micrographs for treated and untreated samples, we modeled three wing architectures, consisting of: (1) nanopillars with variable height together with cuticle-based nipple nanostructures on the wing membrane, (2) cuticle-based nipple nanostructures on the wing membrane and (3) the wing membrane without any nanostructures, to simulate the optical properties for different conditions (Fig. 6E). The simulated reflectance data of the untreated and treated conditions in Fig. 6F closely resembled the experimental ones. In untreated wings of *G. oto*, we found that transparent regions have a low total diffuse reflection of approximately 2%, which is in line with previous reflectance measurements of this species (Siddique et al., 2015) (Fig. 6F). By contrast, the hexane-treated wings without the upper layer of wax nanopillars had approximately 2.5 times greater reflectance relative to the untreated wings, and generated an iridescent thin film spectra, even though they harbored dome-shaped nipple nanostructures (Fig. 6D,F).

For simulated data, the overall reflectance ratio of the hexane- treated wing to the untreated wing was approximately three, similar to experimental reflectance data (Fig. 6F; see dataset available from Dryad at https://doi.org/10.6078/D1TD7H). Importantly, the simulated results for the untreated wing with wax-based irregular nanopillars make reflectance more uniform across wavelengths, which reduces the iridescent effect of the wing membrane. Finally, we simulated a thin film membrane without any nanostructures, which showed reflectance (averaged from all wavelengths) of the membrane itself to be 8.81±3.46%, whereas the treated and untreated wing reflections were 5.78±2.82% and 1.93±0.77%, respectively (Fig. 6F). While treated wings harboring dome-shaped nipple nanostructures reduced the overall reflectance relative to the membrane only, their effect was not strong enough to reduce reflectance spectra oscillation. The wax-based irregular nanopillars on top introduced a more gradual transition between refractive indices to lessen the oscillation by approximately five-fold, in addition to reducing overall reflection (Fig. 6F). Additionally, we simulated the three wing architecture models considering different mean membrane thicknesses and variance in membrane thickness (Fig. S3). We found that variance in wing membrane thickness reduced reflectance spectra oscillations, rather than mean membrane thickness alone, and more peaks appear in the visible spectrum with increasing thickness of the membrane. (Fig. S3; Dryad dataset https://doi.org/10.6078/D1TD7H). Overall, these results demonstrate that the non-

constant architecture of the wing membrane and wax-based irregular nanopillars on the wing surface of *G. oto* function to dramatically enhance anti-reflective properties.

Solubility of wing surface nanostructures in clearwing Lepidoptera

We investigated additional species of clearwing Lepidoptera by assessing the solubility of wing surface nanostructures with hexane treatments, including (A) an additional glasswing butterfly, *Godyris duilia* (Nymphalidae: Ithomiini), (B) the amber phantom butterfly, *Haetera piera* (Nymphalidae: Haeterini), (C) the longtail glasswing, *Chorinea faunus* (Riodinidae: Riodinini), and (D) the clearwing hawkmoth, *Hemaris thysbe* (Sphingidae: Dilophonotini) (Fig. 7). For both *G. duilia* and *H. piera*, we found that the clear wing membrane surface is covered in irregularly arranged nanopillar structures (Fig. 7A,B). After hexane treatments, the wings became more reflective, the upper layer of irregularly arranged nanopillars was removed, while nipple-like structures remained, supporting that nanopillars are likely wax-based, similar to *G. oto*. Conversely, for both *C. faunus* and *H. thysbe*, the reflectivity of the wings and the regularly arranged nipple array-like nanostructures on the membrane surface appeared unaffected after hexane treatment, suggesting that the structures are chitin-based (Fig. 7C,D). These results indicate that wing surface nanostructures can be either chitin-based, which morphologically resemble the nipple array type of nanostructure, or wax-based, which morphologically resemble irregularly arranged nanopillars, and both types appear to have arisen in phylogenetically distant lineages of Lepidoptera.

Discussion

Butterflies and moths have evolved sub-wavelength anti-reflective structural innovations on their wings that enable them to be transparent. Here, we report the details of pupal wing development and scale cytoskeletal organization in the glasswing butterfly, *G. oto*, as well as insights into the ontogeny and biochemical basis of wing surface nanostructures that reduce reflection in clearwing Lepidoptera.

The arrangement of unicellular projections in insect integument, such as bristles and scales, has been a model for research on cellular pattern formation (Ghiradella and Butler, 2009). Shortly after pupation, SOP cells develop from a monolayer of epithelial cells into orderly arrangements, then differentiate into scale and socket cells. In the present study, we found that early SOP cell patterning affects the final adult scale density in *G. oto*, and this feature of spacing scale cells farther apart, and therefore reducing the overall density of scales, is an initial step to generate clear wings. During early pupal development, the receptor molecule *Notch* is expressed in a grid-like pattern in the wing epithelium (Reed, 2004). This may contribute to the parallel rows of uniformly spaced SOP cells, which express a homolog of the achaete-scute proneural transcription factors that likely function in scale precursor cell differentiation (Galant et al., 1998). *Notch*-mediated lateral inhibition could establish a dense population of ordered SOP cells in the developing wing, resulting in a characteristic ratio of scale-building and epithelial cells (Escudero et al., 2003; Couturier et al., 2019). Future studies should investigate whether modifications in *Notch* signaling play a role in scale

cell patterning in clearwing butterflies and moths, many of which contain reduced densities of scale cells (Gomez et al., 2020 preprint; Pinna et al., 2020 preprint).

The range of morphological diversity among scales and bristles within Lepidoptera likely results developmentally from components or modifiers of the cytoskeletal structures and cell membrane. One study surveyed a wide range of developing butterfly and moth scales and identified that F-actin is required for several aspects of scale development, including scale cell elongation and proper orientation (Dinwiddie et al., 2014). In the developing bristle-like scales in *G. oto*, we find relatively symmetrical actin bundles distributed throughout the periphery and a large population of longitudinally running interior microtubules. This is similar to what has been described for developing bristles in *Drosophila melanogaster* pupae, which contain peripheral bundles of cross-linked actin filaments and a large population of microtubules that run longitudinally along the bristle (Tilney et al., 2000). It was recently shown that actin bundles play different roles in shaping scales and bristles in the mosquito *Aedes aegypti*, in which developing bristles contained symmetrically organized actin bundles, while actin bundle distribution in scales became more asymmetrically organized (Djokic et al., 2020). Given that actin dynamics play a variety of roles in regulating the development of bristles and scales (Dinwiddie et al., 2014; Day et al., 2019; Tilney et al., 2000; Djokic et al., 2020), we hypothesize that modifications in F-actin organization of scales in the transparent wing of *G. oto* are responsible in part for their narrow bristle-like and forked morphologies.

In an analysis of moth scale development, major shape changes were found to be correlated with changes to the orientation of the cytoplasmic microtubules (Overton, 1966). In the present study, we identified large populations of microtubules organized throughout developing scales and found that microtubules exhibit different distributions and orientations relative to distinct scale morphologies, namely between bristle, forked and flat, round scales. In *D. melanogaster*, microtubules may play a role in bristle development by adding bulk to the bristle cytoplasm, contributing to proper axial growth, and aiding organelle and protein distribution (Bitan et al., 2010, 2012). It would be interesting for future studies to functionally characterize the role microtubules play in the development of lepidopteran scales. Our findings lend further support to the observations that general patterns of scale development, including patterns of F-actin localization and microtubule distribution, seem to be well conserved in Lepidoptera, and that modifications of scale morphology to achieve clearwing phenotypes, such as narrow bristle-like and forked scales, likely involve alteration of cytoskeletal organization during scale growth.

Chitinous wing membrane has a higher refractive index than air, which generates glare under natural light conditions. Some clearwing species have evolved sub-wavelength anti-reflective nanostructures, which reduces glare and likely aids in crypsis (Yoshida et al., 1997; Siddique et al., 2015). In this study, we identified the early developmental processes of nanostructures that arise in the wing epithelium. We also note interesting parallels of our observations to previous descriptions of developing nanostructures on the surface of insect cornea. Early data on pupal development of corneal nanostructures were produced by detailed electron microscopy studies, showing that corneal nipples emerge during lens formation (Gemne, 1971; Fröhlich, 2001). In these observations, development of initial laminar patches formed on top of

underlying microvilli. Subsequently, nanostructures (termed nipple structure array) formed on the surface, with the tips of microvilli still attached to the inner surface. Gemne (1971) proposed that the corneal nanostructures originate from secretion by the regularly spaced microvilli of the cone lens cells, although there is still debate about the exact nature of how microvilli pre-pattern nanostructure arrays (Kryuchkov et al., 2017). Our TEM results provide insight into the early developmental processes of anti-reflective nanostructure formation in the wings of *G. oto*, highlighting certain similarities to nipple array development in insect cornea. It would be interesting for future work to explore whether features of nanostructure formation arose independently in insect cuticle as a mechanism to reduce surface reflection.

In contrast to previously described highly ordered nipple arrays found on insect eyes and some clearwing lepidopteran wings (Stavenga et al., 2006; Kryuchkov et al., 2017), the irregularly sized anti-reflective nanopillars in the clear regions of *G. oto* wings appear to consist of an upper layer of wax-based epicuticle sitting above procuticle-based nipple nanostructures. Insect cuticle is an extracellular matrix formed by the epidermis and is composed of three layers: the outermost envelope, the middle epicuticle and the inner procuticle (Moussian, 2010). The envelope and the epicuticle are composed mainly of lipids and proteins, while the procuticle contains the polysaccharide chitin. Many terrestrial arthropods deposit a layer of wax lipids on the surface of their cuticle, which reduces evaporative water loss (Gibbs, 1998). In some species of dragonfly, epicuticular wax-based nanostructures have also been demonstrated to play a role in generating optical properties, such as an ultraviolet reflection (Futahashi et al., 2019). In mature males of these dragonflies, a dense wax secretion composed of long-chain methyl ketones, in particular 2-pentacosanone, was found to contribute to the UV reflection properties (Futahashi et al., 2019). The chemical composition of nanopillars on the wing surface of cicadas, which contribute to hydrophobicity and antimicrobial properties, was found to consist of epicuticular components such as fatty acids and hydrocarbons ranging from C17 to C44(Romań-Kustas et al., 2020). Another study exploring the molecular organization of dragonfly wing epicuticle found that the major components identified were fatty acids and n-alkanes with even- numbered carbon chains ranging from C14 to C30 (Ivanova et al., 2013). Here, we identified that the epicuticular layer of irregularly sized anti-reflective nanopillars in *G. oto* appears to be composed mainly of n-alkanes, including pentacosane (C25) and heptacosane (C27) and showed the importance of these structures in attaining better transparency. Interestingly, we found that butterflies belonging to the tribe *Haeterini* also contain irregularly arranged hexane-soluble nanopillars on the wing membrane surface, suggesting that wax-based anti-reflective structures have arisen multiple times independently.

Turing reaction–diffusion mechanisms have been proposed as a model for the formation of various corneal nanostructure morphologies (such as spacing, height, and spatial organization) during insect eye development (reviewed in Kryuchkov et al., 2017). Although the degree of height irregularity of nanopillars is important for achieving omnidirectional anti-reflection in *G. oto*, we do not yet understand how such variability in height is generated. Perhaps the pressure of the wax secretion varies across the area of microvillar extensions, similar to how nozzle area plays a role in the propulsion force, and tunes the height of the nanopillars in the process. In such a scenario, the degree of

the height variation could be synthetically engineered depending on the two-dimensional nanopatterned mask design in the biomimetic processes, such as molding or imprinting techniques. Additionally, others have generated three-dimensional wax structures using n-alkanes, noting that wax-based crystals can generate different shapes, sizes and densities depending on the chain length (Gorb et al., 2014). Future work should investigate the possible role of alkanes, and the two-dimensional surface growth geometry, in generating three- dimensional anti-reflective nanostructures and potential applications for biomimetics.

Taken together, these results enable us to form a hypothesis that the origin of anti-reflective nanopillars may have involved a two-step evolutionary process. First, regions of wing membrane may have become increasingly exposed through a reversion of dense, flat, wing scales to fewer, narrow more bristle-like scales. Next, membrane surface nanostructures may have arisen and reduced surface reflection, which became an advantageous phenotype owing to enhanced crypsis and reduced predation. Interestingly, some basal ithomiines contain nanostructures on the membrane surface, despite having opaque wings (C.P., unpublished observations). Wing surface nanostructures are also known to provide antibacterial and hydrophobicity properties in insects, which may explain why they are present in some opaque species. This presents an interesting question of whether wing surface nanostructures in clearwings were already present in an opaque ancestor and were selected for anti-reflective properties, or whether they arose *de novo*. In either scenario, this potential two-step evolutionary process may have required different sets of developmental programs or gene networks that co-occurred to generate wing transparency. Future studies of scale and nanostructure development and evolutionary histories of transparent species and their opaque ancestors will help to elucidate how transparency repeatedly arose in Lepidoptera. Our exploration of *G. oto* wing development can serve as a model for understanding how transparent phenotypes evolved within Ithomiini, a diverse tribe of neotropical butterflies that act as mimicry models for numerous species of Lepidoptera Elias et al. (2008), as well as more distantly related butterfly and moth species.

Acknowledgments

We would like to thank the Angie Serrano, Paola Betucci, Idoia Quintana-Urzainqui and Helena Bilandzija from the MBL Embryology Course and Cao Lu Yan from the MBL Physiology Course for their preliminary work on scale morphology of clearwing Lepidoptera, and additional work by Jaap van Krugten, Raymundo Picos, and Johnny On at UC Berkeley. We also thank Rachel Thayer, Arnaud Martin, Damien Gailly, Melanie Mcclure, Luca Livraghi, Oscar Paneso, Rémi Mauxion and Owen McMillan for their assistance with fieldwork, rearing, and preliminary experiments. We thank Fred Gagnon of Magic Wings Butterfly Conservatory and Gardens for assistance with butterfly rearing and Neil Tsutsui for support with the GC-MS experiments. Members of the Patel Lab, Craig Miller, and Noah Whiteman provided helpful feedback on the manuscript. RHS acknowledges the support from the Beckman Institute of the California Institute of Technology to the Molecular Materials Research Center.

Competing interests

The authors declare they have no competing interests.

Author contributions

Conceptualization: A.F.P., N.H.P.; Methodology: A.F.P., R.H.S., E.I.C., K.H., D.G., M.E., N.H.P.; Formal analysis: A.F.P., R.H.S., E.I.C.; Investigation: A.F.P., R.H.S., E.I.C., Y.K., C.P., K.H.; Resources: A.F.P., R.H.S., E.I.C., K.H., N.H.P.; Writing - original draft: A.F.P.; Writing - review & editing: A.F.P., R.H.S., E.I.C., Y.K., C.P., D.G., M.E., N.H.P.; Visualization: A.F.P.; Supervision: N.H.P.; Funding acquisition: M.E., N.H.P.

Funding

This work was supported by a grant from the Human Frontier Science Program (RGP0014/2016), a France-Berkeley fund grant (FBF 2015-58) and an ANR grant (CLEARWING project, ANR-16-CE02-0012). Open access funding provided by UC Berkeley. Deposited in PMC for immediate release.

Data availability

Spectrometry data on *Greta oto* untreated and hexane treated clear wing regions and simulated reflectance spectra have been deposited in Dryad (Pomerantz, 2021): https://doi.org/10.6078/D1TD7H

References

1. Arias M, Elias M, Andraud C, Berthier S, Gomez D. Transparency improves concealment in cryptically coloured moths. Journal of Evolutionary Biology. 2020 Feb;33(2):247-52.
2. Arias M, Mappes J, Desbois C, Gordon S, McClure M, Elias M, Nokelainen O, Gomez D. Transparency reduces predator detection in mimetic clearwing butterflies. Functional Ecology. 2019 Jun;33(6):1110-9.
3. Bagge LE. Not as clear as it may appear: Challenges associated with transparent camouflage in the ocean. Integrative and Comparative Biology. 2019 Dec 1;59(6):1653-63.
4. Beldade P, Brakefield PM. The genetics and evo–devo of butterfly wing patterns. Nature Reviews Genetics. 2002 Jun;3(6):442-52.
5. Bernhard CG. A corneal nipple pattern in insect compound eyes. Acta physiol. scand. 1962;56:385-6.
6. Binetti VR, Schiffman JD, Leaffer OD, Spanier JE, Schauer CL. The natural transparency and piezoelectric response of the *Greta oto* butterfly wing. Integrative Biology. 2009 Apr 1;1(4):324-9.
7. Bitan A, Guild GM, Bar-Dubin D, Abdu U. Asymmetric microtubule function is an essential requirement for polarized organization of the *Drosophila* bristle. Molecular and cellular biology. 2010 Jan 15;30(2):496-507.

8. Bitan A, Rosenbaum I, Abdu U. Stable and dynamic microtubules coordinately determine and maintain *Drosophila* bristle shape. Development. 2012 Jun 1;139(11):1987-96.
9. Couturier L, Mazouni K, Corson F, Schweisguth F. Regulation of *Notch* output dynamics via specific E (spl)-HLH factors during bristle patterning in *Drosophila*. Nature communications. 2019 Aug 2;10(1):1-3.
10. Day CR, Hanly JJ, Ren A, Martin A. Sub-micrometer insights into the cytoskeletal dynamics and ultrastructural diversity of butterfly wing scales. Developmental Dynamics. 2019 Aug;248(8):657-70.
11. Deparis O, Mouchet SR, Dellieu L, Colomer JF, Sarrazin M. Nanostructured surfaces: bioinspiration for transparency, coloration and wettability. Mater. Today Proc. S. 2014 Jan 1;1:122-9.
12. Dinwiddie A, Null R, Pizzano M, Chuong L, Krup AL, Tan HE, Patel NH. Dynamics of F-actin prefigure the structure of butterfly wing scales. Developmental biology. 2014 Aug 15;392(2):404-18.
13. Djokic, S., Bakhrat, A., Tsurim, I., Urakova, N., Rasgon, J. L. and Abdu, U. (2020). Actin bundles play a different role in shaping scales compared to bristles in the mosquito Aedes aegypti. Sci. Rep. 10, 1-5. doi:10.1038/s41598-020-71911-0
14. Elias M, Gompert Z, Jiggins C, Willmott K. Mutualistic interactions drive ecological niche convergence in a diverse butterfly community. PLoS Biol. 2008 Dec 2;6(12):e300.
15. Escudero LM, Wei SY, Chiu WH, Modolell J, Hsu JC. Echinoid synergizes with the *Notch* signaling pathway in *Drosophila* mesothorax bristle patterning. Development. 2003 Dec 22;130(25):6305-16.
16. Fröhlich A. A scanning electron-microscopic study of apical contacts in the eye during postembryonic development of *Drosophila melanogaster*. Cell and tissue research. 2001 Jan 1;303(1):117-28.
17. Futahashi R, Yamahama Y, Kawaguchi M, Mori N, Ishii D, Okude G, Hirai Y, Kawahara-Miki R, Yoshitake K, Yajima S, Hariyama T. Molecular basis of wax-based color change and UV reflection in dragonflies. Elife. 2019 Jan 15;8:e43045.
18. Galant R, Skeath JB, Paddock S, Lewis DL, Carroll SB. Expression pattern of a butterfly achaete-scute homolog reveals the homology of butterfly wing scales and insect sensory bristles. Current Biology. 1998 Jul 2;8(14):807-13.
19. Gemne G. Ontogenesis of corneal surface ultrastructure in nocturnal Lepidoptera. Philosophical Transactions of the Royal Society of London. B, Biological Sciences. 1971 Aug 20;262(843):343-63.
20. Ghiradella HT, Butler MW. Many variations on a few themes: a broader look at development of iridescent scales (and feathers). Journal of the Royal Society Interface. 2009 Apr 6;6(suppl_2):S243-51.
21. Gibbs AG. Water-proofing properties of cuticular lipids. American Zoologist. 1998 Jun 1;38(3):471-82.
22. Gilbert LE, Singer MC. Butterfly ecology. Annual Review of Ecology and Systematics. 1975 Jan 1:365-97.

23. Gomez D, Pinna C, Pairraire J, Arias M, Barbut J, Pomerantz A, Nous C, de Marcillac WD, Berthier S, Patel N, Andraud C. Transparency in butterflies and moths: structural diversity, optical properties and ecological relevance. bioRxiv [preprint]. 2020 May 15. doi: https://doi.org/10.1101/2020.05.14.093450.
24. Goodwyn PP, Maezono Y, Hosoda N, Fujisaki K. Waterproof and translucent wings at the same time: problems and solutions in butterflies. Naturwissenschaften. 2009 Jul 1;96(7):781-7.
25. Gorb E, Böhm S, Jacky N, Maier LP, Dening K, Pechook S, Pokroy B, Gorb S. Insect attachment on crystalline bioinspired wax surfaces formed by alkanes of varying chain lengths. Beilstein Journal of Nanotechnology. 2014 Jul 14;5(1):1031-41.
26. Hooper IR, Vukusic P, Wootton RJ. Detailed optical study of the transparent wing membranes of the dragonfly Aeshna cyanea. Optics express. 2006 May 29;14(11):4891-7.
27. Huang YF, Jen YJ, Chen LC, Chen KH, Chattopadhyay S. Design for approaching cicada-wing reflectance in low-and high-index biomimetic nanostructures. ACS nano. 2015 Jan 27;9(1):301-11.
28. Ivanova EP, Nguyen SH, Webb HK, Hasan J, Truong VK, Lamb RN, Duan X, Tobin MJ, Mahon PJ, Crawford RJ. Molecular organization of the nanoscale surface structures of the dragonfly Hemianax papuensis wing epicuticle. PLoS One. 2013 Jul 9;8(7):e67893.
29. Johnsen S. Hidden in plain sight: the ecology and physiology of organismal transparency. The Biological Bulletin. 2001 Dec;201(3):301-18.
30. Johnsen S. Hide and seek in the open sea: pelagic camouflage and visual countermeasures. Annual review of marine science. 2014 Jan 3;6:369-92.
31. Kryuchkov M, Blagodatski A, Cherepanov V, Katanaev VL. Arthropod corneal nanocoatings: diversity, mechanisms, and functions. InFunctional Surfaces in Biology III 2017 (pp. 29-52). Springer, Cham.
32. Mcclure M, Clerc C, Desbois C, Meichanetzoglou A, Cau M, Bastin-Héline L, Bacigalupo J, Houssin C, Pinna C, Nay B, Llaurens V. Why has transparency evolved in aposematic butterflies? Insights from the largest radiation of aposematic butterflies, the Ithomiini. Proceedings of the Royal Society B. 2019 Apr 24;286(1901):20182769.
33. Moussian B. Recent advances in understanding mechanisms of insect cuticle differentiation. Insect biochemistry and molecular biology. 2010 May 1;40(5):363-75.
34. Narasimhan V, Siddique RH, Lee JO, Kumar S, Ndjamen B, Du J, Hong N, Sretavan D, Choo H. Multifunctional biophotonic nanostructures inspired by the longtail glasswing butterfly for medical devices. Nature nanotechnology. 2018 Jun;13(6):512-9.
35. Overton J. Microtubules and microfibrils in morphogenesis of the scale cells of Ephestia kuhniella. The Journal of cell biology. 1966 May 1;29(2):293-305.
36. Pinna CS, Vilbert M, Borenztajn S, de Marcillac WD, Piron-Prunier F, Pomerantz AF, Patel NH, Berthier S, Andraud C, Gomez D, Elias M. Convergence in light transmission properties of transparent wing areas in clearwing mimetic butterflies. bioRxiv. 2020 July 2. doi: https://doi.org/10.1101/2020.06.30.180612.

37. Pomerantz, A. (2021). Supplementary table from: Spectrometry of *Greta oto* untreated and hexane treated clear wing regions and simulated reflectance spectra. Dryad. Dataset. doi:10.6078/D1TD7H
38. Prum RO, Quinn T, Torres RH. Anatomically diverse butterfly scales all produce structural colours by coherent scattering. Journal of Experimental Biology. 2006 Feb 15;209(4):748-65.
39. Reed RD. Evidence for *Notch*-mediated lateral inhibition in organizing butterfly wing scales. Development Genes and Evolution. 2004 Jan 1;214(1):43-6.
40. Román-Kustas J, Hoffman JB, Reed JH, Gonsalves AE, Oh J, Li L, Hong S, Jo KD, Dana CE, Miljkovic N, Cropek DM. Molecular and Topographical Organization: Influence on Cicada Wing Wettability and Bactericidal Properties. Advanced Materials Interfaces. 2020 May;7(10):2000112.
41. Siddique RH, Gomard G, Hölscher H. The role of random nanostructures for the omnidirectional anti-reflection properties of the glasswing butterfly. Nature communications. 2015 Apr 22;6(1):1-8.
42. Siddique RH, Vignolini S, Bartels C, Wacker I, Hölscher H. Colour formation on the wings of the butterfly Hypolimnas salmacis by scale stacking. Scientific reports. 2016 Nov 2;6:36204.
43. Stavenga DG, Foletti S, Palasantzas G, Arikawa K. Light on the moth-eye corneal nipple array of butterflies. Proceedings of the Royal Society B: Biological Sciences. 2006 Mar 22;273(1587):661-7.
44. Stavenga DG, Leertouwer HL, Wilts BD. Coloration principles of nymphaline butterflies–thin films, melanin, ommochromes and wing scale stacking. Journal of Experimental Biology. 2014 Jun 15;217(12):2171-80.
45. Stavenga DG. Thin film and multilayer optics cause structural colors of many insects and birds. Materials Today: Proceedings. 2014 Jan 1;1:109-21.
46. Thayer RC, Allen FI, Patel NH. Structural color in Junonia butterflies evolves by tuning scale lamina thickness. Elife. 2020 Apr 7;9:e52187.
47. Tilney LG, Connelly PS, Vranich KA, Shaw MK, Guild GM. Actin filaments and microtubules play different roles during bristle elongation in *Drosophila*. Journal of Cell Science. 2000 Apr 1;113(7):1255-65.
48. Vukusic P, Sambles JR, Lawrence CR, Wootton RJ. Quantified interference and diffraction in single Morpho butterfly scales. Proceedings of the Royal Society of London. Series B: Biological Sciences. 1999 Jul 22;266(1427):1403-11.
49. Yoshida A, Motoyama M, Kosaku A, Miyamoto K. Antireflective nanoprotuberance array in the transparent wing of a hawkmoth, Cephonodes hylas. Zoological science. 1997 Oct;14(5):737-41.

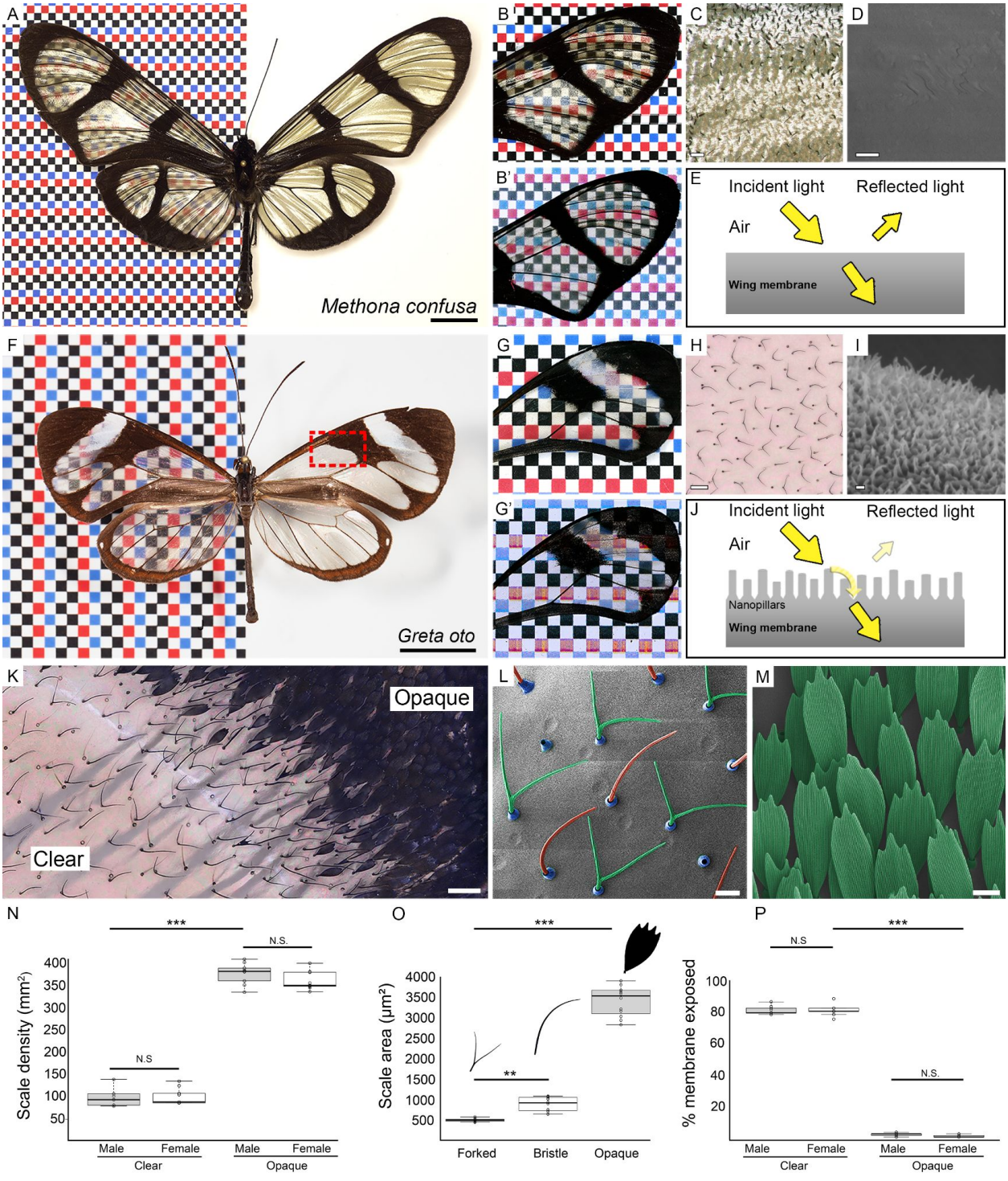


Figure 2.1. Examples of clearwing butterflies and wing scale features in *Greta oto.*

(A) Giant glasswing *Methona confusa* (Nymphalidae: Ithomiini). Scale bar, 1 cm. Wings under (B) reflected and (B′) transmitted light, illustrating general transparency, but strong light reflectance off the wing surface. (C) High magnification of the clear wing

region, showing reflection off the membrane surface. Scale bar, 100 μm. (D) Scanning electron microscopy (SEM) of the wing membrane demonstrates that the surface is smooth and devoid of nanostructures. Scale bar, 1 μm. (E) Simplified diagram of reflection and transmission on the smooth wing membrane of *M. confusa*. Owing to the higher refractive index of the wing membrane, light is reflected at the surface. (F) Glasswing (Nymphalidae: Ithomiini). The red box indicates the representative clear and opaque dorsal forewing regions investigated in this study. Scale bar, 1 cm. Wings under (G) reflected and (G′) transmitted light and (H) high magnification of the clear wing region, illustrating minimal reflectance. Scale bar, 100 μm. (I) SEM of the wing membrane surface reveals irregularly sized nanopillars that enable omnidirectional anti-reflective properties (Siddique et al., 2015). Scale bar, 200 nm. (J) Simplified diagram of reflection and transmission on the wing of *G. oto* containing wing surface nanostructures, which reduce reflection by creating a smoother gradient of refractive indices between air and chitin. (K) High magnification of a transition boundary between a clear and opaque wing region. Scale bar, 100 μm. (L) SEM of adult scales in a clear wing region of *G. oto*, revealing alternating forked (green false coloring) and bristle-like (red false coloring) scale morphologies (socket false colored in blue). Scale bar, 20 μm. (M) SEM of scales in an opaque wing region, highlighting typical large, flat scale morphologies. Scale bar, 20 μm. (N) Measurements of scale density in clear and opaque wing regions, (O) scale surface area for forked, bristle-like, and opaque scale morphologies, and (P) percent of wing membrane exposed in *G. oto* clear and opaque regions. Error bars indicate means+s.d. of three measurements taken from wings in three different individuals (P-values based on t-tests for N and O, and ANOVA for P; ***P<0.001; **P<0.01).

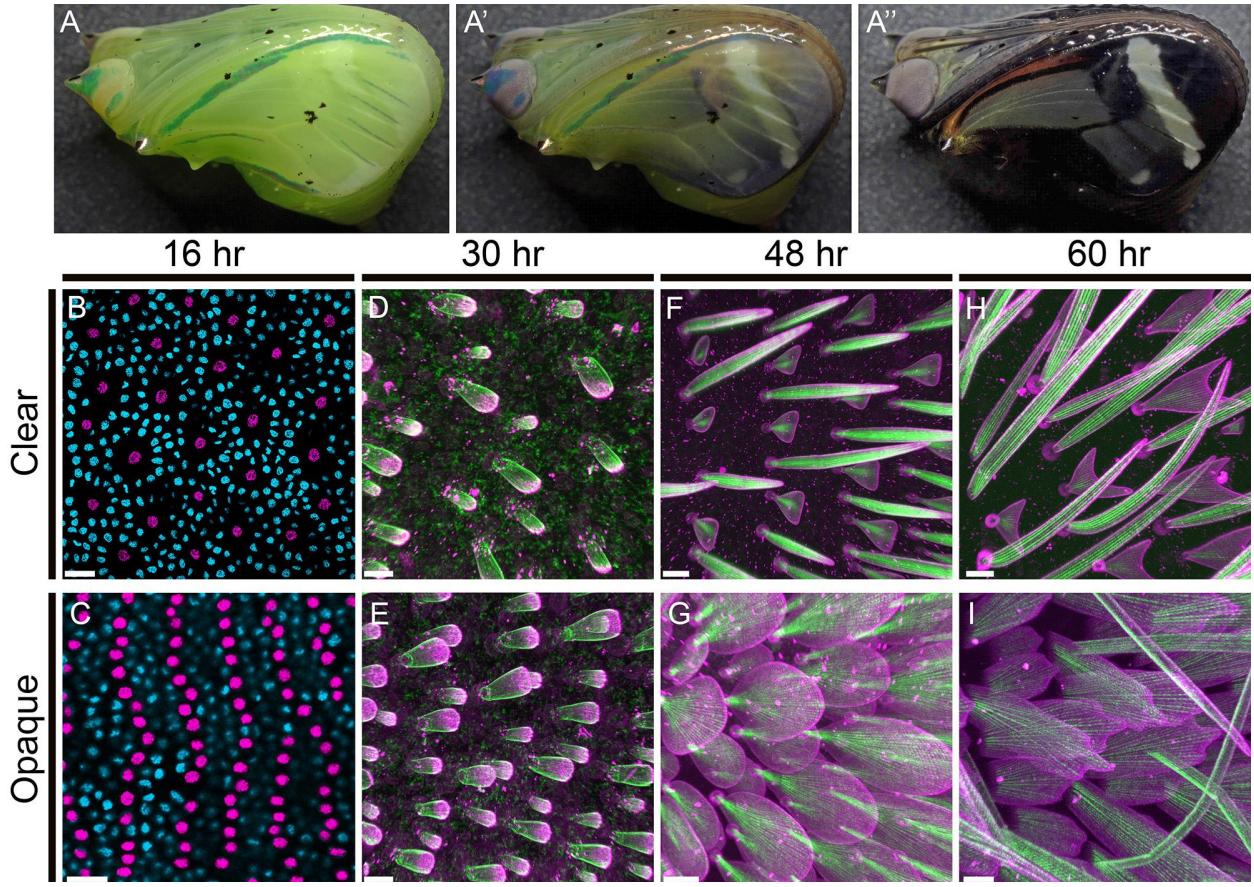


Figure 2.2. Pupal wing development and cytoskeletal organization of scales in clear and opaque regions.

(A) Representative image of a *G. oto* pupa ~5 days after pupal formation (APF), (A′,A″) developing up to the melanic stage ~7 days APF, just prior to eclosion. (B,C) Early wing development 16 h APF stained with DAPI (nuclei) in (B) a clear wing region and (C) an opaque wing region. The clear region contains a reduced number of sensory organ precursor (SOP) cells (the precursor cells to the scale and socket cells) relative to the opaque region. Scale bars, 20 μm. SOP cells are false-colored magenta for better viewing. (D–I) Fluorescently labeled scale cell membrane (wheat germ agglutinin; *WGA*, magenta) and F-actin (phalloidin, green), comparing clear wing regions (D,F,H) to opaque wing regions (E,G,I). (D,E) At 30 h AFP, *WGA* and phalloidin staining reveal early scale buds extending from the wing epithelium and loosely organized parallel actin filaments. (F,G) At 48 h APF, scales have grown and changed in morphology. Short actin filaments have reorganized and formed smaller numbers of thick, regularly spaced parallel bundles under the cell membrane surface. (F) In the clear wing region, scale cells alternate between triangular shapes and bristles. (H,I) At 60 h APF, developing scales have become more elongated. (H) The triangular-shaped scales in the clear wing region have proceeded to generate two new branches, which fork and elongate bidirectionally. (I) In the opaque region, scales are longer and have developed serrations at the tips. Scale bars, (D–I) 10 μm.

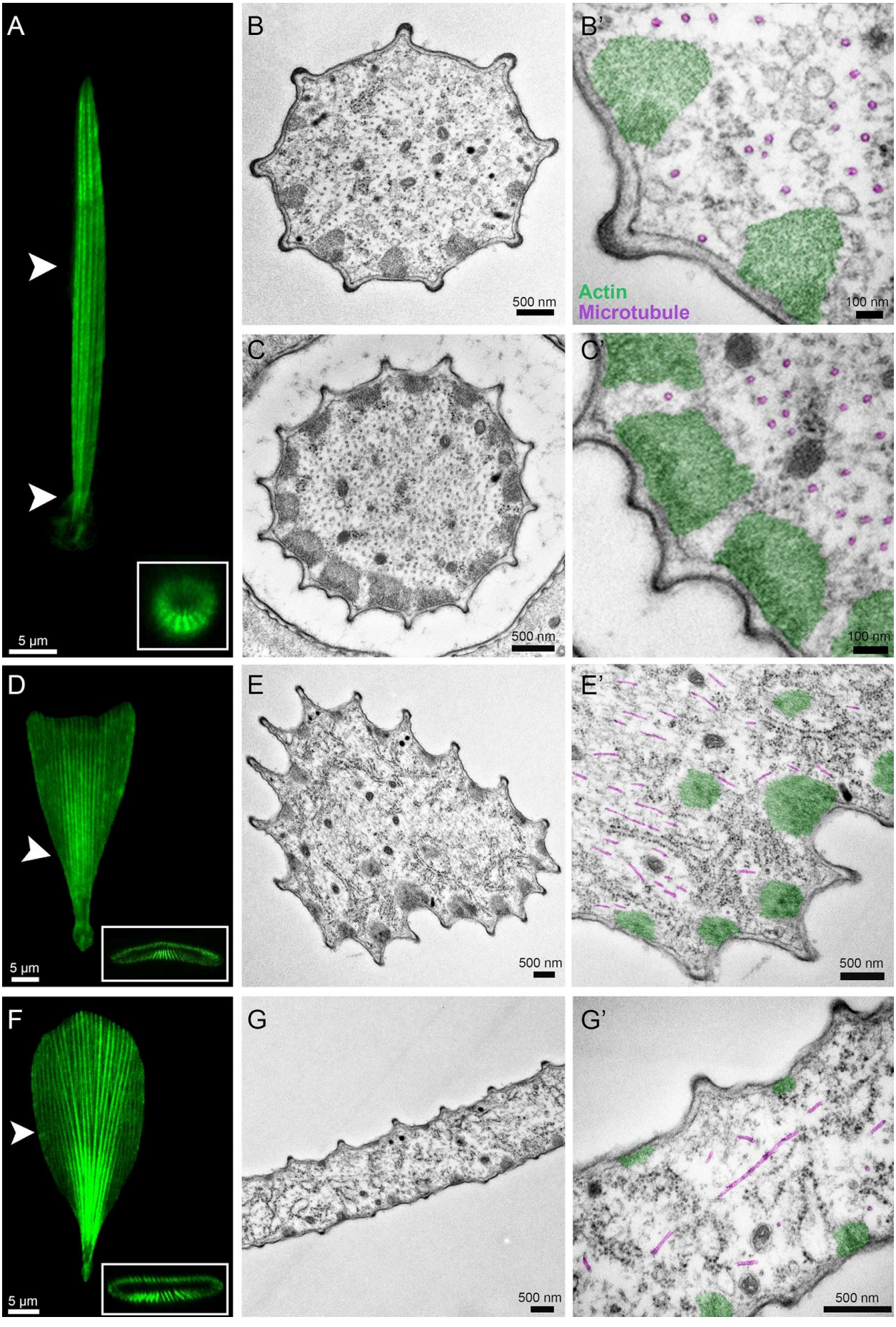

A
5 µm
B
500 nm
B'
Actin
Microtubule
100 nm
C
500 nm
C'
100 nm
D
5 µm
E
500 nm
E'
500 nm
F
5 µm
G
500 nm
G'
500 nm

Figure 2.3. Confocal and transmission electron microscopy (TEM) transverse sections of developing bristle (top), forked (middle) and flat (bottom) scales 48 h APF in *G. oto*.

(A) Confocal projection of a bristle-like scale morphology (phalloidin) in a clear wing region. White arrowheads show representative regions of transverse TEM sections, one near a distal region of the bristle-like scale, and one near the base of the bristle-like scale, which correspond to B and C, respectively. Scale bar, 5 μm. (B,C) TEM of a bristle-like scale in a distal region (B,B′) and a basal region near the socket cell (C,C′). Note the peripheral actin bundles (false-colored green) and internal microtubule rings (false-colored magenta). Scale bars, (B,C) 500 nm, (B′,C′) 100 nm. (D) Confocal projection of a developing forked scale (phalloidin) in a clear wing region. White arrowhead shows a representative region of transverse TEM sections. Scale bar, 5 μm. (E,E′) TEM of a forked scale reveals peripheral bundles of actin (false-colored green), with thicker actin bundles on the ventral side of the scale and internal microtubules (false-colored magenta). Two internal bundles of actin filaments can be observed in the cytoplasm (E′). Scale bars, 500 nm. (F) Confocal projections of developing flat, round scale (phalloidin) in an opaque wing region. White arrowhead shows a representative region of transverse TEM sections. Scale bar, 5 μm. (G,G′) TEM reveals asymmetry in the actin bundles (false-colored green), which are larger on the bottom side of the scale relative to the upper surface. Microtubules (false-colored magenta) are found in various orientations. Scale bars, 500 nm. The insets in A, D and F indicate confocal projections of the scales stained with phalloidin rotated horizontally.

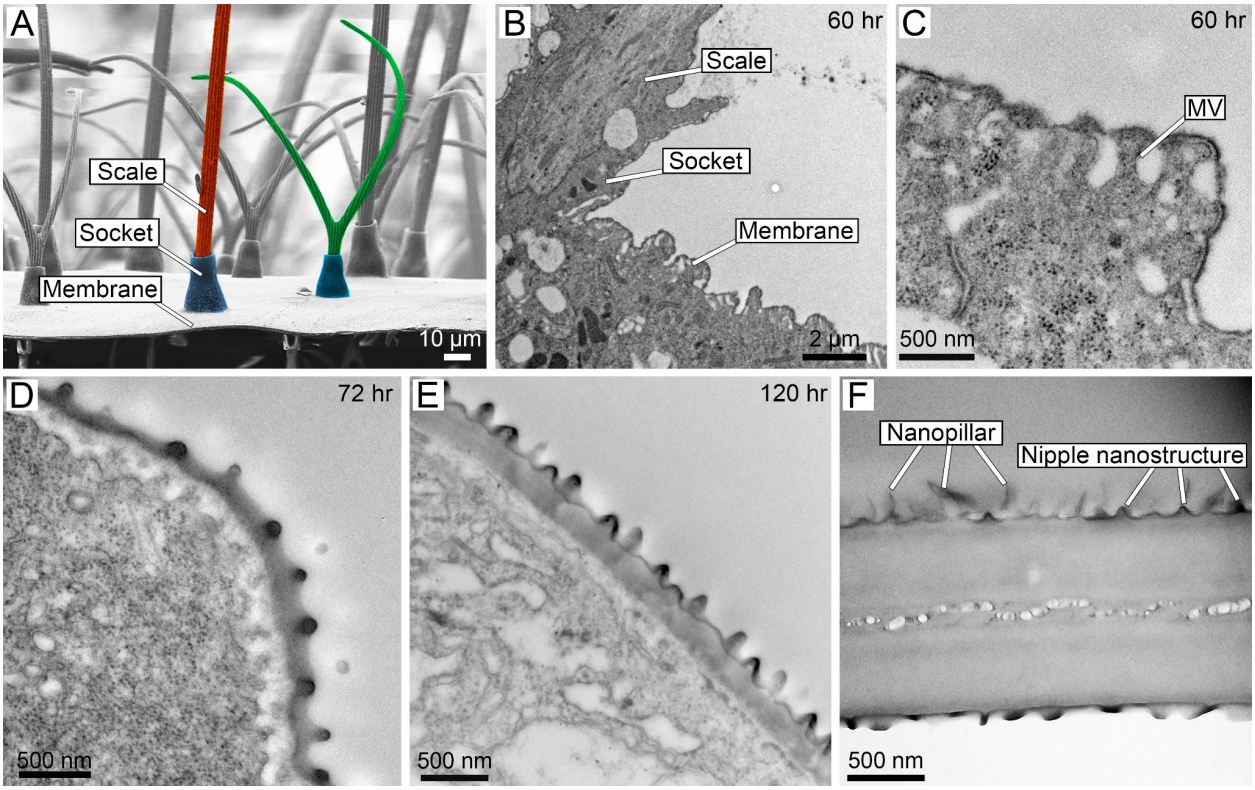


Figure 2.4. Ontogeny of wing membrane surface nanostructures.

(A) SEM cross-section (side view) of an adult *G. oto* clear wing region. Scale bar, 10 µm. Bristle-like scale false colored in red, forked scale false colored in green, sockets false colored in blue. (B) TEM transverse section of epithelial tissue 60 h APF, showing lateral scale growth and wing membrane cells. Scale bar, 2 µm. (C) Higher magnification of developing wing epithelial cells at 60 h APF show microvilli (MV) projections, which appear as slender linear extensions from the inner margins of the developing cells that insert into a thin layer of electron-dense material. Lamina evaginations appear in the section as domes. (D,E) TEM of epithelial tissue (D) 72 h APF and (E) 120 h APF shows wing surface nanostructures protruding from the surface, with tips of microvilli still attached to the inner surface of the wing membrane. (F) TEM of the adult wing membrane. The surface contains dome-shaped nipple nanostructures and an upper layer of nanopillars. Scale bars, (C–E) 500 nm.

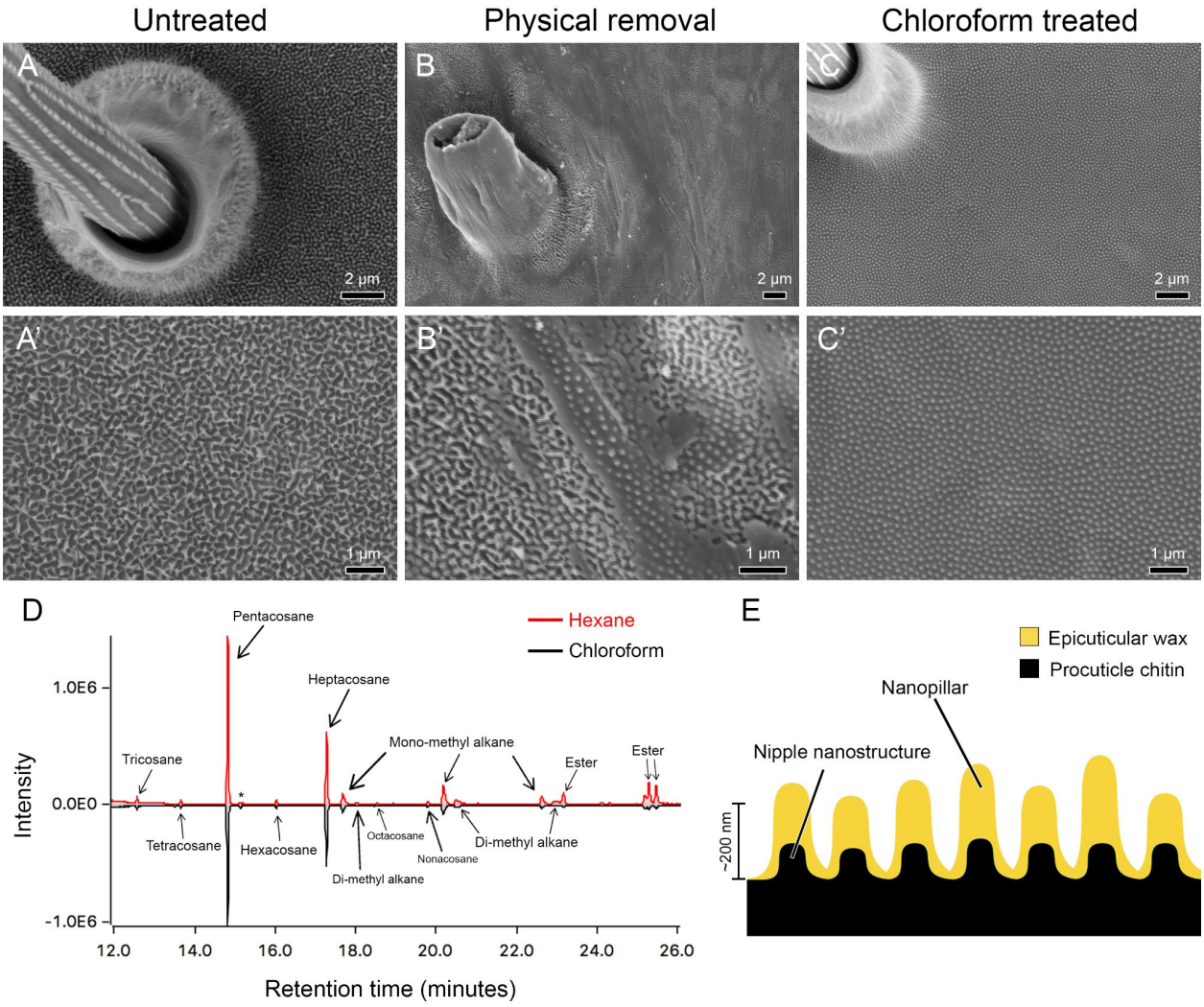


Figure 2.5. Topographical organization and biochemical composition of wing surface nanostructures.

SEM of the transparent wing membrane surface of *G. oto* under (A,A′) the untreated condition, highlighting the presence of irregularly arranged nanopillar structures covering the surface, (B,B′) the physical treated condition, revealing partial removal of surface nanopillars, and a lower layer of more regularly arranged nipple-like nanostructures and (C) the chloroform-treated condition, revealing complete removal of the upper layer of nanopillars, and remaining lower layer of nipple-like nanostructures. Scale bars, (A–C) 2 µm, (A′–C′) 1 µm. (D) Chromatogram of hexane-treated (top; red line) and chloroform-treated (bottom; black line) clearwing extracts. x-axis shows the retention time in minutes and y-axis shows the abundance of total ion current. (E) Schematic of proposed wing surface membrane nanostructures in *G. oto*, composed of chitin-based procuticle and wax-based epicuticle.

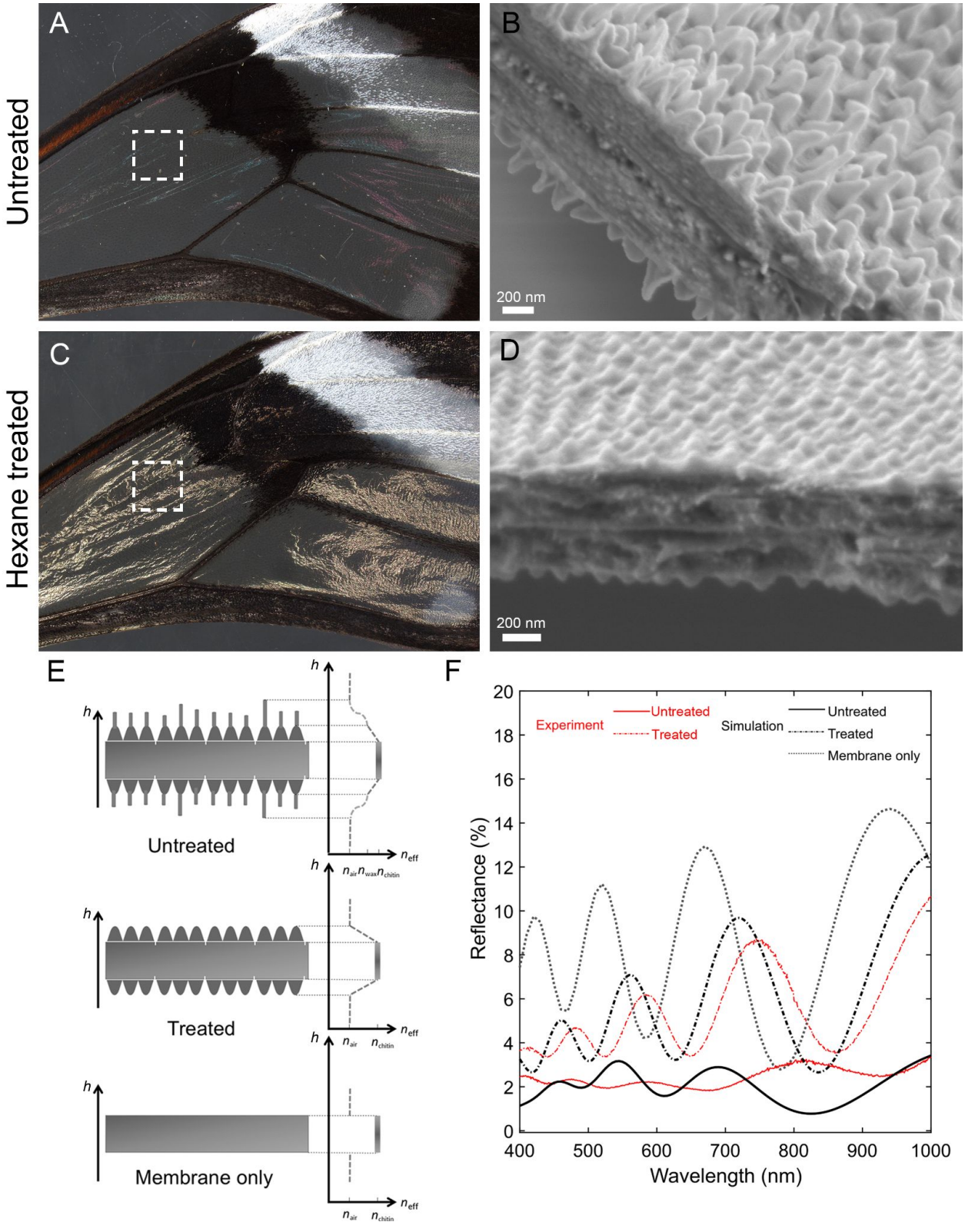

A
B
C
D
Untreated
Hexane treated
200 nm
200 nm
E
h
Untreated
Treated
Membrane only
n_{eff}
n_{air} n_{wax} n_{chitin}
n_{air} n_{chitin}
F
Experiment
Untreated
Treated
Simulation
Untreated
Treated
Membrane only
Reflectance (%)
Wavelength (nm)
0
2
4
6
8
10
12
14
16
18
20
400
500
600
700
800
900
1000

Figure 2.6. Structural elements, reflectance spectra and optical modeling of anti-reflective nanostructures.

Optical images and cross-section SEM of *G. oto* (A,B) untreated wings, illustrating low reflectance and the presence of nanopillars on the wing membrane surface, and (C,D) hexane-treated wings, illustrating increased reflectance and the loss of nanopillars on the wing membrane, but presence of nipple-like nanostructures on the surface. The dashed squares in A and C indicate approximate regions of the wing used for SEM and spectral reflectance measurements. Scale bars, (B,D) 200 nm. (E) Optical modeling of effective refractive index conditions for (top) untreated wings, with nanopillars of variable height together with cuticle-based nipple nanostructures on the wing membrane, (middle) treated wings, with cuticle-based nipple nanostructures on wing membrane, and (bottom) wing membrane without any nanostructure. y-axis represents height h and x-axis represents effective refractive index condition of air (nair), chitin (n_{chitin}) and wax (n_{wax}). (F) Representative reflectance spectra of experimental (red) and simulation data (black) for untreated wings with nanopillars on the membrane surface (solid line), hexane-treated wings with the wax-based layer of nanopillars removed (dashed line) and membrane only (dotted line).

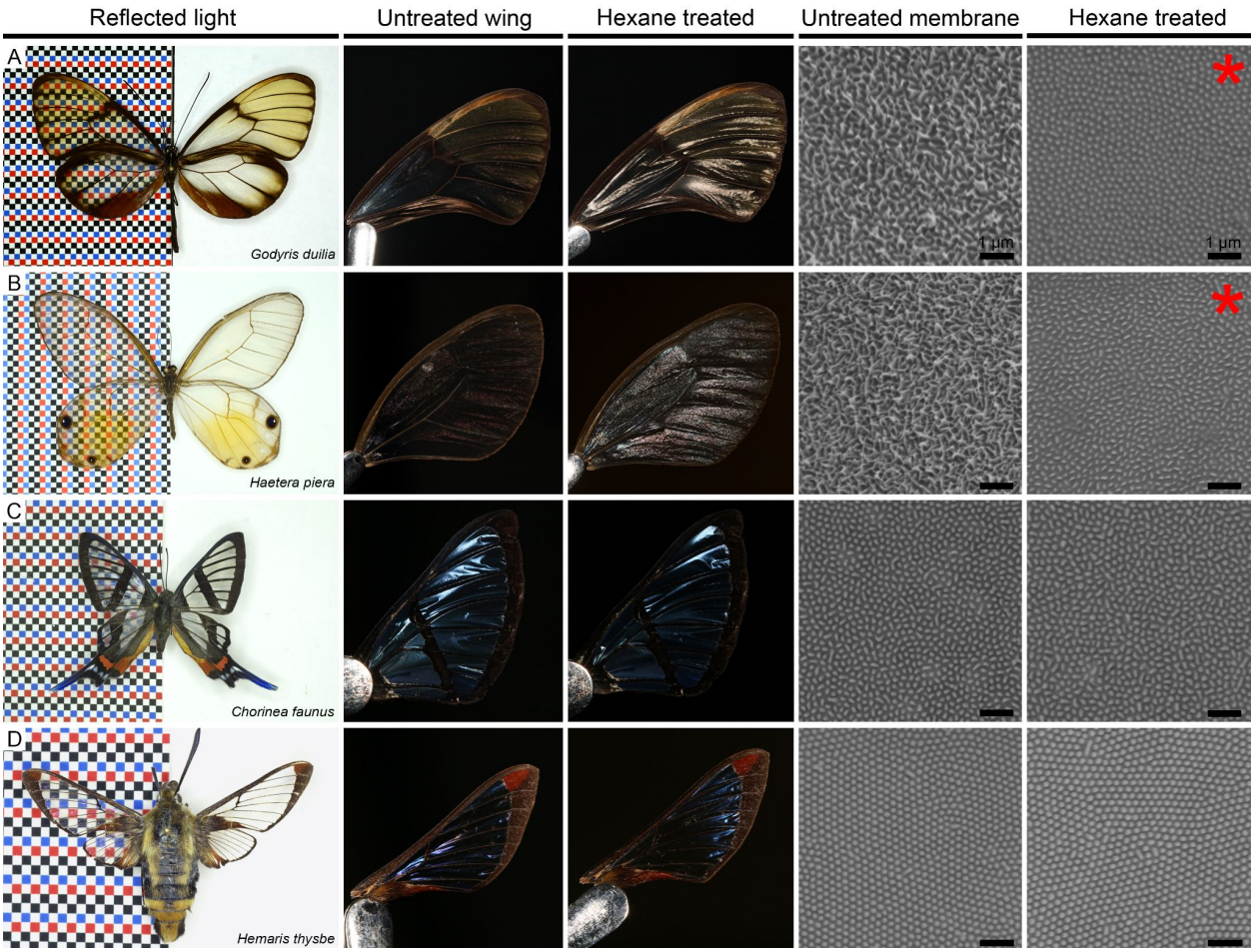


Figure 2.7. Solubility of wing surface nanostructures in additional species of clearwing Lepidoptera.

Untreated wings, hexane-treated wings, SEM of untreated membrane and SEM of hexane-treated membrane for (A) *Godyris duilia* (Nymphalidae: Ithomiini), (B) *Haetera piera* (Nymphalidae: Haeterini), (C) *Chorinea faunus* (Riodinidae: Riodinini) and (D) *Hemaris thysbe* (Sphingidae: Dilophonotini). For both (A) *G. duilia* and (B) *H. piera*, the membrane surface contains irregularly arranged nanopillar structures. After hexane treatments, the wings become more reflective and the upper layer of irregularly arranged nanopillars is removed, while nipple-like structures remain (indicated by red asterisks). For both (C) *C. faunus* and (D) *H. thysbe*, the reflectivity of the wings and the regularly arranged nipple array-like nanostructures on the membrane surface appear unaffected after hexane treatment. Scale bars, 1 µm.
Supplementary Information

Fig. S2.1 Movie. 3D projection of developing scales in a clear wing region 48 hours after pupal formation.

3D projection and rotation of the same scales shown in Fig. 2F, 48 hours APF in a clear wing region. *WGA* (magenta) stains cell membranes and phalloidin (green) stains F-actin and DAPI (blue) stains nuclei. Short actin filaments have reorganized and formed smaller numbers of thick, regularly spaced parallel bundles just under the surface of the cell membrane. Scales alternate with future forked scales appearing as triangular shapes and longer future bristle-like shapes.
https://movie.biologists.com/video/10.1242/jeb.237917/video-1

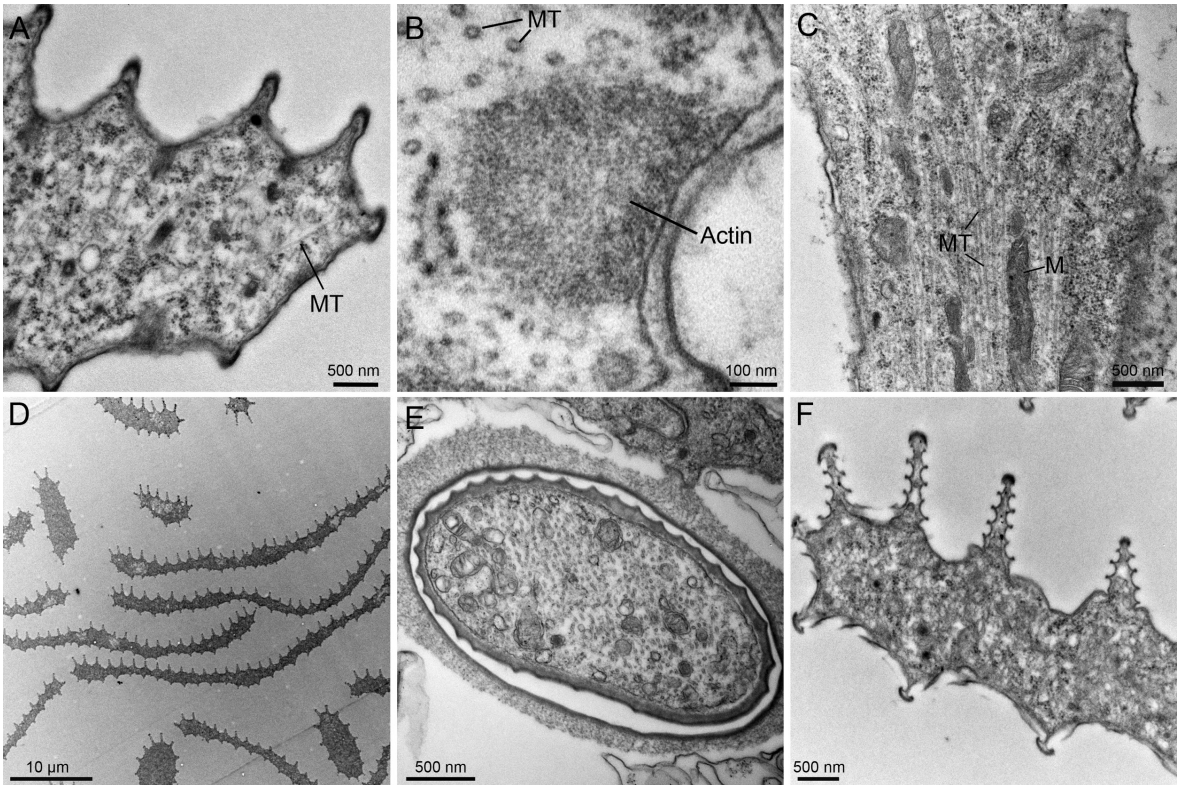
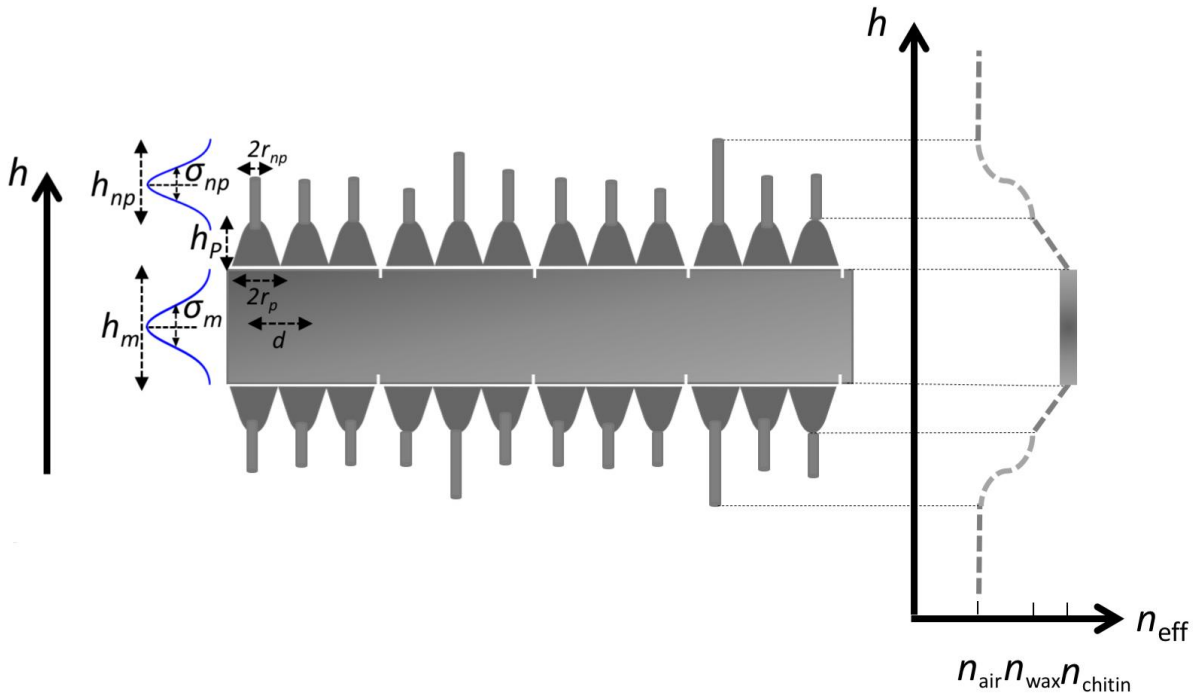


Fig. S2.1 Fig. TEM micrographs of scales 72 hours (top) and 120 hours (bottom) after pupal formation

(A) TEM micrograph of a developing opaque scale 72 h APF, highlighting microtubule arrangement (MT). **(B)** Thick actin bundles contain dense, hexagonally packed F-actin filaments. **(C)** Basal region of a developing scale outgrowth and socket cell. Developing scales 72 h APF contain dense populations of microtubules (MT) and numerous internal organelles, including mitochondria (M), electron dense vesicles and free single ribosomes. **(D)** Transverse section of developing scales around 120 h APF, highlighting both flat and thin, bristle-like scale morphologies. Cross section near the **(E)** base and **(F)** distal region of scales 120 h APF, showing thickened cuticle and ridge morphologies.



S2 Fig. Optical modeling parameters and effective refractive index condition for untreated transparent wing of *Greta oto*

Schematic representation for the optical modeling parameters of wing membrane and surface nanostructures. Average distance between two nanostructures represented as *d*, conical shaped cuticular nipple nanostructures height as h_p, wax-based irregular nanopillars radius as r_{np}, mean height as h_{np} and variance σ_{np}, and membrane thickness as h_m and variance σ_m. Y-axis represents height *h* and X-axis represents effective refractive index condition of air (n_{air}), chitin (n_{chitin}), and wax (n_{wax}).

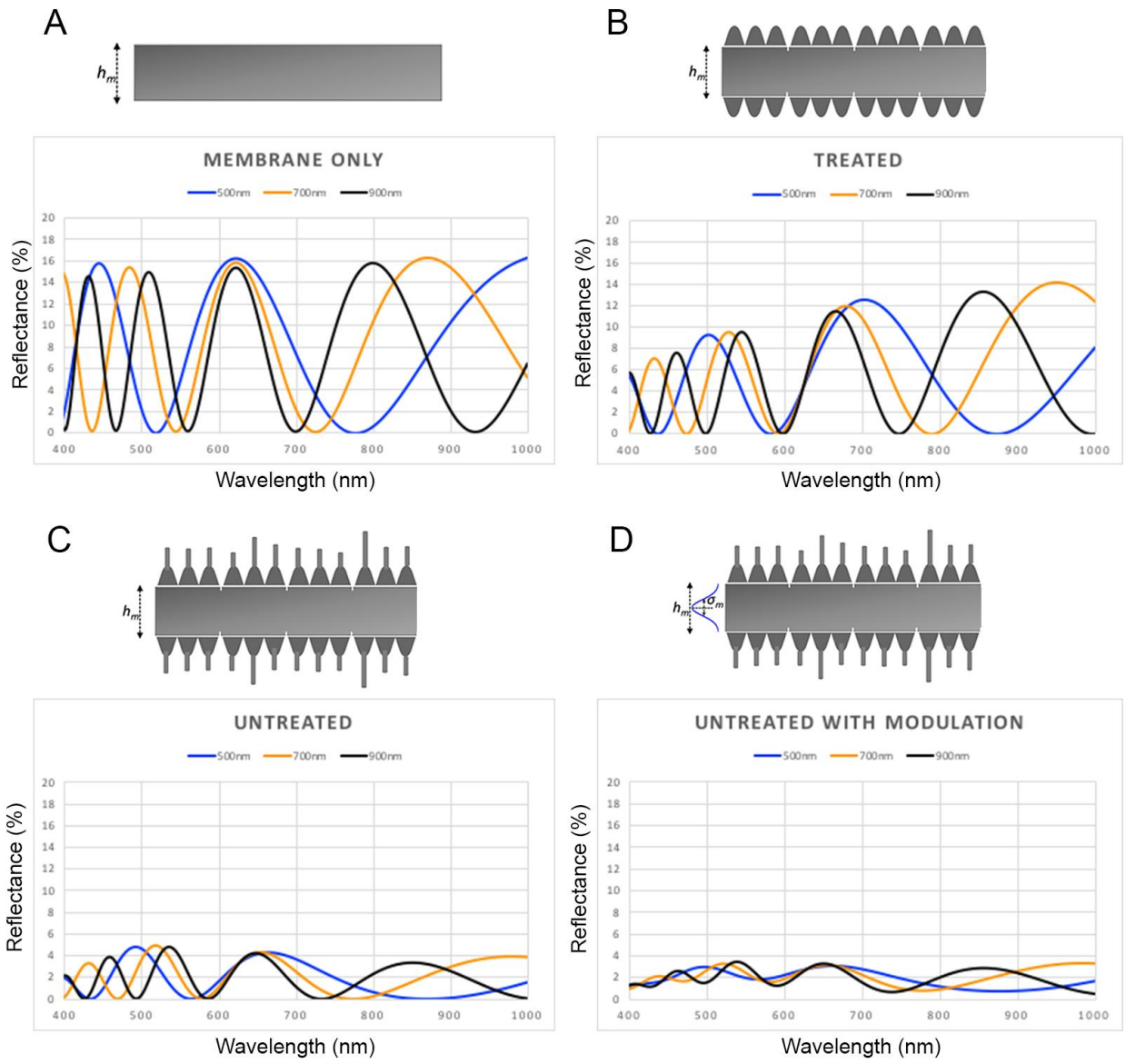


Fig. S2.3 Fig. Optical simulations for mean membrane thickness and modulation of thickness under different wing architecture models

Simulation reflectance spectra of **(A)** Membrane only (lacking surface nanostructures) with varying mean membrane thickness. **(B)** Treated wings (containing cuticle-based nipple nanostructures but lacking wax-based irregular nanopillars) with varying mean membrane thickness. **(C)** Untreated wings (containing wax-based irregular nanopillars and nipple nanostructures) with varying mean membrane thickness and no modulation in thickness. **(D)** Untreated wings with variable mean membrane thickness and modulation of 43 nm variance in thickness.

Preface to Chapter 3

In our efforts to understand the genetic basis of wing transparency, we asked ourselves: what pathways regulate the development of a scale cell, and what are the factors involved in a species that contains transparent 'windows' in which scales do not develop? Lepidoptera wing scales and *Drosophila melanogaster* bristles are considered homologous structures based on similarities in their cell lineages and developmental programs. However, the molecular mechanisms underlying scale development in Lepidoptera are less well characterized. To better understand gene expression associated with wing scale formation, we characterized the transcriptomic landscape in pupal wing tissue during timepoints when scale precursor cells differentiate from epithelial tissue and when early scale morphogenesis occurs. We carried out stage-specific RNA sequencing of micro-dissected wing tissue in the buckeye butterfly *Junonia coenia* and the giant silkmoth *Antheraea polyphemus* to identify conserved differentially expressed genes. Additionally, we investigated differential expression between two regions within the wing of *A. polyphemus*: a transparent region we refer to as a 'window' in which scale cells do not develop, and an adjacent region that undergoes canonical scale development. We chose *Antheraea polyphemus* as a candidate species to probe wing scale development because numerous species belonging to the giant silkmoth family Saturniidae contain transparent eyespots in their wings located along the crossvein. Initial investigation of this phenotype by members of our lab (led by Yuriko Kishi) through pupal wing dissections and fluorescent staining revealed that this wing region is transparent in some saturniids due to a lack of scale precursor cell formation during development (resulting in the absence of scales). The molecular mechanism responsible for this apparent suppression of scale precursor cell development was unknown.

Analyzing lepidopteran wing transcriptomes to search for genes that are differentially expressed between tissue types has been previously applied to understand the molecular differences in pigmentation patterns in butterflies. Utilizing a similar approach, we set out to characterize the gene regulatory landscape associated with scale development in *Junonia coenia* and wing transparency in the giant silkmoth *Antheraea polyphemus*. We then applied fluorescent *in situ* hybridization and CRISPR/Cas9 induced knockouts to characterize the spatiotemporal expression and function of genes involved in scale cell development.

Chapter 3: Transcriptomics of pupal wings and transparent 'windows' provide insight into the genetic basis of scale development in Lepidoptera

Aaron Pomerantz[1,2,*], Yuriko Kishi[3,4], Kyle DeMarr[1,2], Nipam Patel[2,*]

Affiliation
[1]Department of Integrative Biology, University of California, Berkeley, Berkeley, CA 94720, USA
[2]Marine Biological Laboratory, Woods Hole, MA 02543, USA
[3]Department of Molecular & Cell Biology, University of California, Berkeley, Berkeley, CA 94720, USA
[4]Department of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125, USA
*Authors for correspondence (pomerantz_aaron@berkeley.edu, npatel@mbl.edu)

Key words
scales; sensory organ; neurogenesis; development; RNA-seq, butterfly wing; CRISPR/Cas9

Abstract

Background

The wings of butterflies and moths (Lepidoptera) are densely covered by scales; flat cuticular projections that derive from epidermal cells and arise during the early stages of pupal development. Wing scales are homologous to insect sensory bristles and are a specialized type of non-innervated macrochaete that serve as the basis of color production in Lepidoptera. To better understand gene expression associated with wing scale formation, we characterized the transcriptomic landscape in pupal wing tissue during timepoints when scale precursor cells differentiate from epithelial tissue and when early scale morphogenesis occurs. We carried out stage-specific RNA sequencing of micro-dissected wing tissue in the buckeye butterfly *Junonia coenia* and the giant silkmoth *Antheraea polyphemus* to identify conserved gene expression. Additionally, we investigated differential expression between two regions within the wing of *A. polyphemus*: a region we refer to as a 'window' in which scale cells do not develop, and an adjacent region that undergoes canonical scale development. We then applied fluorescent *in situ* hybridization and CRISPR/Cas9 induced knockouts to characterize the spatiotemporal expression and function of genes involved in scale cell development.

Results

RNA-seq comparisons between *J. coenia* and *A. polyphemus* uncovered genes with equivalent expression patterns during early pupal wing development and scale precursor differentiation, including proneural, cell cycle, and *Notch* signaling factors. At

later pupal stages, when scale cell projections are forming and maturing, we identified genes with equivalent expression patterns related to cytoskeletal organization, melanization, cuticle formation, and chitin-synthesis. Using stage-specific transcriptomic analysis followed by HCR *in situ* hybridization, we uncover a suite of genes that likely play conserved roles in scale cell patterning and morphogenesis in butterflies and moths. We identified two *achaete scute* homologs (*ASH1*, *ASH2*) expressed at the scale cell precursor stage and loss of function of *ASH2* resulted in the loss of scale cells. In contrast, loss of function of the *Notch* receptor led to overproduction and dense clusters of scale cells, likely due to improper lateral inhibition during scale precursor cell differentiation. In wing discs, we observed butterfly proneural and *Notch* signaling genes share similar expression patterns to orthologs in *Drosophila melanogaster*, but in contrast, the spatial regulation shifts and expands to intervein regions shortly after pupation in butterflies, resulting in organized rows of scale precursor cells covering the wing. We also identified that the 'window' scaleless region in *A. polyphemus* is associated with high expression levels of Wnt ligands, including *wingless,* and the bHLH transcription factor *hairy*, a negative regulator of sensory bristles, suggesting how putative co-option of neurogenesis regulatory factors contribute to scale cell patterning in lepidoptera.

Conclusion

The scales that adorn the wings of butterflies and moths are a morphological innovation. Here, we present a comparative transcriptomic analysis of pupal wing tissue between two lepidopterans, toward the identification of biological processes and functional gene activities involved in scale cell formation. Our results confirm previously implicated pathways required for insect macrochaete development, such as *Notch* signaling which plays a key role in patterning scale precursor, and also identify potentially novel genes conserved in butterfly and moth wing scale development. Our dataset suggests various potential mechanisms that drive scale cell development, which will help facilitate the characterization of scale cell evolution and provide insight into how changes in gene expression during wing development play a role in cellular morphological diversity. Spatial regulation of neurogenesis factors during Lepidopteran evolution likely promoted scale cell formation across the wings.

Introduction

The diversity of colorful wing patterns in butterflies and moths (Lepidoptera) have captivated humans for centuries. Butterfly and moth wings are covered with thousands of flat overlapping scales, each one of which derives from a single cell during pupal development. Wing scales represent a multifunctional evolutionary innovation that can function in color production, antipredator strategies (e.g. camouflage, deflection, aposematism, mimicry), attracting mates, thermoregulation, and water repellency (Nijhout, 2001; Perez Goodwyn et al., 2009; Mazo-Vargas et al., 2017; Deshmukh et al., 2018; Tsai et al., 2020). Although Lepidopteran wing scales are non-innervated cellular structures, they are considered homologous to insect sensory organs, which can give rise to a multitude of different cell types such as neurons, glia, sheath cells, and cells

generating external cuticular projections, including sockets and bristles or scales (Overton, 1966; Galant et al., 1998; Lai and Orgogozo, 2004; Klann et al., 2021). For instance, mechanosensitive bristles (microchaetes and macrochaetes) are distributed in regularly spaced rows on the dorsal thorax of *Drosophila melanogaster* and derive from the sensory organ precursor (SOP) cell lineage (Calleja et al., 2002; Furman and Bukharina, 2008; García-Bellido and de Celis, 2009).

The first step of SOP differentiation involves relatively large clusters of contiguous ectodermal cells, called proneural clusters, becoming competent to develop as sensory organ precursors (Cubas et al., 1991). Proneural genes are essential for the development of SOPs; in the absence of these genes, SOPs will not form. In *D. melanogaster*, several different proneural genes have been identified which belong to a class of basic helix-loop-helix (bHLH) transcriptional regulators, in particular the *achaete-scute* complex (AS-C). The As-c locus contains four genes, *achaete* (*ac*), *scute* (*sc*), *lethal of scute* (*l'sc*), *asense* (*ase*), but only *ac* and *sc* are required for cells to adopt the neural fate and form bristles (Cubas et al., 1991; Hinz et al., 1994; García-Bellido and de Celis, 2009). Both *ac* and *sc* genes are co-expressed, share cis-regulatory elements and act redundantly to specify bristles (Gomez-Skarmeta et al., 1995). Proneural bHLH factors are initially expressed in ectodermal cells, giving the cells competence to undergo neural commitment, and are then upregulated upon neural commitment. In mutants with strongly reduced or missing proneural clusters of *ac*/*sc* expression, the corresponding sensory organ precursors (SOPs) to bristles do not form, whereas in other mutants with enlarged domains of *ac*/*sc* expression extra SOPs and corresponding bristles emerge in ectopic positions (Cubas et al., 1991; Skeath and Carroll, 1991; Cubas and Modolell, 1992).

During the subsequent neurogenic stage of bristle formation, the *Notch* pathway is critical for the progression of SOP cell development. Proneural gene expression activates the *Notch/Delta* pathway, which limits neural fate adoption to a single cell surrounded by neighboring epidermal cells, through a process called lateral inhibition (Simpson, 1990; Campos-Ortega, 1995; Parks et al., 1997). Genes encoding elements of the lateral inhibition pathway include the gene *Notch* that encodes the receptor, *Delta* that encodes the *Notch* ligand, and the *Enhancer of split* complex (*E(spl)*) that belong to the group of *Hairy*-related proteins, a distinct subfamily of bHLH proteins that generally operate as DNA-binding transcriptional repressors. Upon ligand (e.g. *Delta*) binding to a *Notch*, the cytoplasmic domain of the *Notch* receptor is cleaved and translocates to the nucleus where it activates target genes such as the *Enhancer of split* complex, which promotes the development of epidermal cells, as opposed to SOPs, by inhibiting the transcriptional activator proneural genes (Couturier et al., 2019). Together, the activator and repressor genes that encode these proteins have co-evolved as a regulatory gene ‘‘cassette’’ involved in many cell fate decisions, including the formation of sensory organs. In normal development, cells which turn down proneural genes due to *Notch/Delta* signaling will join the cells outside the proneural clusters and develop as epidermal cells. If the neurogenic signaling pathway is absent or deficient, all cells of the proneural cluster become SOPs, resulting in supernumerary SOPs and holes in the epidermis (Hartenstein and Posakony, 1990; Parks et al., 1997). In SOPs that give rise to external or internal mechanosensory organs, the SOP cell divides to produce two daughter cells, which then undergo another round of asymmetric division to ultimately

yield the sensory neuron, glial, sheath, socket, and macrochaete cells. The binary decisions between these asymmetric divisions and cell fates are regulated by *Notch* and Numb (Schweisguth, 2015).

Distinct combinations of transcription factors can generate a variety of sensory organ cell types from the SOP lineage, such as the choice between external and internal sensory organ identities. In *D. melanogaster*, several transcription factors are switched on following SOP differentiation, and changes in cell subtype identity occur if particular genes are mutated. For example, *cut*, which encodes homeobox domain transcription factor, is expressed in SOP cells that give rise to external sensory organs. Loss of function in *cut* results in transformation of external sensory organs into a different cell type that gives rise to internal stretch receptors called chordotonal organs, whereas ectopic *cut* expression results in the reciprocal transformation (Bodmer, 1987; Blochlinger et al., 1990, 1991).

Research on insect macrochaete development has yielded insight into genetic changes associated with cellular morphological diversity. Interestingly, recent work has shown that a single nucleotide substitution in an enhancer of the *scute* gene contributes to both the gain of bristles on the legs and loss of bristles on the genitalia in *Drosophila santomea*, demonstrating how an evolved mutation via a pleiotropic enhancer can impact the development of bristles on multiple appendages at once (Nagy et al., 2018). Different types and subtypes of sense organs exist on arthropod bodies, which raises the question of how this variety of cells emerged and what molecular mechanisms facilitated their evolution. Work remains to elucidate the factors involved in SOP identity beyond *D. melanogaster* and if the same pathways and processes are deployed in other insect species such as butterflies and moths.

The wings of butterflies and moths are established during the embryonic stage, forming within the body of the caterpillar as sac-like epithelial structures known as imaginal discs. During larval development, wing imaginal discs enlarge, veins form, and pre-patterns (such as eyespots and color boundaries) are established through the expression of transcription factors and signaling molecules (Macdonald et al., 2010; Martin and Reed, 2014; Monteiro, 2015; Mazo-Vargas et al., 2017). Upon pupation, the imaginal discs become everted to the external surface of the larva where proliferation, patterning, and differentiation of scale precursor cells occur within the first few hours of pupal development. During this process, select epidermal cells differentiate to become scale precursor cells, which undergo subsequent differential divisions that yield a highly polyploid scale and a socket cell. Following specification of the scale and socket, the scale projection begins as a small cytoplasm-filled extension, surrounded by an active cell membrane that extends out of the wing epithelium. The extensions begin as cylinders that become flattened during growth. The scale cell extension rapidly grows during development and components of the cytoskeleton and cell membrane form the overall shape and architecture, ultimately producing the surface extension that we recognize as an adult wing scale (Dinwiddie et al., 2014; Day et al., 2019). In the late stages of pupation the scale cell dies, leaving a non-living skeleton of chitin filled with air and pigment.

In order to determine if the developmental pathways leading to *D. melanogaster* sensory bristle and butterfly scale formation use similar genetic circuitry, a homolog of the *Drosophila achaete-scute* (AS-C) genes, which encode transcription factors that

promote neural precursor formation, was cloned from the butterfly *Junonia coenia* and its expression pattern was examined during development (Galant et al., 1998). During embryonic and larval development, the expression pattern of this AS-C homolog (*ASH*) was detected in clusters of cells corresponding to the central and peripheral nervous system. The *ASH* is also expressed in larval wing discs along the future wing veins and margin, corresponding to sensory scale precursor cells. *ASH* is later expressed in pupal wings in evenly spaced rows of enlarged cells that segregate from the underlying epidermis, but, rather than giving rise to neural structures, each cell contributes to an individual scale. If scales and bristles are in fact homologous structures, then the non-innervation of scales is consistent with the proposed programmed cell death of the basal daughter cell of the putative scale precursor, the cell which in *D. melanogaster* is fated to produce the neuron and glia. These observed expression patterns support the notion that *ASH* may perform multiple functions throughout butterfly development, such as promoting the initial events of selection and formation of both neural and scale precursor cells.

Similarities in the cellular and molecular processes of scale and neural precursor formation suggests that the spatial regulation of *achaete-scute* homologs have been modified during Lepidopteran evolution to promote scale cell formation. Additionally, observation of *Notch* expression in *Heliconius* butterflies suggests that scale precursors differentiate from their surrounding epithelia in evenly-spaced rows via lateral inhibition and asymmetric cell division for scale and socket cell fate (Reed, 2004). A series of studies have investigated a scaleless mutant of *Bombyx mori*, which displays a severe reduction in wing scales (Zhou et al., 2004, 2006, 2009). This recessive mutantation was reported to result in severely reduced expression of an *achaete-scute* homolog (*Bm-ASH2*) was severely reduced in the scaleless mutant pupal wings. The authors provided evidence that a 26 bp deletion within a putative promoter of *Bm-ASH2* is closely linked to the scaleless locus, and potentially caused the reduction of *Bm-ASH2* expression and the scaleless wing phenotype. Thus, an *achaete-scute* homolog appears to play a critical role in scale formation in *Bombyx mori*, supporting the proposed homology of lepidopteran scales and dipteran bristles. Since these experiments on a scaleless mutant and expression pattern of a single *ASH* gene, little work has gone into assessing proneural and neurogenic pathways further, and their functional roles in Lepidoptera wing development have not been directly assessed.

While researchers have made progress in understanding genetic pathways responsible for pigment production and genetic factors that demarcate wing pattern positions during imaginal disc development, less is known about the genetic underpinnings of scale cells, which serve as the basis of color production in Lepidoptera wings. This study therefore aimed to (1) characterize genes implicated in scale cell precursor differentiation and scale morphogenesis using comparative transcriptomics, (2) test if arthropod bristle development genes also play similar roles in lepidopteran wing scale formation, and (3) characterize genes potentially involved in repression of scale development in the ‘window’ region of the giant silkmoth. To achieve these goals, we employed RNA sequencing (RNA-seq) in the buckeye butterfly *Junonia coenia* and the giant silkmoth *Antheraea polyphemus*, and were able to identify candidate scale formation genes expressed during wing development. We then employed HCR *in situ*

hybridization and CRISPR/Cas9 genome editing to assess the spatiotemporal expression and function of several putative scale development genes.

Results

Stage specific morphogenesis of wing scale development

Lepidopteran wing tissue undergoes a series of morphological transitions during early pupal development. We therefore carried out stage-specific RNA sequencing of micro-dissected wing tissue in the buckeye butterfly *Junonia coenia* and the giant silkmoth *Antheraea polyphemus* at four stages of wing development, which we term 1) epithelial, 2) scale precursor, 3) scale early, and 4) scale late (Fig 1). Our morphological observations of pupal wing tissue at stages dissected for RNA-seq mirror previous descriptions (Dinwiddie et al., 2014). At the 'epithelial stage' (approximately 8-12 h after pupa formation (APF) in *J. coenia*), we observed that the wing is composed of a bilayer of undifferentiated epithelial (EP) cells (Fig 1B). At the 'scale precursor' stage (approximately 24 hours after pupal formation in *J. coenia*), we see that certain cells have segregated and differentiated into scale precursors (SP) cells (outlined in white for better visibility) surrounded by neighboring undifferentiated epithelial cells (Fig 1C). At the 'scale early' stage (approximately 72 hours APF in *J. coenia*), we observed round scale cell projections labelled with phalloidin (green) that mark large bundles of actin filaments and the outer cell membrane that extends out of the wing epithelium marked with wheat germ agglutinin (magenta) (Fig 1D). Finally, at the 'scale late' stage (approximately 120 hours APF in *J. coenia*), we observed maturing scale projections, with actin bundles extending to the finger-like tips at the end of the scale and more prominent ridges that have formed in between the existing actin bundles (Fig 1E). Additionally, we investigated a region we refer to as a 'window' in the wings of the giant silkmoth *A. polyphemus*, in which scale cells do not develop. At all stages of the 'window' region (bottom panel) in *A. polyphemus* we find that scale precursor cells do not appear to initiate, resulting in scale-less membrane surrounding the crossvein region of the wing in the adult moth (Fig 1 M-Q).

Transcriptome analysis of wing scale development in butterflies and moths.

To identify candidate genes associated with scale development in *J. coenia* and *A. polyphemus*, we quantified gene expression across the four distinct stages of pupal wing development (epithelial, scale precursor, scale early, and scale late) in both species using RNA-sequencing (Fig 1). In total, we sequenced twelve wing tissue samples from *J. coenia* and eight wing tissue samples from *A. polyphemus*. After initial filtering, we performed *de novo* assemblies based on a total of 362 million paired-end reads for *J. coenia* and 233 million paired-end reads for *A. polyphemus*. The overall alignment rate ranged from 81.9% to 86.8% for *J. coenia* and 78.9% to 90.8% for *A. polyphemus*. We then performed transcript quantification and differential gene expression (DGE) analyses for the two species separately. For each species, we performed PCA which showed clustering of samples by stage (Fig 2 A,B). Clustering of the gene expression data based with hierarchical dendrograms of differential gene

expression further indicated that samples clustered by their appropriate developmental stage (Fig 2 C,D). A total of 2188 transcripts were identified by DESeq2 as differentially expressed in *J. coenia* (Fig 2C) and 1319 in *A. polyphemus* (Fig 2D). Sets of genes showed clear stage-specific expression patterns that displayed similar expression profiles between both species.

Shared DEGs and GO enrichment highlight conserved scale development pathways in butterflies and moths

To gain insights into transcriptome dynamics during pupal wing development, pairwise differential expression analyses between stages (epithelial vs SOP, SOP vs scale early, scale early vs scale late) were conducted using DESeq2. We next performed Gene Ontology enrichment analysis based on DEGs that displayed similar expression patterns in both *J. coenia* and *A. polyphemus*. We identified a total of 406 differentially expressed genes that displayed similar expression levels in both species, highlighting pathways that are likely conserved in early wing development across Lepidoptera.

For 'epithelial' versus 'SP' differential expression comparisons, we identified 1646 transcripts in *J. coenia* and 1042 transcripts in *A. polyphemus*. Of these DEGs, 290 transcripts shared equivalent expression patterns between both species. Gene Ontology (GO) enrichment analysis produced 26 biological processes (BP) enriched at the epithelial stage, including terms such as 'cell-matrix adhesion' (GO:0007160), 'chitin catabolic process' (GO:0006032) and 'integrin-mediated signaling pathway' (GO:0007229) (Fig 2E). For the scale precursor stage, GO enrichment analysis produced 59 biological processes including 'peripheral nervous system development' (GO:0007422), 'sensory organ development' (GO:0007423), and 'mitotic cell cycle' (GO:0000278) (Fig 2F). For 'scale early versus scale late' differential expression comparisons, we identified 2332 transcripts in *J. coenia* and 882 transcripts in *A. polyphemus*. Of these DEGs, 193 transcripts had similar expression levels between both species. For the scale early stage, GO enrichment analysis produced 11 biological processes including 'apical constriction' (GO:0003383), 'cuticle pattern formation' (GO:0035017) and 'actin filament organization' (GO:0007015) (Fig 2G). Finally, for the scale late stage, GO enrichment analysis produced 18 biological processes including 'chitin metabolic process' (GO:0006030), 'chitin-based cuticle development' (GO:0040003) and 'cuticle pigmentation' (GO:0048067). (Fig 2H). Overall, we find shared sets of genes in butterflies and moths at particular stages of wing development, which include biological processes related to actin organization, cell adhesion, cell cycle, proneural, *Notch* signaling, chitin binding, cuticular protein and pigmentation.

Differential gene expression associated with epithelial and scale precursor stages

The *Notch* signaling pathway is well characterized based on macrochaete development in *D. melanogaster*, which is required for lateral inhibition and proper sensory organ precursor formation (Fig 3A). From our differential expression results in *J. coenia* and *A. polyphemus*, we identified a list of differentially expressed transcripts that included notable genes such as *Integrin*, *U-shaped* and *prospero* that displayed highest expression levels at the epithelial stage of pupal wing development. Genes including

members of the *achaete scute* complex (two *achaete scute* homologs, *asense*) *senseless*, several *cyclins*, *neuralized*, and *enhancer of split* genes displayed highest expression levels at the SOP stage (Fig 3B). *Delta*, *Notch* and *numb* displayed more dynamic expression across developmental stages (Fig 3B).

Differential gene expression associated with scale early and scale late stages

Cytoskeletal dynamics and chitin synthesis play essential roles in wing scale morphogenesis and aspects of scale ultrastructure during pupal development (Fig 3C). From our differential expression results in *J. coenia* and *A. polyphemus* at later stages of wing development, we identified cytoskeletal-related genes such as *forked*, *singed* (aka *fascin*) and microtubule associated protein *Jupiter*, as well as ZP domain protein coding genes, which play roles in cuticle formation, such as *dusky-like*, *miniature* and *trynity* with highest expression levels at the scale early stage. Finally, we observed genes including *Chitin synthase-kkv* and several cuticular proteins with highest expression levels at the scale late stage (Fig 3D). For instance, we identified Cuticular proteins such as *CPLCP12*, *CP92F*, and *CP62Bc* differentially expressed in the scale late stage of both *J. coenia* and *A. polyphemus*. These expression results support the importance of cytoskeletal and chitin regulators that likely play roles in shaping the complex morphologies of these cuticular projections.

Spatiotemporal expression and functional assessment of proneural factors

Members of the *achaete-scute* complex (AS-C) are proneural genes that play essential roles in neurogenesis and are required for the formation of all external sensory organs. From our RNA-seq data, we identified two *achaete-scute* homologs *ASH1* and *ASH2* that were highly expressed at the scale precursor stage. Using HCR *in situ* hybridization, we assessed the spatiotemporal expression of these two achaete scute homologs. In the larval wing disc, we observed *ASH1* mRNA expressed primarily along the wing margin (Fig 4A). In pupal wing tissue at the scale precursor stage (24 hrs APF), *ASH1* expression had shifted into organized rows with highest expression in scale precursor cells and absent in the undifferentiated epithelial cells (Fig 4B). In the wing disc, we observed *ASH2* mRNA expression in a small number of cells along the wing margin and future wing veins, marking future sensory bristles, which is in line with previous observations (Fig 4C, Galant et al. 1998). Similar to *ASH1*, in pupal wing tissue at the scale precursor stage, the expression of *ASH2* had shifted into organized regularly spaced rows across the wing, with high expression corresponding to scale precursor cells and absence of expression in epithelial cells (Fig 4D). We did not detect expression of the third lepidopteran achaete scute homolog (*ASH3*) in the wing disc or pupal wings via RNA-seq or HCR *in situ* hybridization (Fig S1).

In *D. melanogaster*, *cut* expression is observed in the sensilla of the wing margin, the third longitudinal vein, and the anterior cross vein of pupal wings. In *J. coenia*, we observed strong *cut* mRNA expression in the peripheral tissue along the margin of wing discs (Fig 4E, in line with previous antibody stainings by Macdonald et al., 2010). In pupal wing tissue at the scale precursor stage, *cut* was still expressed at high levels in peripheral tissue. Interestingly, in pupal wing tissue we also observed *cut* expression in

a small subset of scale precursor cells (Fig 4F). Based on the position of these cells, these appear to be precursors to large hair-like scales present on the hindwings. If so, this expression could be the first observation of *cut* expression corresponding to putative subtype specification of hair-like scales in butterflies.

Expression patterns have been observed for some butterfly *achaete-scute* homologs, but functional testing of these genes was lacking. To assay the function of these genes during wing development, we designed CRISPR guides against *ASH2* and generated G0 somatic mosaic mutants. In resulting adults, we observed three individuals that had mosaic patterns of missing scales on their wings. Higher magnification of the wings revealed that the region missing scales was also devoid of sockets, suggesting that SOP cells did not form and as a result no socket or scale cells formed in these mutant regions during pupal wing development (Fig 4G-H). These results using CRISPR/Cas9 knockouts provide evidence for the functional role of *ASH2* in scale cell precursor formation.

Spatiotemporal expression and functional assessment of neurogenic factors

In the imaginal disc, the expression of the *Notch*-ligand, *Delta***,** appeared restricted to the future vein territories, similar to expression in *D. melanogaster* wing discs (Fig 5A). In pupal wing tissue at the scale precursor stage, *Delta* had shifted expression into organized rows with high expression levels in scale precursor cells, while expression was not detected in epithelial cells (Fig 5B). In contrast, we observed *Notch* expression at high levels within intervein midline territories and lower levels within future vein territories in the imaginal disc (Fig 5C). In pupal wing tissue at the scale precursor stage, *Notch* expression had shifted and appeared in a grid-like pattern, occurring at the highest levels in epithelial cells and lower levels within scale precursor cells (Fig 5D). This expression pattern differs from expression in *D. melanogaster* pupal wings, in which *Notch* expression becomes restricted to stripes localized at the vein-intervein boundaries. Finally, we investigated expression of a homolog of *Enhancer of split*, which is known to be involved in *Notch* signaling. We observed high levels of *Enh(spl)* flanking both sides of the developing wing veins in the imaginal disc, similar to expression patterns found in *D. melanogaster* wing discs (Fig 5E). In pupal wing tissue at the scale precursor stage, *Enh(spl)* appeared to be expressed at higher levels in epidermal cells surrounding scale precursor cells (Fig 5F).

Expression patterns have been documented for the receptor molecule *Notch* in butterflies, but functional validation of this gene was lacking. To assay the function of *Notch* in wing development, we targeted multiple exons and generated G0 somatic mosaic mutants using CRISPR/Cas9 knockouts and observed two adult individuals that contained mosaic wing defects. Upon closer inspection, we observed regions of the wing that lacked scales, as well as regions where multiple wing scales were densely clustered together, likely due to improper lateral inhibition during scale precursor cell differentiation, as well as thickened wing veins (Fig 5G-H, Fig S2). Taken together, these results provide direct evidence for the functional role of *Notch*-mediated lateral inhibition in spatial regulation of scale precursor cells.

Differential expression between the "window" and "non-window" regions of *A. polyphemus* wings

In *Antheraea polyphemus*, the 'window' region straddles the crossvein and tissue in this region fails to initiate scale development, whereas adjacent non-window regions develop scales normally (Fig 1). We extended our transcriptomic analysis to the 'window' versus 'non-window' and identified 31 differentially expressed genes between the two tissue types (Fig 6A). These DEGs were primarily expressed at high levels in the window region, including members of the Wnt signaling pathway, such as *wingless*, *Wnt6, Wnt10, frizzled*, *arrow*, *odd paired* and *divisions abnormally delayed*, as well as the bHLH transcription repressor *hairy* (Fig 6A,B). GO enrichment analysis of these DEGs generated biological processes such as 'Wnt signaling pathway', 'neuron differentiation', and 'cell fate commitment'.

Based on our RNA-seq results implicating several Wnt molecules expressed in the window region, we injected *A. polyphemus* pupae with high concentrations of heparin, a highly sulfated form of heparan sulfate glycosaminoglycans which is thought to interact with Wnt signaling (among other molecular pathways) in Lepidoptera. In line with findings from Sourakov & Shirai (2020), we observed what appear to be strong effects of heparin, in which certain wing pattern elements of the treated individuals had dramatically shifted (Fig 6C). For instance, silvery scales that are normally circling the window region had expanded to encompass the entire discal cell and a pattern normally present as a small stripe of red scales had expanded all the way to the distal margin of the wing, and melanic patches had expanded, especially in the hindwing (Fig 6). We did not observe a noticeable change in the size of the window wing pattern in heparin treated adults relative to control adults.

Based on RNA-seq results and the fact that *wingless* is known to be expressed along the crossvein in insect wings, we designed HCR probes for *A. polyphemus wingless* to determine expression in pupal wings. We observed a strong expression pattern of *wingless* at the crossvein which coincided with the window wing pattern (Fig 6D,E). Finally, our RNA-seq results pointed to a promising candidate gene known as *hairy*, which was highly expressed in the window region (Fig 6A,B). Previous studies have implicated *hairy* as a direct transcriptional repressor of *achaete scute* in *D. melanogaster*, which made it a gene of interest as a possible repressor of scales in the window region of *A. polyphemus*. We therefore designed HCR probes and detected higher expression of *hairy* largely overlapping with *wingless* expression in the 'window' region surrounding the discal crossvein (Fig 6F,G). Notably, *hairy* expression was strongly correlated with the absence of scale precursor cell formation and expression was lower in adjacent 'non-window' regions in which scale precursor cells form (Fig 6 G,H). While functional analysis of *hairy* has yet to be performed in *A. polyphemus*, our results identify a correlation between high *hairy* expression and absence of scale precursor cell formation in the 'window' region during pupal wing development, which would be in line with its known role as a repressor of SOP formation in *D. melanogaster*.

Discussion

This study utilized a developmental transcriptomics approach to elucidate components of the gene networks deployed during scale cell development in Lepidoptera. Several lines of evidence suggest a cellular homology between sensory bristles and the scales of butterflies. From a morphological perspective, both bristles and scales are large, unicellular structures associated with a basal socket cell, and there are similarities in the cell lineage and ontology of asymmetric divisions leading to scales and sensory bristles. Expression of proneural and neurogenic orthologs in developing scales and bristles, such as *Achaete/Scute* and *Notch/Delta*, as well as a cytoskeletal factors such as F-Actin cables leading to surface ridging are also conserved between *D. melanogaster* and butterflies. Finally, there is evolutionary evidence from the sister lineage to Lepidoptera, the Trichoptera, in which several lineages appear to have independently gained bristles and scale-like structures covering the wings (Overton, 1966; Galant et al., 1998).

Proneural factors during pupal wing development in Lepidoptera

The *achaete-scute homolog* (*ASH*) genes are present in all metazoans and are involved in neural development and the specification of sensory organs, making them a model for the study of development and pattern formation (Negre and Simpson, 2009). The ancestral AS-C in insects was composed of two genes: an *achaete-scute homolog* gene and an *ase* gene. *ASH* genes have subsequently undergone independent duplications in different lineages (Negre and Simpson, 2009; Ayyar et al., 2010). This complex is typically clustered and surrounded by CytP450 and yellow genes in insect genomes. Coleoptera and Hymenoptera show the ancestral configuration with one *ASH* and one *ase* gene. In the beetle *Tribolium castaneum*, *Tc-Ash* and *Tc-ase* are expressed in all neural precursors and play roles in neural precursor formation and regulating nervous system development, respectively, with recent functional evidence from RNAi indicating that *Tc-Ash* is required for the formation of all morphological and functional classes of external sensory organs (Wheeler et al., 2003; Klann et al., 2021). In semi-aquatic bugs (Heteroptera, Gerromorpha) the *achaete-scute* complex contains only a single gene that controls bristle development (Finet et al., 2018). In the Diptera, the number of *achaete-scute* genes vary as a result of independent duplication events, which appear to correlate with the emergence and stereotyped patterning of macrochaetes. For instance, basal dipterans such as those in the Nematocera (which includes mosquitoes) contain only two genes in the AS-C and species in this group display a random distribution of bristles on the notum. On the other hand, two tandem duplications in the lineage leading to the Schizophora (which includes species such as *Drosophila melanogaster*) have led to a total of four genes in the AS-C and species in this group display more spatially organized patterns of machochaetes on the notum (Calleja et al., 2002; Skaer et al., 2002). In Lepidoptera, *Bombyx mori* has been shown to contain one *ase* and three *ASH* genes, which represent independent duplications from those of Diptera (Negre and Simpson, 2009; Zhou et al., 2009). These data raised a question of whether *ASH* duplications or differential gene regulation could be related to the origin of wing scales. In *B. mori*, *ASH1* and *ASH2* were found to be expressed in all organs derived from the mesoderm, ectoderm and endoderm, while *ASH3* and *ase* were expressed only in organs derived from mesoderm and ectoderm and the *ASH* genes

were suggested to play important roles in regulating neurogenesis of embryo and formation of pupal wing (Zhou et al., 2008).

Our data provides additional support for the role of the *achaete scute* homologs in Lepidopteran wing scale development and provides a direct functional role of *ASH2* in scale cell precursor formation, as mutations in this gene result in the complete loss of scale and socket cells in the wing. While *ASH1* and *ASH2* appear to have different expression patterns within the larval wing disc, the two homologs appeared to be mostly co-expressed in the pupal wing scale precursor cells, suggesting a potential redundancy in expression and function for scale cell patterning. We additionally attempted to knockout *ASH1* via CRISPR/Cas9 but did not generate any clear mosaic mutant adults, which could point to functional redundancy of the two homologs, which would echo similarities with *achaete* and *scute* being functionally redundant for bristle formation in *D. melanogaster*. Interestingly, we were unable to identify expression of the *ASH3* gene via RNA-seq in either *A. polyphemus* or *J. coenia*, and despite designing probes for HCR in situ hybridization, were unable to observe apparent expression within the wing disc or pupal wings. Future work should determine if *ASH3* is expressed in different developmental contexts, such as embryonic CNS or PNS development.

In the last common ancestor of arthropods, a single *achaete scute* homolog was likely the predominant proneural gene for sense organ development and that *ASH* likely endowed epidermal cells with the potential to develop into external sense organs without simultaneously specifying subtype identity. Other genes may have been subsequently recruited to regulate early processes of cell specification, such as decisions to form accessory (e.g. socket, shaft, sheath cells) or neuronal (e.g. neurons, glia) types. Future work utilizing lineage tracing via live-imaging and functional assessment of additional neurogenic regulators in Lepidoptera should help to reveal the series of cell divisions and molecular toolkit involved in scale cell development that results in a scale cell and socket cell, and why neuronal cell types do not form in this type of SOP lineage.

Cell cycle progression and differentiation must be highly coordinated for proper development, which relies primarily on the activity of cyclin-dependant kinases (Cdk) that are regulated by their association with factors such as cyclins or cyclin kinase inhibitors. Cyclins are the best-known positive regulators of cell proliferation, and their molecular mechanisms in the cell-cycle transition are conserved in eukaryotes. Similar to the *Drosophila* bristle lineage, scale cell development in Lepidoptera is likely closely related to the cell cycle progression. Our RNA-seq data suggests that several G2/mitotic-specific cyclins and Cyclin-dependent kinases, including *cyclin-A*, *cyclin-B*, *cyclin-E* and *Cyclin-dependent kinase 4* are highly expressed at the scale precursor stage in butterflies and moths, implying their conserved role in coordinating cell cycle progression that lead to highly polyploid differentiated scale and socket cells in the wings. Previous work on Lepidoptera scale cells have suggested that ploidy levels correlate with scale size and pigmentation state, in which higher ploidy results in larger, darker scales (Cho and Nijhout, 2013; Iwata and Otaki, 2016). Recent evidence suggests that *cortex*, a gene with homology to a family of cell cycle regulators, is a key scale cell specification gene, but it remains unclear how it could act to modulate ploidy levels or scale identity (Livraghi et al., 2021). The *cortex* gene could potentially function to modulate ploidy levels by inducing endoreplication cycles in developing scale cells,

but further investigations are required to clarify if there is a direct a causal relationship between ploidy state and scale size/type.

The Cut protein is localized in the nuclei of all four types of cells composing campaniform sensilla during the developmental stages and has an instructive role in execution of an external-type peripheral sense organ program in *D. melanogaster* (Blochlinger et al., 1990, 1991). Several additional genes including *cut, futsch, embryonic lethal abnormal vision (elav),* and *Suppressor of Hairless (Su(H))* are expressed in the campaniform sensilla on pupal wings in *D. melanogaster* (Koseki et al., 2021). We observed *cut* expression in pupal wings that appeared to correlate with subsets of scale precursor cells that become sensory hairs on the hindwing, suggesting that *cut* has maintained an ancestral role in SOP subtype specification but plays a novel role in a modified scale cell type in Lepidoptera wings.

Neurogenic factors and the role of *Notch* during pupal wing development in Lepidoptera

The *Notch* gene encodes a transmembrane protein that functions as a receptor of intercellular signals controlling cell fate choices in numerous developmental processes. The stereotyped arrangement of sensory bristles on the thorax of *Drosophila* arises from a self-organized process in which inhibitory *Notch* signaling both delimits proneural stripes and singles out sensory organ precursor cells (SOPs). In proneural clusters, the *Delta-Notch* signaling pathway prohibits adoption of the neural fate by repressing expression of the proneural genes in cells not adopting the neural fate. Additionally, loss-of-function alleles of *Notch* result in thickened wing veins, whereas *Notch* gain-of function alleles cause thinner and incomplete veins in *Drosophila* and in other species (de Celis and García-Bellido, 1994). Our observations of *Notch* CRISPR mosaic knockouts in butterflies are in line with the functional role of this gene in lateral inhibition to establish and limit fates for vein cells and scale cells within the developing wing. Once proneural clusters or stripes have formed, inhibitory cell–cell interactions mediated by *Notch* restrict the potential to become an SOP to one or a few cells per cluster (or stripe) (Simpson, 1990). *Notch* inhibits the competence to become neural via the E(spl)-HLH family of transcriptional repressors which act redundantly to antagonize the activity and expression of *Ac* and *Sc* (Delidakis and Artavanis-Tsakonas, 1992; Jennings et al., 1994; Couturier et al., 2019). Therefore, adoption of the SOP fate depends on a balance between the activity of *Ac* and *Sc*, acting synergistically with *Senseless*, and the anti-proneural activity of the E(spl)-HLH proteins. One of the major genetic orchestrators of peripheral nervous system development is *Senseless*, which is both necessary and sufficient for SOP differentiation. In our RNA-seq data, we observed high expression levels of *Senseless* at the scale precursor stage, indicating a conserved role in lepidopteran scale cell formation.

Delta (*Dl*) is expressed in microchaeta proneural cells in *Drosophila*, is detected prior to the onset of *achaete* expression and arises normally even in the absence of *achaete/scute* function, indicating that initial *Delta* expression in the notum is not dependent on proneural gene function (Parks et al., 1997). Within microchaeta proneural stripes, *Delta-Notch* signaling prohibits adoption of the SOP fate by repressing the expression of proneural genes. *Dl* expression is detected at high levels

within proneural stripes and at low levels within interstripes. Conversely, *Notch* expression is higher in interstripes than in stripes. These asymmetries suggest that stripe cells (expressing high levels of *Delta*) may interact with adjacent interstripe cells (expressing high levels of *Notch*), resulting in higher levels of *Delta-Notch* signaling activity in interstripe cells than in stripe cells. This asymmetric expression may be essential for normal repression of *achaete* expression in interstripe but not stripe cells. Recent work has demonstrated that prior to the onset of proneural gene expression, a bimodal gradient of *Delta* expression drove *Notch* activity; as a result of cis-inhibition, *Notch* was activated only in regions of intermediate *Delta* levels. This defined an inhibitory template for a first series of three proneural stripes. The spatial pattern of *Notch* activity dynamically changed as these first proneural stripes emerged, forming a negative template for a second group of intercalating stripes. Eventually, each stripe resolved into a row of isolated SOPs through *Notch* signaling. Thus, *Notch* mediated both proneural stripe patterning and SOP selection via a self-organized process (Corson et al., 2017). A dynamic balance between proneural factors and Enhancer of split-HLH (E(spl)-HLH) *Notch* targets underlies patterning. Recent work has identified two classes of partially redundant E(spl)-HLH factors, whose expression both precedes and delimits proneural activity, and is dependent on proneural activity and required for proper SOP spacing within the stripes, respectively (Couturier et al., 2019).
The gene *u-shaped* (*ush*) encodes a novel zinc finger that acts as a transregulator of *achaete* and *scute* in *D. melanogaster*, which is required to pattern the bristles on the dorsal notum. Mutants for *u-shaped* display additional dorsocentral and scutellar bristles that result from overexpression of *achaete* and *scute*. In contrast, overexpression of *u-shaped* causes a loss of *achaete–scute* expression and consequently a loss of dorsal bristles (Cubadda et al., 1997). Based on our RNA-seq data, we observe high levels of *ush* expression in the epithelial stage of wing development, which then appears to drop off by the scale precursor stage in both *J. coenia* and *A. polyphemus*. Based on its role in regulating the position of proneural ac/sc genes in *D. melanogaster*, *ush* would be an interesting candidate to assess functionally in Lepidoptera.

The transcription factor prospero plays a fundamental role in establishing binary SOP sibling cell fates without being asymmetrically localized during SOP division. In the wild-type SOP lineage, the SOP produces the IIa and IIb precursors; IIb divides first to generate a neuron and a sheath cell; IIa divides soon after to generate a hair cell and socket cell. Prospero (P) is first detected at low levels in the IIb cell just prior to its division and it is inherited by both neuron and sheath cell; the sheath cell maintains a high level of Prospero, while the neuron has transient Prospero. In the prospero mutant SOP lineage there is a IIb-to-IIa cell fate transformation, resulting in duplicate external cell types (hair/socket) and loss of internal cell types (neuron/sheath cell) (Manning and Doe, 1999). Based on our RNA-seq data, we observe high levels of *prospero* expression in the epithelial stage of wing development, which then appears to drop off at the scale precursor stage in both *J. coenia* and *A. polyphemus*. Based on its role in determining cell fate in the external sense organ lineage in *D. melanogaster*, *prospero* would be an interesting candidate to assess functionally in Lepidoptera to determine if it plays a role in determining scale cell identity in the wings.

In butterflies during early pupal development, the receptor molecule *Notch* is expressed in a grid-like pattern in the wing epithelium (Fig 5, Reed, 2004). This may

contribute to the parallel rows of uniformly spaced scale precursor cells, which express a homolog of the *achaete-scute* proneural transcription factors that likely function in scale precursor cell differentiation. There is a regulatory loop between *Notch* and *Delta* that is under the transcriptional control of the *E(spl)-C* and *AS-C* genes in *D. melanogaster*, and based on our expression and functional data, these factors appear to be conserved in Lepidoptera wing scale patterning. What has diverged between these insect lineages appears to be the spatial regulation of these molecular pathways in Lepidoptera to produce scale precursor cells across the entire wing.

Cytoskeletal and cuticular factors during scale cell morphogenesis

The morphogenesis of sensory bristles in *D. melanogaster* relies on the function of actin and microtubule cytoskeleton during pupal development. For instance, actin bundles outlining the cell periphery are essential for bristle for elongation and curvature (Tilney et al., 2000a). Actin cross linking proteins, such as *forked* and *singed*, are required to bundle the actin filaments that support elongating bristle cells. The forked protein initiates the process by assembling actin bundles into a loose array which are secondarily assembled by cross-linking into a highly periodic crystalline array by *Fascin* (encoded by *singed*). Mutants of singed and forked lead to morphological defects in the adult bristle including reduction in length and improper shape (Tilney et al., 1995, 2000b, 2000a). The protein *Javelin* may also play a role in bundle maintenance or assembly. Bundles without Javelin splinter and become fragmented, resulting in poorly curved bristles with an inflated, spear-like tip (Shapira et al., 2011). *Stubble-stubbloid* is required for hormone-dependent epithelial morphogenesis of imaginal discs of *Drosophila*, including the formation of bristles, legs, and wings. For instance, mutations affect the organization of microfilament bundles during bristle morphogenesis and it has been characterized as a transmembrane serine protease with potential influence of normal actin organization and cleavage of the extracellular membrane (Appel et al., 1993). The morphological differences between scales and bristles likely result from components or modifiers of the cytoskeleton and cell membrane. Butterfly and moth wing scales are the remains of secreted chitin laid down by retracting trichogen cells during pupal wing development. Scales can reliably produce intricate structures down to the nanometer level across their surfaces thousands of times over in each wing. How a cell is able to produce such a complex shape is largely unknown, but recent studies point to the importance of accumulations of actin filaments, termed ‘rods’, that prefigure ridges along the abwing surface. In butterflies, F-actin bundles are degraded once the ridges have grown at approximately 60% of pupal development (Dinwiddie et al., 2014). How these rods are assembled from individual filaments and the relative contribution of other cytoskeletal components, such as microtubules, is still unclear, but our results suggest that cytoskeletal organizers such as *forked*, *fascin*, *javelin* and *stubble* are highly expressed during early wing scale development, suggesting they play roles in scale morphogenesis, as they do in insect bristle development.

Microtubules (MTs) are dynamic polymers with critical roles in many biological processes such as cell division, organelle trafficking, cell polarity and morphogenesis. The dynamic organization and stability of microtubules is thought to be regulated primarily by microtubule-associated proteins (MAPs), attached either along the

microtubule wall or at microtubule ends (Bitan et al., 2010). We identified the expression of several microtubule associated proteins with dynamic expression during pupal wing development. This includes *Jupiter*, *futsch*, *Mink*, *Fascetto* and *Spindly* that displayed higher levels of expression at 'scale early' timepoints, suggesting that these may play a role in scale morphogenesis. Previous studies have identified correlations between microtubule distribution and scale shape. For instance, during wing development, bristle-like scale morphologies predominantly contain vertically-oriented microtubules that run parallel to actin filaments and the direction of growth (Tilney et al., 2000a; Pomerantz et al., 2021). On the other hand, flat scale morphologies contain microtubules that 'fan out' in various orientations. Our study points to promising microtubule associated candidate genes that require functional studies to assess their potential contribution to scale morphogenesis.

The evidence present in the *Drosophila* literature suggests that ZPs act to connect the chitinous ECM to the plasma membrane. Several ZP domain proteins are important for cuticle formation during insect development, including *dusky-like*, *dusky*, *miniature*, *trynity*, and *morpheyus*. For instance, mutant embryos for ZP genes *dusky* have tracheal development defects, which are similar to mutants affecting chitin secretion like *kkv*, a chitin synthase (Jaźwińska et al., 2003). These genes were found to be expressed at high levels during wing development in *D. melanogaster* (Sobala and Adler, 2016). Likewise, we identified many of these genes expressed at their highest levels at the early wing scale stage in both *J. coenia* and *A. polyphemus*, suggesting that ZP domain genes play evolutionarily conserved roles in cuticle formation during lepidopteran development as well.

Cuticular proteins (CPs) and chitin are produced by scale cells during development and form the overall architecture of what remains as the hard, external adult wing scale. Recent work by (Liu et al., 2021) has contributed to our understanding of cuticular protein (CP) composition and function in lepidopteran scales, including a distinctive class of histidine rich (His-rich) CPs. We identified several of these CPs in our RNAseq dataset, further highlighting the conserved role of cuticular proteins in lepidoptera wing formation. Functional studies using RNAi revealed CPs with different histidine content played distinct roles in constructing the microstructure of the scale surface. Knockdown experiments of His-rich CPs and BmlacA2 resulted in damaged compound eyes and mandibular bristles, as well as scale abnormalities, with notable effects on the ridges and microribs, highlighting that CPs identified in wing scales also play important roles in the formation of cuticle structures in other organs (Liu et al., 2021). In addition to cuticle protein genes, a number of other genes expressed at relatively high and varying levels in pupal wings are known to be important for normal cuticle deposition, such as actin proteins and chitinases. We also identified expression of cuticle and chitin-synthesis related genes that showed conserved expression profiles in the butterfly and moth. These genes were primarily upregulated at the 'scale late' developmental stage, suggesting that they play a role in scale morphogenesis through chitin processing. This includes genes such as chitin deacetylase, chitinase and chitin synthase (encoded by the gene *krotzkopf verkehrt* (*kkv*)), which protects chitin from degradation. Similar to previous RNAseq results from insect pupal wing tissue, we also identified the expression several genes involved in pigmentation and sclerotization, including *pale, ebony, Dopa decarboxylase*, and *yellow* (Sobala and Adler, 2016; Zhang

et al., 2017). In addition to pigmentation processes that generate color, these genes are also important for the synthesis of normal cuticle and in some cases forming specific wing scale morphologies (Zhang et al., 2017; Matsuoka and Monteiro, 2018). Still, much remains to be tested with regard to direct interactions of cytoskeletal factors, CPs and chitin and their contributions to scale morphology.

Signaling molecules, neural regulators and the scaleless wing pattern in the giant silkmoth

Fruit fly species have served as models to understand the regulatory evolution of wing patterns. For instance, the polka-dotted fruit fly *Drosophila guttifera* contains wing patterns in which dark melanic spots form at specific positions, such as around mechanosensory organs (campaniform sensilla) and around crossveins. This wing color pattern has been shown to be induced by the expression of the Wnt signaling gene *wingless*, which then positively regulates pigmentation genes such as *yellow* through the evolutionary gain of novel enhancer activities (Werner et al., 2010; Koshikawa et al., 2015). However *Drosophila melanogaster* also expresses *wingless* around the crossvein and yet there are no wing pigmentation patterns, and it has also been shown that expression of the melanin synthesis gene *yellow* is not sufficient to induce pigmentation in the *D. melanogaster* wing (Gompel et al., 2005; Fukutomi et al., 2021). It has therefore been proposed that the novel pigmentation pattern of *D. guttifera* emerged through multistep co-options of multiple gene regulatory networks, signaling pathways, and effector genes, potentially including *Notch* signaling genes, melanin synthesis genes and Wnt signaling genes (Fukutomi et al., 2021).

In our study, it is interesting to find some of the same signaling pathways expressed in the wing patterns of the giant silkmoth *Antheraea polyphemus*. The Wnt signaling gene *wingless* was indeed expected, as it is known to be expressed by cells at the discal crossvein in other insects and act as a morphogen, in which diffusible intercellular signal transduction molecules are secreted to the surrounding epidermal cells and induce expression of genes in the recipient cells (Martin and Reed, 2014). *Wingless* has previously been shown to play a role in wing patterning in butterflies and also shown to be involved in synapse development (Macdonald et al., 2010; Martin and Reed, 2010; Özsu et al., 2017; Koseki et al., 2021). *wingless* has also been shown to be associated with eyespot formation during pupal wing development *Bicyclus anynana* butterflies, as well as pigmentation patterns in the silkworm, *Bombyx mori* (Yamaguchi et al., 2013; Özsu et al., 2017). Our results confirm that *wingless* is expressed at high levels surrounding the crossvein in *A. polyphemus* pupa wings, diffusing outward reminiscent of a classical morphogen, and is associated with the window scaleless pattern found in many saturniids. The question becomes, how did this novel scaleless wing pattern evolve, and did co-option events occur in a parallel fashion to what has been identified in the polka-dotted fruit fly?

One such gene in our window vs non-window comparisons caught our attention, the bHLH transcriptional repressor gene, *hairy*. In *D. melanogaster*, *hairy* has been shown to act as an antagonizing factors that regulates the transcription of *achaete* and *scute* by binding to their upstream regulatory sequences and repress their transcriptional activity (Skeath and Carroll, 1991; Ohsako et al., 1994; Van Doren et al.,

1994). In *hairy* mutants, ectopic sensory bristles are produced in *D. melanogaster*, indicating that *hairy* represses sensory organ formation by directly repressing transcription of proneural genes (Ohsako et al., 1994). From what we have gathered in the literature, *hairy* is not known to be expressed at high levels surrounding the crossvein in other insect species (Fukutomi et al., 2021). Therefore, our findings of the development and transcriptional activities in the 'window' and 'non-window' regions of the wing in *A. polyphemus* point to a unique process of scale cell pattern formation in which a proneural repressor gene was putatively co-opted, resulting in suppression of scale cells in the 'window' region of some saturniid moths. Future work could shed additional light on this novel patterning process by functionally characterizing *hairy* in a lepidopteran system and investigating the evolution of crossvein-associated pattern formation in a more comparative context using multiple species.

Did shifting spatial expression patterns of conserved proneural and neurogenic pathways lead to the evolution of wing scales in Lepidoptera?

A long-standing problem in evolutionary biology is how genetic variation arises within populations and evolves to make species anatomically different. Many of the morphological differences in body plans between animal groups are thought to result from changes in gene expression during development. For instance, the genes *scute*, *pannier*, *u-shaped*, *hairy* and *extramacrochaetae* are functionally conserved in regulating the development of macrocheates in two distantly related species of fly, *Drosophila melanogaster* and *Musca domestica*. However, the mRNA spatial expression patterns of these genes differ during the larval development stage between the two species, leading to species-specificity in how the machochaete bristles are patterned (Liang et al., 2018). Our findings of dynamic expression patterns associated with lepidopteran wing scale cell development contributes to the growing set of observations suggesting that shifting temporal dynamics of conserved genes represents a mechanism through which different morphologies can be achieved (e.g. Sakamoto et al., 2009; Lemmon et al., 2016).

Conclusion

This work presents a comparative transcriptome analysis between two lepidopterans, toward the identification of biological processes and functional gene activities on scale cell formation that occur during the early stages of pupal development. Moreover, it provides information on the molecular processes contributing to the diversification of cell specification and patterning, a central topic of evolutionary research. Our dataset suggests potential mechanisms that drive scale cell development, such as shifting temporal dynamics of conserved neurogenesis pathways, which can help facilitate the characterization of scale cell evolution and provide insight into how changes in gene expression during wing development play a role in cellular morphological diversity.

Materials and methods

Insect husbandry

Junonia coenia butterflies were derived from a laboratory colony maintained at Duke University (gift of Laura Grunert and Fred Nijhout), maintained in a 16:8 hr light/dark cycle at 26°C, 60% relative humidity, were fed on a multi-species artificial diet (Southland) supplemented with dried leaves of the host plant *Plantago lanceolata*, and were induced to oviposit on fresh leaves *of P. lanceolata.* Pupae of *Antheraea polyphemus* were obtained from breeders (Bill Oehlke and suppliers) and temporarily stored at 4°C to break diapause.

RNA isolation and sequencing

For RNA isolations, we performed micro-dissections of pupal forewing tissue immersed in ice-cold PBS (taking care to isolate tissue of interest and avoid cuticle or wing veins) and stored tissue in Trizol reagent (Invitrogen) at −80°C until downstream extraction steps. For *Junonia coenia*, we performed three biological replicates for each developmental stage, which consisted of: ~8-12 hours after pupal development (APF) when wing tissue is composed of undifferentiated epithelial cells, ~24 h APF when scale precursor cells have differentiated, ~72 h APF when early scale projections have elongated, and ~120 h APF when late scale projections are maturing. For *A. polyphemus*, pupal development proceeded at more viable rates, so we performed wing tissue dissections for RNA isolation on half of the dissected wings and, for the other half of the wings, fixed and stained the tissue with DAPI and phalloidin (as in Dinwiddie et al. 2014) to confirm approximately equivalent developmental stages via fluorescence microscopy, selecting two replicates each for epithelial, scale early, and scale late stages. Total RNA was isolated using the Direct-zol RNA miniprep kit (Zymo Research), dissolved in nuclease-free water and RNA quantity and integrity was evaluated using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). The cDNA library was then prepared using the SMART-Seq v4 Ultra Low Input RNA Kit (Clonetech) at QB3 Genomics (Berkeley, CA). Sequences for *A. polyphemus* were generated using a HiSeq4000 at 100 base paired-end reads and sequences for *J. coenia* were generated using a NovaSeq at 150 base paired-end reads (Illumina Inc.). We additionally sequenced the same full-length cDNA that was generated for the Illumina short read RNA-seq using the Oxford Nanopore MinION R9.4 flow, Native Barcoding Kit, LSK109 library preparation kit according to manufacturer's protocols (Oxford Nanopore Technologies Inc.).

Assembly and analysis of reference transcriptomes

Quality assessment of raw paired-end reads were conducted using FastQC and quality trimming and adaptor removal on raw reads were performed with Trimmomatic (Bolger et al., 2014). The paired-end reads were merged, in silico normalized and *de novo* assembled using Trinity v.2.5.1 using the software's default parameters (Grabherr et al., 2011; Haas et al., 2013). The assembled sequences identified by Trinity were filtered and the longest transcript was retained. Cluster database at high identity with tolerance (CD-HIT) was used for further clustering and redundancy removal with minimum similarity cut-off of 95% (Li and Godzik, 2006). For *J. coenia*, we used the CDS v1.0

dataset obtained from LepBase, which contains 19,237 transcripts (www.lepbase.org). One sample from *J. coenia* at the 'epithelial' stage was omitted, as it was an outlier between the remaining 'epithelial' and 'scale precursor' samples, perhaps due to the fact that it was an intermediate development stage. For *A. polyphemus*, the Trinity assembler generated 144,206 transcripts and CD-HIT reduced this further to 119,685 transcripts used for downstream read mapping. Reads for each library were aligned and transcript abundances estimated using the RSEM pipeline script included with Trinity, which converts the total number of aligned reads to an estimated number of transcripts in the library while accounting for transcript length and total number of reads in the library (Li and Dewey, 2011; Haas et al., 2013). Sample information and alignment statistics are reported in Supplementary Table 1. Expression count data were analyzed for differential expression with DESeq2 using the Wald test for significant log(fold change) in expression between conditions (here, developmental stages) using DEBrowser v1.16.3 (Love et al., 2014; Kucukural et al., 2019). Differential expression was determined with an FDR-adjusted P-values cut-off <0.05 and |log2FoldChange| ≥ 1. Hierarchical clustered heatmaps and principal component analysis (PCA) was used to assess normalized transcriptomic read profiles of each sample.

Gene annotations and gene ontology enrichment

Genomes and annotations for *J. coenia* are publicly available at www. Lepbase.org (van der Burg et al., 2019). TransDecoder, a package included in the Trinity software, was used for the prediction of coding regions in *A. polyphemus* (Haas et al., 2013). Orthology to *Drosophila melanogaster* genes were established first by reciprocal best hits using BLASTX and BLASTP (cutoff E value of 1e-5) (Camacho et al., 2009). Where reciprocal BLAST was not determinative, OrthoFinder was used, with the input being the *Drosophila* reference proteome from UniProt (downloaded January, 2020) and the annotated proteins of *J. coenia* and *A. polyphemus* (Emms and Kelly, 2019). In some cases where similar copies of a gene are present, for example the Achaete-scute homologs, genes were manually curated to ensure their correct homology. See Supplementary Table 1 for the list of gene names and transcript IDs for each species. Gene ontology (GO) enrichment analysis (Huang et al., 2009) was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (Dennis et al., 2003). Unique *D. melanogaster* orthologs of identified differentially expressed genes served as the query gene list and *D. melanogaster* orthologs to Lepidoptera were used as the universal background list.

Wing fixation, Hybridization Chain Reaction (HCR) *in situ* hybridization (ISH) and immunohistochemistry

Wing discs and pupal wings were dissected and fixed with 3.2% paraformaldehyde in PEM buffer for 30 min at room temperature, as described previously (Dinwiddie et al., 2014). Fixed wings were incubated in 1X PBS+0.1% Triton-X 100 (PT) with 1:200 dilution of phalloidin, Alexa 555 conjugated (Invitrogen A34055), and wheat germ agglutinin, Alexa 647 conjugated (Invitrogen W32466) at a dilution of 1:200 overnight at

4°C. Wings were washed in PT and then placed in 50% glycerol:PBS with DAPI overnight at 4°C.

For Hybridization Chain Reaction (HCR) in situ hybridization, version 3.0 HCR probes (20 pairs per gene) and fluorescently-labelled hairpins were produced by Molecular Instruments, Inc. for *ASH1*, *ASH2*, *cut* and *Notch* in *J. coenia* and for *wingless* and *hairy* in *A. polyphemus*. Additional probes were designed using the insitu_probe_generator software (https://github.com/rwnull/insitu_probe_generator) for *Delta* and enhancer of split in *J. coenia*. All required buffers were made according to the instructions provided by Molecular Instruments and HCR ISH was then carried out as per the Molecular Instruments HCR v3.0 protocol for whole-mount fruit fly embryos with some modifications (Choi et al., 2018; Bruce et al., 2021). For immunohistochemistry, subsequent to fixation, washes were carried out in wash buffer (0.1% Triton-X 100 in PBS) before blocking the wings at 4˚C in block buffer (0.05 g bovine serum albumin, 10 ml PBS 0.1% Triton-X 100). Wings were then incubated in primary antibodies against C17.9C6 *Notch* antibody (1:30) at 4˚C overnight, washed, and added in secondary antibody (1:500, goat anti-mouse IgG, AlexaFlour 555, ThermoFisher Scientific). Before mounting, wings were incubated in 50% glycerol and 1.0µg/mL DAPI solution, then tissue was transferred into 70% glycerol. Slides of fixed tissue were examined with an LSM 880 confocal microscope (Carl Zeiss, Germany).

Cas9-mediated genome editing / Design of sgRNA targets

Target sequences of *J. coenia* were obtained from LepBase (van der Burg et al., 2019). sgRNAs were designed against 5’ ends of genes in order to produce a frameshift mutation due to imperfect non-homologous end joining repair. Target guide RNAs were generated using ChopChop software (https://chopchop.cbu.uib.no/) and were identified by searching for N_{20}NGG patterns targeting the first exon (or generally near the 5’ end of the coding sequence), then tested for uniqueness by BLAST against the genome or transcriptome reference. sgRNAs were ordered from Synthego and mixed with Cas9 prior to injection. Injection mixes had a final concentration of 333ng/uL Cas9 protein, 150ng/uL sgRNA. Butterfly eggs laid on host plant leaves were collected after 1–4 hours (following procedures described in Zhang et al., 2017). Eggs were then glued to a glass slide and mounted with the micropyle facing up. CRISPR mixtures containing preassembled sgRNAs and recombinant Cas9 protein were injected using pulled glass needles. Post injection, butterfly eggs were rested in a humidity chamber until hatching and addition of artificial diet. For genotyping, DNA was extracted from mutant leg tissue and amplified using oligonucleotides flanking the sgRNAs target region. PCR amplicons were column purified and run on a Fragment Analyzer (Agilent).

Imaging

Wings were imaged with the Keyence VHX-5000 digital microscope at 50X-500x on a VH-Z00T lens.

Data availability

The datasets generated and analyzed in the current study will be available in the NCBI SRA repository upon publication.

Acknowledgements

This work was supported by a grant from the Human Frontier Science Program (RGP0014/2016). We thank UC Berkeley QB3 Genomics for library preparation and sequencing assistance, as well as Katie Sanko for assistance with insect husbandry and egg microinjections. We thank Noah Whiteman, Craig Miller, Peter Sudmant, Ryan Null and Rachel Thayer their thoughtful comments and suggestions on the manuscript.

References

1. Appel, L. F., Prout, M., Abu-Shumays, R., Hammonds, A., Garbe, J. C., Fristrom, D., et al. (1993). The *Drosophila* Stubble-stubbloid gene encodes an apparent transmembrane serine protease required for epithelial morphogenesis. *Proceedings of the National Academy of Sciences* 90, 4937–4941. doi:10.1073/pnas.90.11.4937.
2. Ayyar, S., Negre, B., Simpson, P., and Stollewerk, A. (2010). An arthropod cis-regulatory element functioning in sensory organ precursor development dates back to the Cambrian. *BMC Biol* 8, 127. doi:10.1186/1741-7007-8-127.
3. Bitan, A., Guild, G. M., Bar-Dubin, D., and Abdu, U. (2010). Asymmetric Microtubule Function Is an Essential Requirement for Polarized Organization of the *Drosophila* Bristle. *Mol Cell Biol* 30, 496–507. doi:10.1128/MCB.00861-09.
4. Blochlinger, K., Bodmer, R., Jan, L. Y., and Jan, Y. N. (1990). Patterns of expression of cut, a protein required for external sensory organ development in wild-type and cut mutant *Drosophila* embryos. *Genes & Development* 4, 1322–1331. doi:10.1101/gad.4.8.1322.
5. Blochlinger, K., Jan, L. Y., and Jan, Y. N. (1991). Transformation of sensory organ identity by ectopic expression of Cut in *Drosophila*. *Genes & Development* 5, 1124–1135. doi:10.1101/gad.5.7.1124.
6. Bodmer, R. (1987). Transformation of sensory organs by Mutations of the cut locus of *D. melanogaster*. *Cell* 51, 293–307. doi:10.1016/0092-8674(87)90156-5.
7. Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi:10.1093/bioinformatics/btu170.
8. Bruce, H., Jerz, G., not provided, S., not provided, J., Pomerantz, A., Senevirathne, G., et al. (2021). Hybridization Chain Reaction (HCR) In Situ Protocol v1. doi:10.17504/protocols.io.bunznvf6.
9. Calleja, M., Renaud, O., Usui, K., Pistillo, D., Morata, G., and Simpson, P. (2002). How to pattern an epithelium: lessons from achaete-scute regulation on the notum of *Drosophila*. *Gene* 292, 1–12. doi:10.1016/S0378-1119(02)00628-5.
10. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10, 421. doi:10.1186/1471-2105-10-421.

11. Campos-Ortega, J. A. (1995). Genetic mechanisms of early neurogenesis in*Drosophila melanogaster*. *Mol Neurobiol* 10, 75–89. doi:10.1007/BF02740668.
12. Cho, E. H., and Nijhout, H. F. (2013). Development of polyploidy of scale-building cells in the wings of Manduca sexta. *Arthropod Structure & Development* 42, 37–46. doi:10.1016/j.asd.2012.09.003.
13. Choi, H. M. T., Schwarzkopf, M., Fornace, M. E., Acharya, A., Artavanis, G., Stegmaier, J., et al. (2018). Third-generation *in situ* hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* 145, dev165753. doi:10.1242/dev.165753.
14. Corson, F., Couturier, L., Rouault, H., Mazouni, K., and Schweisguth, F. (2017). Self-organized *Notch* dynamics generate stereotyped sensory organ patterns in *Drosophila*. *Science* 356, eaai7407. doi:10.1126/science.aai7407.
15. Couturier, L., Mazouni, K., Corson, F., and Schweisguth, F. (2019). Regulation of *Notch* output dynamics via specific E(spl)-HLH factors during bristle patterning in *Drosophila*. *Nat Commun* 10, 3486. doi:10.1038/s41467-019-11477-2.
16. Cubadda, Y., Heitzler, P., Ray, R. P., Bourouis, M., Ramain, P., Gelbart, W., et al. (1997). u-shaped encodes a zinc finger protein that regulates the proneural genes achaete and scute during the formation of bristles in *Drosophila*. *Genes & Development* 11, 3083–3095. doi:10.1101/gad.11.22.3083.
17. Cubas, P., de Celis, J. F., Campuzano, S., and Modolell, J. (1991). Proneural clusters of achaete-scute expression and the generation of sensory organs in the *Drosophila* imaginal wing disc. *Genes & Development* 5, 996–1008. doi:10.1101/gad.5.6.996.
18. Cubas, P., and Modolell, J. (1992). The extramacrochaetae gene provides information for sensory organ patterning. *The EMBO Journal* 11, 3385–3393. doi:10.1002/j.1460-2075.1992.tb05417.x.
19. Day, C. R., Hanly, J. J., Ren, A., and Martin, A. (2019). Sub-micrometer insights into the cytoskeletal dynamics and ultrastructural diversity of butterfly wing scales. *Developmental Dynamics* 248, 657–670. doi:10.1002/dvdy.63.
20. de Celis, JoséF., and García-Bellido, A. (1994). Roles of the *Notch* gene in *Drosophila* wing morphogenesis. *Mechanisms of Development* 46, 109–122. doi:10.1016/0925-4773(94)90080-9.
21. Delidakis, C., and Artavanis-Tsakonas, S. (1992). The Enhancer of split [E(spl)] locus of *Drosophila* encodes seven independent helix-loop-helix proteins. *Proceedings of the National Academy of Sciences* 89, 8731–8735. doi:10.1073/pnas.89.18.8731.
22. Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., et al. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol* 4, R60. doi:10.1186/gb-2003-4-9-r60.
23. Deshmukh, R., Baral, S., Gandhimathi, A., Kuwalekar, M., and Kunte, K. (2018). Mimicry in butterflies: co-option and a bag of magnificent developmental genetic tricks: Mimicry in butterflies. *WIREs Dev Biol* 7, e291. doi:10.1002/wdev.291.
24. Dinwiddie, A., Null, R., Pizzano, M., Chuong, L., Leigh Krup, A., Ee Tan, H., et al. (2014). Dynamics of F-actin prefigure the structure of butterfly wing scales. *Developmental Biology* 392, 404–418. doi:10.1016/j.ydbio.2014.06.005.

25. Emms, D. M., and Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* 20, 238. doi:10.1186/s13059-019-1832-y.
26. Finet, C., Decaras, A., Armisén, D., and Khila, A. (2018). The *achaete–scute* complex contains a single gene that controls bristle development in the semi-aquatic bugs. *Proc. R. Soc. B.* 285, 20182387. doi:10.1098/rspb.2018.2387.
27. Fukutomi, Y., Kondo, S., Toyoda, A., Shigenobu, S., and Koshikawa, S. (2021). Transcriptome analysis reveals *wingless* regulates neural development and signaling genes in the region of wing pigmentation of a polka-dotted fruit fly. *FEBS J* 288, 115–126. doi:10.1111/febs.15338.
28. Furman, D., and Bukharina, T. (2008). How *Drosophila melanogaster* Forms its Mechanoreceptors. *CG* 9, 312–323. doi:10.2174/138920208785133271.
29. Galant, R., Skeath, J. B., Paddock, S., Lewis, D. L., and Carroll, S. B. (1998). Expression pattern of a butterfly achaete-scute homolog reveals the homology of butterfly wing scales and insect sensory bristles. *Current Biology* 8, 807–813. doi:10.1016/S0960-9822(98)70322-7.
30. García-Bellido, A., and de Celis, J. F. (2009). The Complex Tale of the *achaete–scute* Complex: A Paradigmatic Case in the Analysis of Gene Organization and Function During Development. *Genetics* 182, 631–639. doi:10.1534/genetics.109.104083.
31. Gomez-Skarmeta, J. L., Rodriguez, I., Martinez, C., Culi, J., Ferres-Marco, D., Beamonte, D., et al. (1995). Cis-regulation of achaete and scute: shared enhancer-like elements drive their coexpression in proneural clusters of the imaginal discs. *Genes & Development* 9, 1869–1882. doi:10.1101/gad.9.15.1869.
32. Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A., and Carroll, S. B. (2005). Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature* 433, 481–487. doi:10.1038/nature03235.
33. Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29, 644–652. doi:10.1038/nbt.1883.
34. Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8, 1494–1512. doi:10.1038/nprot.2013.084.
35. Hartenstein, V., and Posakony, J. W. (1990). A dual function of the *Notch* gene in *Drosophila* sensillum development. *Developmental Biology* 142, 13–30. doi:10.1016/0012-1606(90)90147-B.
36. Hinz, U., Giebel, B., and Campos-Ortega, JoséA. (1994). The basic-helix-loop-helix domain of *Drosophila* lethal of scute protein is sufficient for proneural function and activates neurogenic genes. *Cell* 76, 77–87. doi:10.1016/0092-8674(94)90174-0.
37. Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4, 44–57. doi:10.1038/nprot.2008.211.

38. Iwata, M., and Otaki, J. M. (2016). Spatial patterns of correlated scale size and scale color in relation to color pattern elements in butterfly wings. *Journal of Insect Physiology* 85, 32–45. doi:10.1016/j.jinsphys.2015.11.013.
39. Jaźwińska, A., Ribeiro, C., and Affolter, M. (2003). Epithelial tube morphogenesis during *Drosophila* tracheal development requires Piopio, a luminal ZP protein. *Nat Cell Biol* 5, 895–901. doi:10.1038/ncb1049.
40. Jennings, B., Preiss, A., Delidakis, C., and Bray, S. (1994). The *Notch* signalling pathway is required for Enhancer of split bHLH protein expression during neurogenesis in the *Drosophila* embryo. *Development* 120, 3537–3548. doi:10.1242/dev.120.12.3537.
41. Klann, M., Schacht, M. I., Benton, M. A., and Stollewerk, A. (2021). Functional analysis of sense organ specification in the Tribolium castaneum larva reveals divergent mechanisms in insects. *BMC Biol* 19, 22. doi:10.1186/s12915-021-00948-y.
42. Koseki, M., Tanaka, N. K., and Koshikawa, S. (2021). The color pattern inducing gene *wingless* is expressed in specific cell types of campaniform sensilla of a polka-dotted fruit fly, *Drosophila* guttifera. *Dev Genes Evol*. doi:10.1007/s00427-021-00674-z.
43. Koshikawa, S., Giorgianni, M. W., Vaccaro, K., Kassner, V. A., Yoder, J. H., Werner, T., et al. (2015). Gain of *cis* -regulatory activities underlies novel domains of *wingless* gene expression in *Drosophila*. *Proc Natl Acad Sci USA* 112, 7524–7529. doi:10.1073/pnas.1509022112.
44. Kucukural, A., Yukselen, O., Ozata, D. M., Moore, M. J., and Garber, M. (2019). DEBrowser: interactive differential expression analysis and visualization tool for count data. *BMC Genomics* 20, 6. doi:10.1186/s12864-018-5362-x.
45. Lai, E. C., and Orgogozo, V. (2004). A hidden program in *Drosophila* peripheral neurogenesis revealed: fundamental principles underlying sensory organ diversity. *Developmental Biology* 269, 1–17. doi:10.1016/j.ydbio.2004.01.032.
46. Lemmon, Z. H., Park, S. J., Jiang, K., Van Eck, J., Schatz, M. C., and Lippman, Z. B. (2016). The evolution of inflorescence diversity in the nightshades and heterochrony during meristem maturation. *Genome Res.* 26, 1676–1686. doi:10.1101/gr.207837.116.
47. Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 323. doi:10.1186/1471-2105-12-323.
48. Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658–1659. doi:10.1093/bioinformatics/btl158.
49. Liang, Q., Peng, T., Sun, B., Tu, J., Cheng, X., Tian, Y., et al. (2018). Gene expression patterns determine the differential numbers of dorsocentral macrochaetes between *Musca domestica* and *Drosophila melanogaster*. *genesis* 56, e23258. doi:10.1002/dvg.23258.
50. Liu, J., Chen, Z., Xiao, Y., Asano, T., Li, S., Peng, L., et al. (2021). Lepidopteran wing scales contain abundant cross-linked film-forming histidine-rich cuticular proteins. *Commun Biol* 4, 491. doi:10.1038/s42003-021-01996-4.

51. Livraghi, L., Hanly, J. J., Van Bellghem, S. M., Montejo-Kovacevich, G., van der Heijden, E. S., Loh, L. S., et al. (2021). Cortex cis-regulatory switches establish scale colour identity and pattern diversity in Heliconius. *eLife* 10, e68549. doi:10.7554/eLife.68549.
52. Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550. doi:10.1186/s13059-014-0550-8.
53. Macdonald, W. P., Martin, A., and Reed, R. D. (2010). Butterfly wings shaped by a molecular cookie cutter: evolutionary radiation of lepidopteran wing shapes associated with a derived Cut/*wingless* wing margin boundary system: Butterfly wings shaped by a molecular cookie cutter. *Evolution & Development* 12, 296–304. doi:10.1111/j.1525-142X.2010.00415.x.
54. Manning, L., and Doe, C. Q. (1999). Prospero distinguishes sibling cell fate without asymmetric localization in the *Drosophila* adult external sense organ lineage. *Development* 126, 2063–2071. doi:10.1242/dev.126.10.2063.
55. Martin, A., and Reed, R. D. (2010). *wingless* and aristaless2 Define a Developmental Ground Plan for Moth and Butterfly Wing Pattern Evolution. *Molecular Biology and Evolution* 27, 2864–2878. doi:10.1093/molbev/msq173.
56. Martin, A., and Reed, R. D. (2014). Wnt signaling underlies evolution and development of the butterfly wing pattern symmetry systems. *Developmental Biology* 395, 367–378. doi:10.1016/j.ydbio.2014.08.031.
57. Matsuoka, Y., and Monteiro, A. (2018). Melanin Pathway Genes Regulate Color and Morphology of Butterfly Wing Scales. *Cell Reports* 24, 56–65. doi:10.1016/j.celrep.2018.05.092.
58. Mazo-Vargas, A., Concha, C., Livraghi, L., Massardo, D., Wallbank, R. W. R., Zhang, L., et al. (2017). Macroevolutionary shifts of *WntA* function potentiate butterfly wing-pattern diversity. *Proc Natl Acad Sci USA* 114, 10701–10706. doi:10.1073/pnas.1708149114.
59. Monteiro, A. (2015). Origin, Development, and Evolution of Butterfly Eyespots. *Annu. Rev. Entomol.* 60, 253–271. doi:10.1146/annurev-ento-010814-020942.
60. Nagaraj, R., and Adler, P. N. (2012). Dusky-like functions as a Rab11 effector for the deposition of cuticle during *Drosophila* bristle development. *Development* 139, 906–916. doi:10.1242/dev.074252.
61. Nagy, O., Nuez, I., Savisaar, R., Peluffo, A. E., Yassin, A., Lang, M., et al. (2018). Correlated Evolution of Two Copulatory Organs via a Single cis-Regulatory Nucleotide Change. *Current Biology* 28, 3450-3457.e13. doi:10.1016/j.cub.2018.08.047.
62. Negre, B., and Simpson, P. (2009). Evolution of the achaete-scute complex in insects: convergent duplication of proneural genes. *Trends in Genetics* 25, 147–152. doi:10.1016/j.tig.2009.02.001.
63. Nijhout, H. F. (2001). Elements of butterfly wing patterns. *J. Exp. Zool.* 291, 213–225. doi:10.1002/jez.1099.
64. Ohsako, S., Hyer, J., Panganiban, G., Oliver, I., and Caudy, M. (1994). *Hairy* function as a DNA-binding helix-loop-helix repressor of *Drosophila* sensory organ formation. *Genes & Development* 8, 2743–2755. doi:10.1101/gad.8.22.2743.

65. Overton, J. (1966). MICROTUBULES AND MICROFIBRILS IN MORPHOGENESIS OF THE SCALE CELLS OF EPHESTIA KÜHNIELLA. *Journal of Cell Biology* 29, 293–305. doi:10.1083/jcb.29.2.293.
66. Özsu, N., Chan, Q. Y., Chen, B., Gupta, M. D., and Monteiro, A. (2017). *Wingless* is a positive regulator of eyespot color patterns in Bicyclus anynana butterflies. *Developmental Biology* 429, 177–185. doi:10.1016/j.ydbio.2017.06.030.
67. Parks, A. L., Huppert, S. S., and Muskavitch, M. A. T. (1997). The dynamics of neurogenic signalling underlying bristle development in *Drosophila melanogaster*. *Mechanisms of Development* 63, 61–74. doi:10.1016/S0925-4773(97)00675-8.
68. Perez Goodwyn, P., Maezono, Y., Hosoda, N., and Fujisaki, K. (2009). Waterproof and translucent wings at the same time: problems and solutions in butterflies. *Naturwissenschaften* 96, 781–787. doi:10.1007/s00114-009-0531-z.
69. Pomerantz, A. F., Siddique, R. H., Cash, E. I., Kishi, Y., Pinna, C., Hammar, K., et al. (2021). Developmental, cellular, and biochemical basis of transparency in clearwing butterflies. *Journal of Experimental Biology*, jeb.237917. doi:10.1242/jeb.237917.
70. Reed, R. D. (2004). Evidence for *Notch*-mediated lateral inhibition in organizing butterfly wing scales. *Development Genes and Evolution* 214, 43–46. doi:10.1007/s00427-003-0366-0.
71. Roch, F., Alonso, C. R., and Akam, M. (2003). *Drosophila miniature* and *dusky* encode ZP proteins required for cytoskeletal reorganisation during wing morphogenesis. *Journal of Cell Science* 116, 1199–1207. doi:10.1242/jcs.00298.
72. Sakamoto, K., Onimaru, K., Munakata, K., Suda, N., Tamura, M., Ochi, H., et al. (2009). Heterochronic Shift in Hox-Mediated Activation of Sonic hedgehog Leads to Morphological Changes during Fin Development. *PLoS ONE* 4, e5121. doi:10.1371/journal.pone.0005121.
73. Schweisguth, F. (2015). Asymmetric cell division in the *Drosophila* bristle lineage: from the polarization of sensory organ precursor cells to *Notch*-mediated binary fate decision: Asymmetric cell division in an epithelium. *WIREs Dev Biol* 4, 299–309. doi:10.1002/wdev.175.
74. Shapira, S., Bakhrat, A., Bitan, A., and Abdu, U. (2011). The *Drosophila* javelin Gene Encodes a Novel Actin-Associated Protein Required for Actin Assembly in the Bristle. *Molecular and Cellular Biology* 31, 4582–4592. doi:10.1128/MCB.05730-11.
75. Simpson, P. (1990). Lateral inhibition and the development of the sensory bristles of the adult peripheral nervous system of *Drosophila*. *Development* 109, 509–519. doi:10.1242/dev.109.3.509.
76. Skaer, N., Pistillo, D., Gibert, J.-M., Lio, P., Wülbeck, C., and Simpson, P. (2002). Gene duplication at the achaete–scute complex and morphological complexity of the peripheral nervous system in Diptera. *Trends in Genetics* 18, 399–405. doi:10.1016/S0168-9525(02)02747-6.
77. Skeath, J. B., and Carroll, S. B. (1991). Regulation of achaete-scute gene expression and sensory organ pattern formation in the *Drosophila* wing. *Genes & Development* 5, 984–995. doi:10.1101/gad.5.6.984.

78. Sobala, L. F., and Adler, P. N. (2016). The Gene Expression Program for the Formation of Wing Cuticle in *Drosophila*. *PLoS Genet* 12, e1006100. doi:10.1371/journal.pgen.1006100.
79. Tilney, L. G., Connelly, P. S., Vranich, K. A., Shaw, M. K., and Guild, G. M. (2000a). Actin filaments and microtubules play different roles during bristle elongation in *Drosophila*. *Journal of Cell Science* 113, 1255–1265. doi:10.1242/jcs.113.7.1255.
80. Tilney, L. G., Connelly, P. S., Vranich, K. A., Shaw, M. K., and Guild, G. M. (2000b). Regulation of Actin Filament Cross-linking and Bundle Shape in *Drosophila* Bristles. *Journal of Cell Biology* 148, 87–99. doi:10.1083/jcb.148.1.87.
81. Tilney, L. G., Tilney, M. S., and Guild, G. M. (1995). F actin bundles in *Drosophila* bristles. I. Two filament cross-links are involved in bundling. *Journal of Cell Biology* 130, 629–638. doi:10.1083/jcb.130.3.629.
82. Tsai, C.-C., Childers, R. A., Nan Shi, N., Ren, C., Pelaez, J. N., Bernard, G. D., et al. (2020). Physical and behavioral adaptations to prevent overheating of the living wings of butterflies. *Nat Commun* 11, 551. doi:10.1038/s41467-020-14408-8.
83. van der Burg, K. R. L., Lewis, J. J., Martin, A., Nijhout, H. F., Danko, C. G., and Reed, R. D. (2019). Contrasting Roles of Transcription Factors Spineless and EcR in the Highly Dynamic Chromatin Landscape of Butterfly Wing Metamorphosis. *Cell Reports* 27, 1027-1038.e3. doi:10.1016/j.celrep.2019.03.092.
84. Van Doren, M., Bailey, A. M., Esnayra, J., Ede, K., and Posakony, J. W. (1994). Negative regulation of proneural gene activity: *hairy* is a direct transcriptional repressor of achaete. *Genes & Development* 8, 2729–2742. doi:10.1101/gad.8.22.2729.
85. Werner, T., Koshikawa, S., Williams, T. M., and Carroll, S. B. (2010). Generation of a novel wing colour pattern by the *Wingless* morphogen. *Nature* 464, 1143–1148. doi:10.1038/nature08896.
86. Wheeler, S. R., Carrico, M. L., Wilson, B. A., Brown, S. J., and Skeath, J. B. (2003). The expression and function of the *achaete-scute* genes in *Tribolium castaneum* reveals conservation and variation in neural pattern formation and cell fate specification. *Development* 130, 4373–4381. doi:10.1242/dev.00646.
87. Yamaguchi, J., Banno, Y., Mita, K., Yamamoto, K., Ando, T., and Fujiwara, H. (2013). Periodic Wnt1 expression in response to ecdysteroid generates twin-spot markings on caterpillars. *Nat Commun* 4, 1857. doi:10.1038/ncomms2778.
88. Zhang, L., Martin, A., Perry, M. W., van der Burg, K. R. L., Matsuoka, Y., Monteiro, A., et al. (2017). Genetic Basis of Melanin Pigmentation in Butterfly Wings. *Genetics* 205, 1537–1550. doi:10.1534/genetics.116.196451.
89. Zhou, Q., Tang, S., Chen, Y., Yi, Y., Zhang, Z., and Shen, G. (2004). A scaleless wings mutant associated with tracheal system developmental deficiency in wing discs in the silkworm, Bombyx mori. *Int. J. Dev. Biol.* 48, 1113–1117. doi:10.1387/ijdb.041845qz.
90. Zhou, Q., Yu, L., Shen, X., Li, Y., Xu, W., Yi, Y., et al. (2009). Homology of Dipteran Bristles and Lepidopteran Scales: Requirement for the *Bombyx mori*

achaete-scute Homologue *ASH2*. *Genetics* 183, 619–627. doi:10.1534/genetics.109.102848.
91. Zhou, Q., Zhang, T., Xu, W., Yu, L., Yi, Y., and Zhang, Z. (2008). Analysis of four achaete-scute homologs in Bombyx mori reveals new viewpoints of the evolution and functions of this gene family. *BMC Genet* 9, 24. doi:10.1186/1471-2156-9-24.
92. Zhou, Q.-X., Li, Y.-N., Shen, X.-J., Yi, Y.-Z., Zhang, Y.-Z., and Zhang, Z.-F. (2006). The scaleless wings mutant in Bombyx mori is associated with a lack of scale precursor cell differentiation followed by excessive apoptosis. *Dev Genes Evol* 216, 721–726. doi:10.1007/s00427-006-0091-6.

Figures

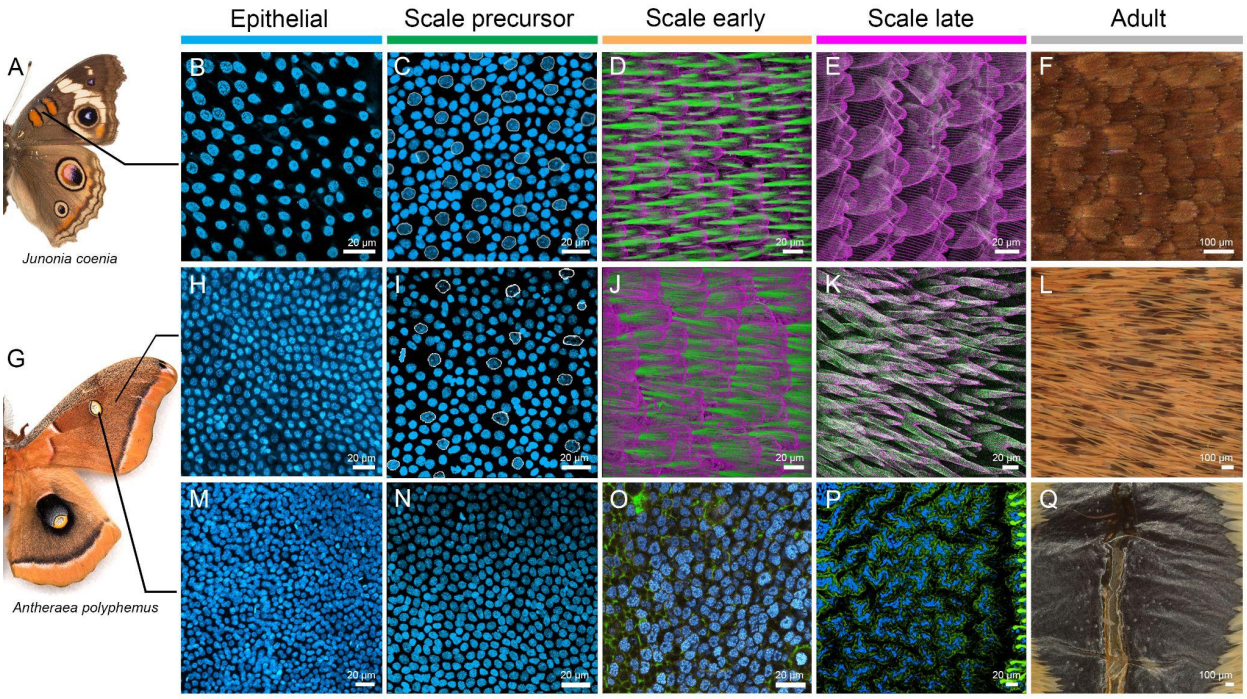


Figure 3.1. Pupal wing developmental stages used for comparative transcriptomic analyses

RNA-seq analysis was performed on the buckeye butterfly *Junonia coenia* and the giant silkmoth *Antheraea polyphemus* at four stages of pupal wing development: epithelial, scale precursor, scale early and scale late. (B) At the 'epithelial stage' the wing is composed of a bilayer of undifferentiated epithelial cells (nuclei of cells marked with DAPI). (C) At the 'scale precursor' stage, certain cells have differentiated into scale precursors (SP cells) outlined in white for better visibility, surrounded by neighboring undifferentiated epithelial cells. (D) At the 'scale early' stage, scale cell projections display large bundles of actin filaments (marked with phalloidin, green) and a round outer membrane (marked with wheat germ agglutinin, magenta). (E) At the 'scale late' stage maturing scale projections contain actin bundles extending to the finger-like tips at the end of the scale and more prominent surface ridges. (M-Q) In the 'window' region (bottom panel) in *A. polyphemus* the scale precursor cells appear to not not initiate, resulting in scale-less membrane surrounding the crossvein region of the wing in the adult moth.

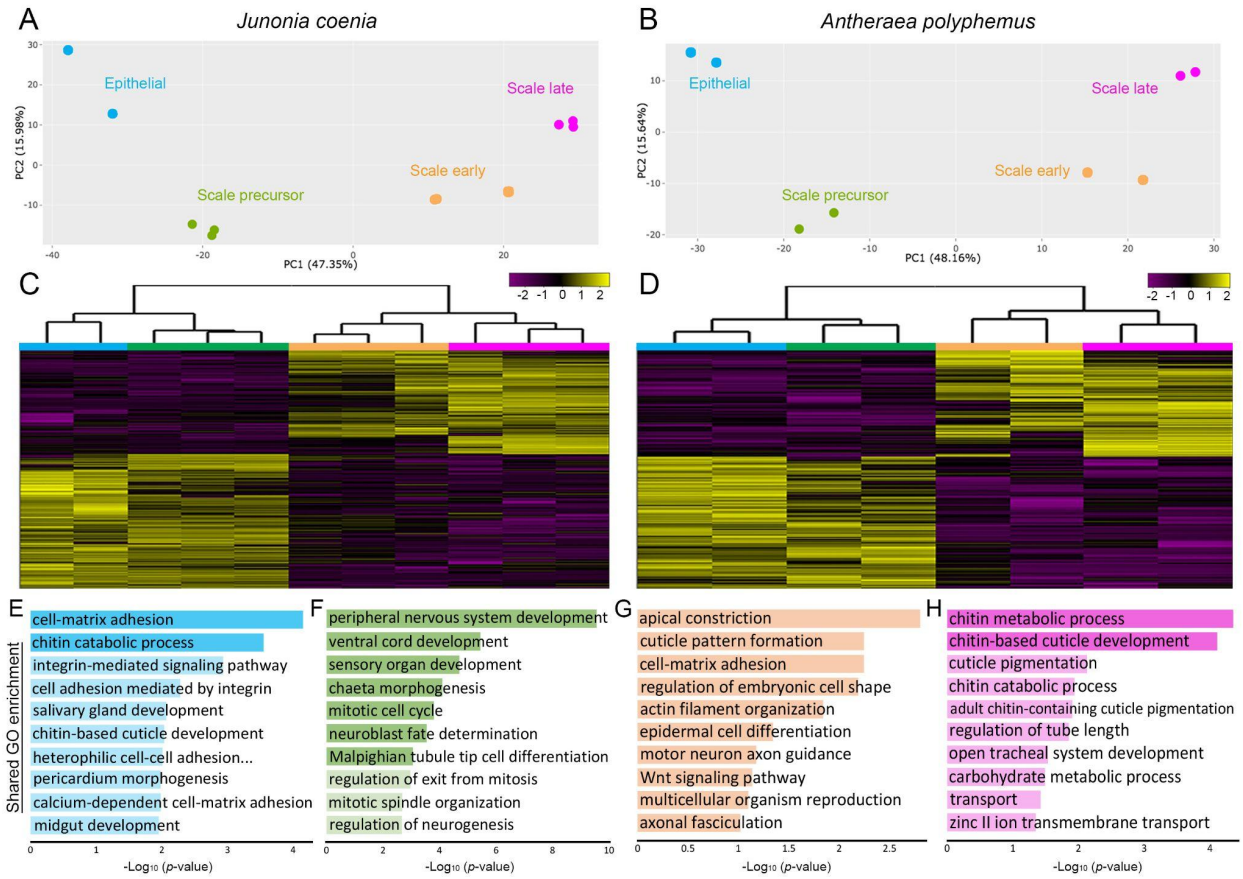


Figure 3.2. Developmental transcriptomic analysis of pupal wing tissue in *J. coenia* and *A. polyphemus*

Principal component analysis (PCA) of RNA-seq samples for *Junonia coenia* (A) and *Antheraea polyphemus* (B) clustered by developmental stage. (C) Hierarchical clustering of samples and expression heatmap of differentially expressed genes. (D-E) hierarchical clustering of samples and expression heatmap of differentially expressed genes. Shared Gene ontology (GO) enrichment of top biological processes between *J. coenia* and *A. polyphemus* from pairwise differential expression analyses corresponding to (E) epithelial (F) scale precursor (G) scale early and (H) scale late stages. The -log10(*P*-value) bar plots of GO enrichment terms are colored by their corresponding developmental stage, solid bars represent FDR < 0.05 and translucent bars represent FDR > 0.05.

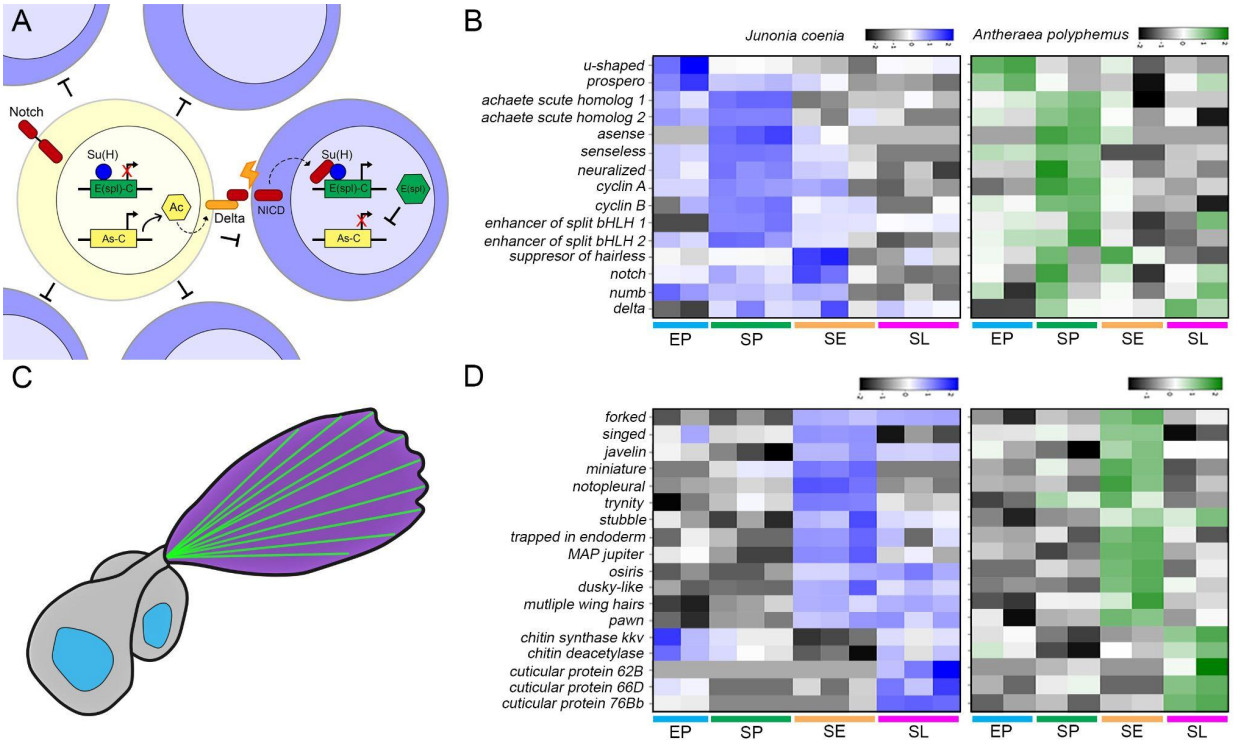


Figure 3.3. Scale development pathways and select differentially expressed genes with equivalent expression patterns between *Junonia coenia* and *Antheraea polyphemus*.

(A) Schematic of the *Notch* signaling pathway, which is required for lateral inhibition and sensory organ precursor patterning. Cells with high proneural activity (yellow; *Achaete/scute* synergize with *Senseless*) inhibit proneural activity in their neighbors (blue) via *Delta-Notch* signaling and its downstream target, the *E(spl)* repressor. (B) Heatmap displaying profiles of DEGs with equivalent expression patterns at early scale cell developmental timepoints, including proneural, cell cycle and *Notch* signaling factors. (C) Cartoon of a typical scale cell and socket cell during pupal development; nuclei colored blue, and internal actin bundles colored green. (D) Heatmap displaying profiles of DEGs with equivalent expression patterns at late scale cell developmental timepoints, including cytoskeletal organization factors, chitin synthesis and cuticle proteins.

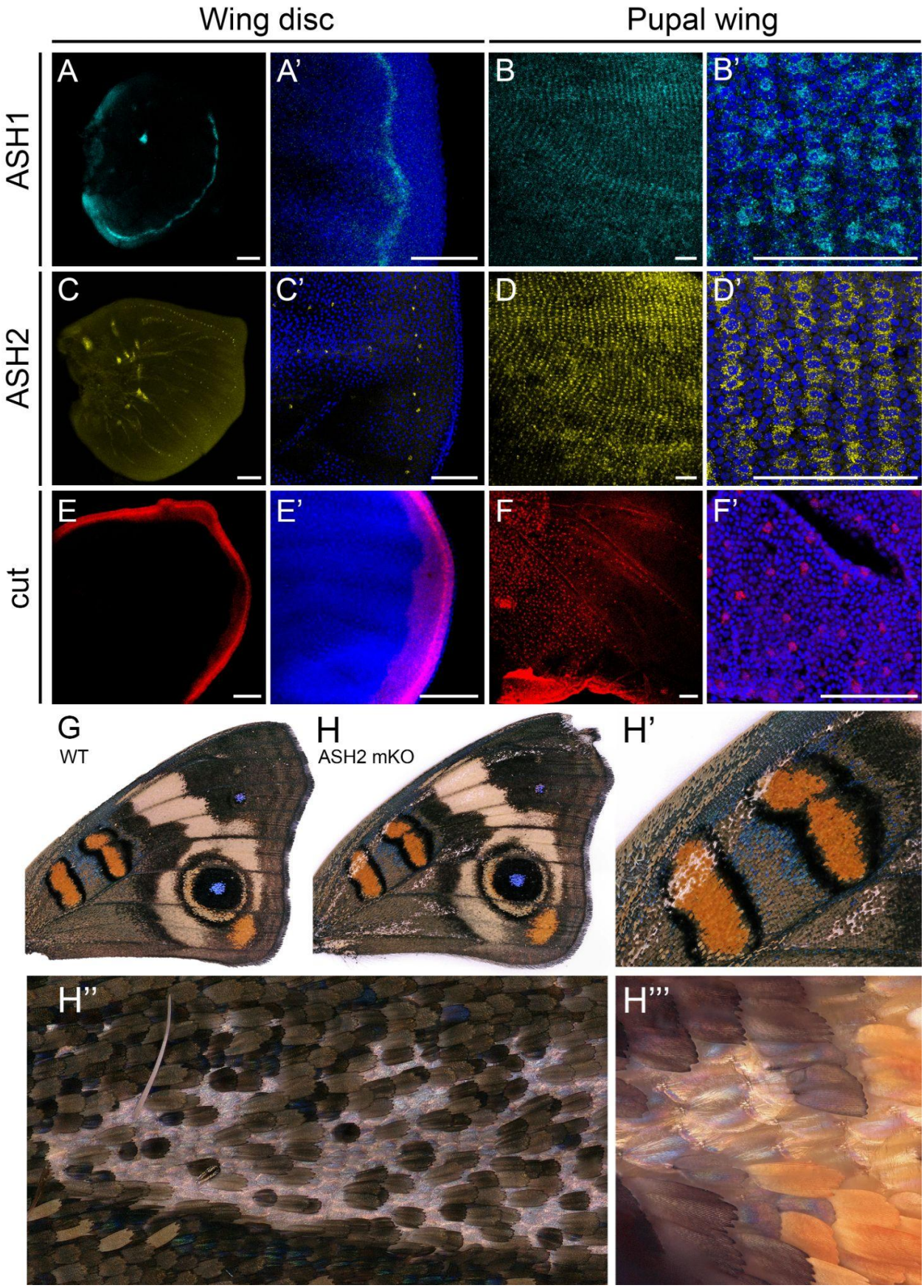

Wing disc
Pupal wing
ASH1
ASH2
cut
A
A'
B
B'
C
C'
D
D'
E
E'
F
F'
G
WT
H
ASH2 mKO
H'
H''
H'''

Figure 3.4. Expression and function of proneural genes in developing wings of *Junonia coenia*

Hybridization Chain Reaction (HCR) *in situ* hybridization for *achaete scute homolog 1* (*ASH1*) in the wing disc (A-A') and pupal wing (B-B'). Expression of *achaete scute homolog 2* (*ASH2*) in the wing disc (C-C') and pupal wing (D-D'). Position-specific expression of *ASH2* in the wing disc determines where sensory bristles develop. *ASH1* and *AHS2* mRNA show high expression in scale precursor cells. Expression of *cut* in the wing disc (E-E') and pupal wing (F-F'). *Cut* expression in 24 h pupal wings appears to correlate with subsets of scale precursor cells that become long sensory hairs on the hindwing. Scale bars = 100 μm. CRISPR/Cas9 induced mosaic knockout of *ASH2* in an adult individual results in absence of both scales and sockets (G-H'").

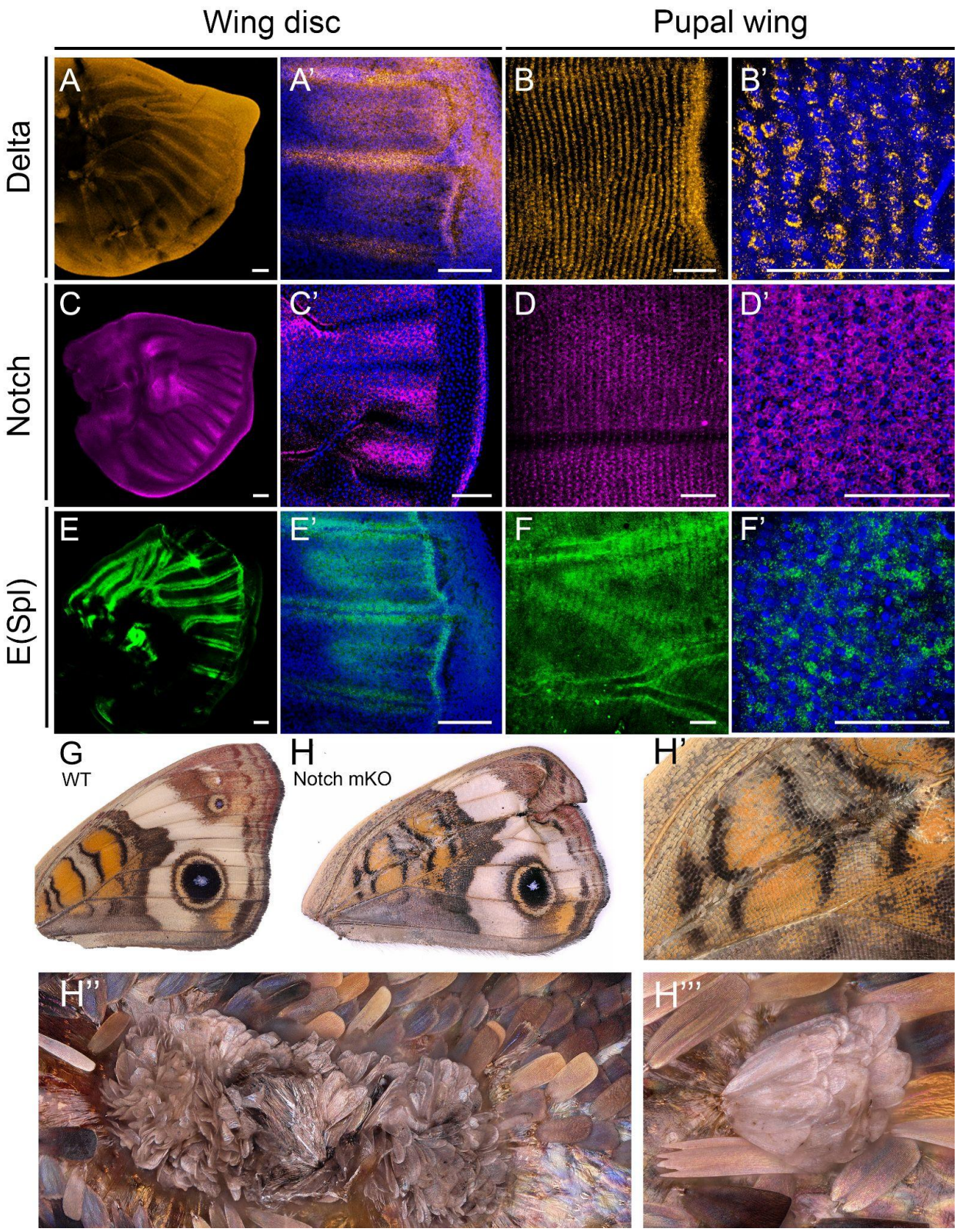

Wing disc
Pupal wing
Delta
Notch
E(Spl)
A
A'
B
B'
C
C'
D
D'
E
E'
F
F'
G
WT
H
Notch mKO
H'
H''
H'''

Figure 3.5. Expression and function of neurogenic factors in developing wings of *Junonia coenia*

Hybridization Chain Reaction (HCR) *in situ* hybridization for *Delta* in the wing disc (A-A') and pupal wing (B-B'). In the imaginal disc, *Delta* is expressed in the future veins, similar to expression in *D. melanogaster* wing discs. 24 hr APF, *Delta* is expressed in rows of scale precursor cells and absent in epithelial cells. Expression of *Notch* in the wing disc (C-C') and pupal wing (D-D'). *Notch* expression occurs at high levels in intervein regions in the imaginal disc and in pupal wings, *Notch* expression appears in a grid-like pattern and occurs at highest levels in the epithelial cells, in line with its role as a proneural repressor. Expression of *Enhancer of split* in the wing disc (E-E') and pupal wing (F-F'). *Enh(spl)* is found in flanking the veins in the imaginal disc and in pupal wings, *Enh(spl)* appears to be expressed at higher levels surrounding scale precursor cells. Scale bars = 100 μm. CRISPR/Cas9 induced mosaic knockout of *Notch* in an adult individual results in dense clusters of scales due to loss of lateral inhibition (G-H''').

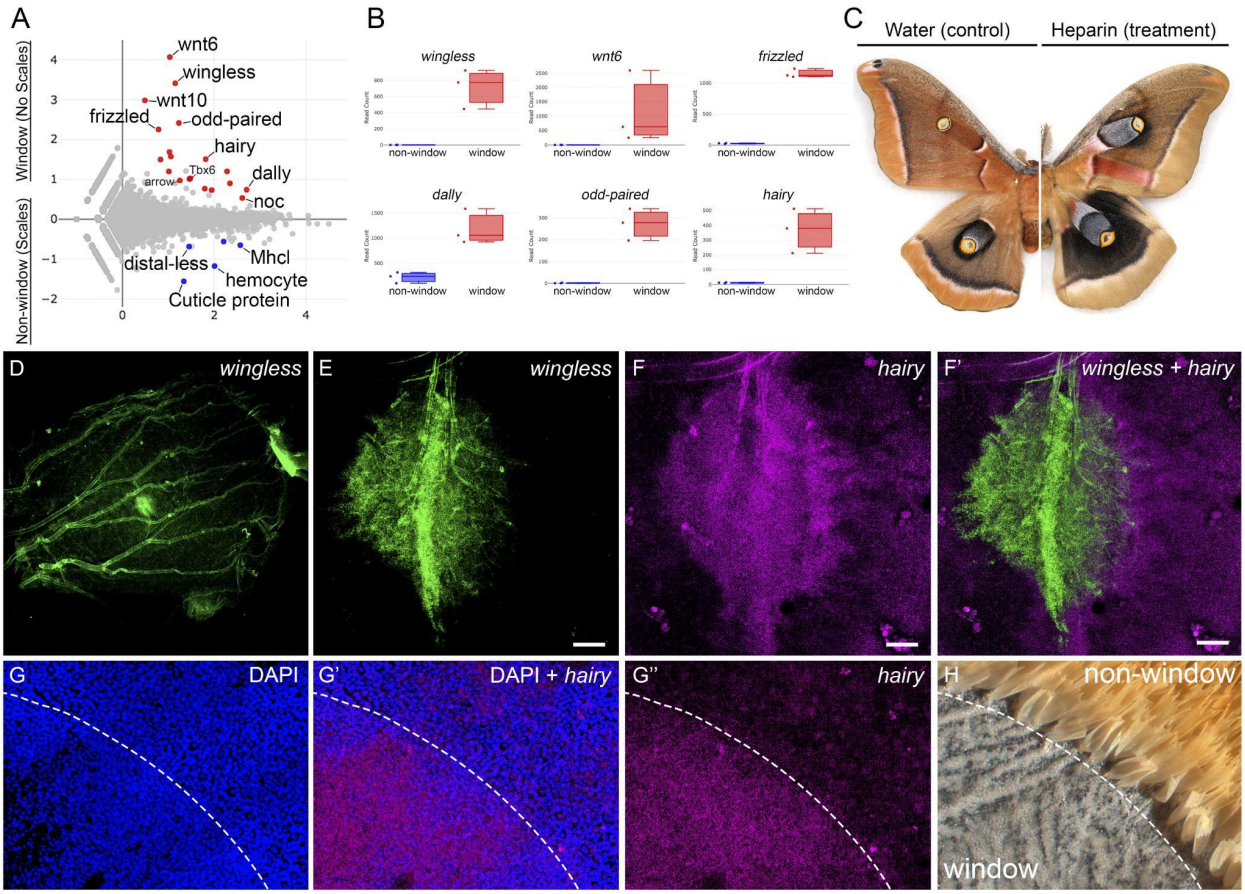


Figure 3.6. Comparative RNA-seq and expression between the 'window' and 'non-window' region in *Antheraea polyphemus* wings

Pairwise differential expression of window versus non-window regions revealed upregulation of genes in the Wnt signaling pathway, such as *wingless, Wnt6, Wnt10, frizzled, arrow, odd paired* and *divisions abnormally delayed*, as well as the bHLH transcription repressor *hairy* (A-B). Heparin treatments reveal shifts in wing pattern elements, likely due to altered Wnt expression profiles (C). In *A. polyphemus wg* is highly expressed in the 'window' region surrounding the discal crossvein (D-E). Higher *hairy* expression can be observed largely overlapping with *wingless* in the 'window' region surrounding the discal crossvein (F). Higher magnification showing the boundary where *hairy* expression drops and scaleprecursor cells form (G-G''). Notably, *hairy* expression is strongly correlated with the absence of scale precursor cell formation and anti-correlated with scale precursor formation adjacent to the 'window' region, where scales develop normally (G-H).

Supplementary Figures

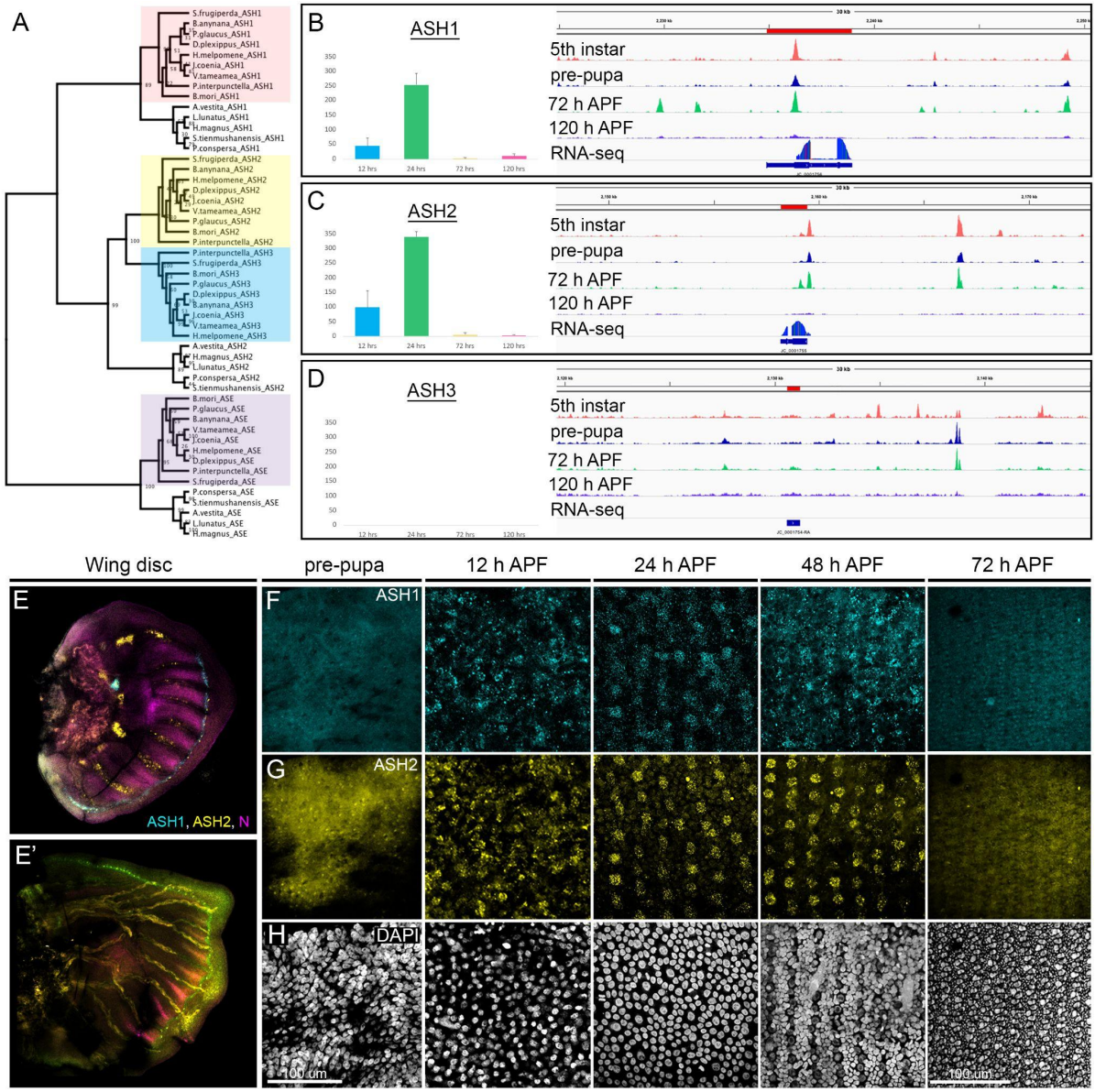


Fig S3.1. *Achaete scute homolog* gene tree and expression patterns across developmental stages.

(A) Simplified *achaete scute homolo*g (*ASH*) gene tree for species within Lepidoptera and Trichotera, highlighting three distinct *ASH* genes and one *ase* gene in Lepidoptera. (B-D) Normalized expression counts from RNA-seq and adjacent ATAC-seq plots for *ASH1*, *ASH2*, and *ASH3*. Note the lack of RNA expression and clear ATAC-seq peaks for *ASH3* in pupal wing tissue. ATACseq data was downloaded from van der Burg et al., 2019 and visualized in Integrative Genomics Viewer (IGV_2.8.9). (E-G) HCR *in situ* hybridization in the wing disc (*ASH1* in cyan, *ASH2* in yellow, *Notch* in magenta), pre-pupa, 12 h, 24 h, 48 h, and 72 h APF, highlighting dynamic expression patterns over time. *ASH1*, *ASH2* expression corresponds to mechanosensory bristles on the wing

margin in the wing disc, then appears stochastically expressed at 12 h, become organized and expressed in scale precursor cells from 24-48 hrs, and expression drops off by 72 hrs.

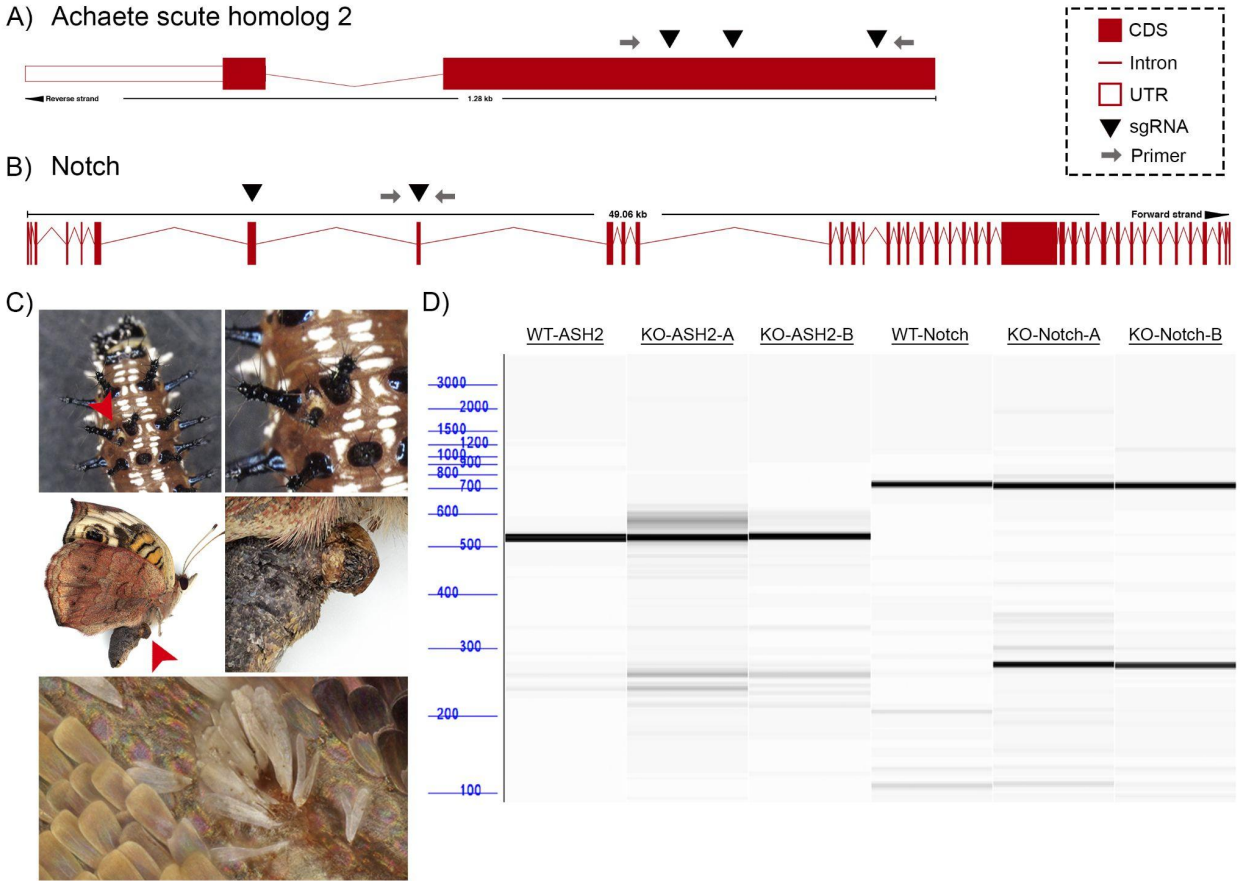


Fig. S3.2. CRISPR-Cas9 mutagenesis for ASH2 and *Notch*

(A-B) Target sites of CRISPR-Cas9 mutagenesis for *ASH2* and *Notch*. (C) Additional phenotypes observed in *Notch*-injected individuals, including a larva which appears to have ectopic setae and effects on adult abdomen and wings. (D) Genomic PCR products visualized on a fragment analyzer, indicating smaller fragments from gene deletions in sgRNA-injected individuals for *ASH2* and *Notch*.

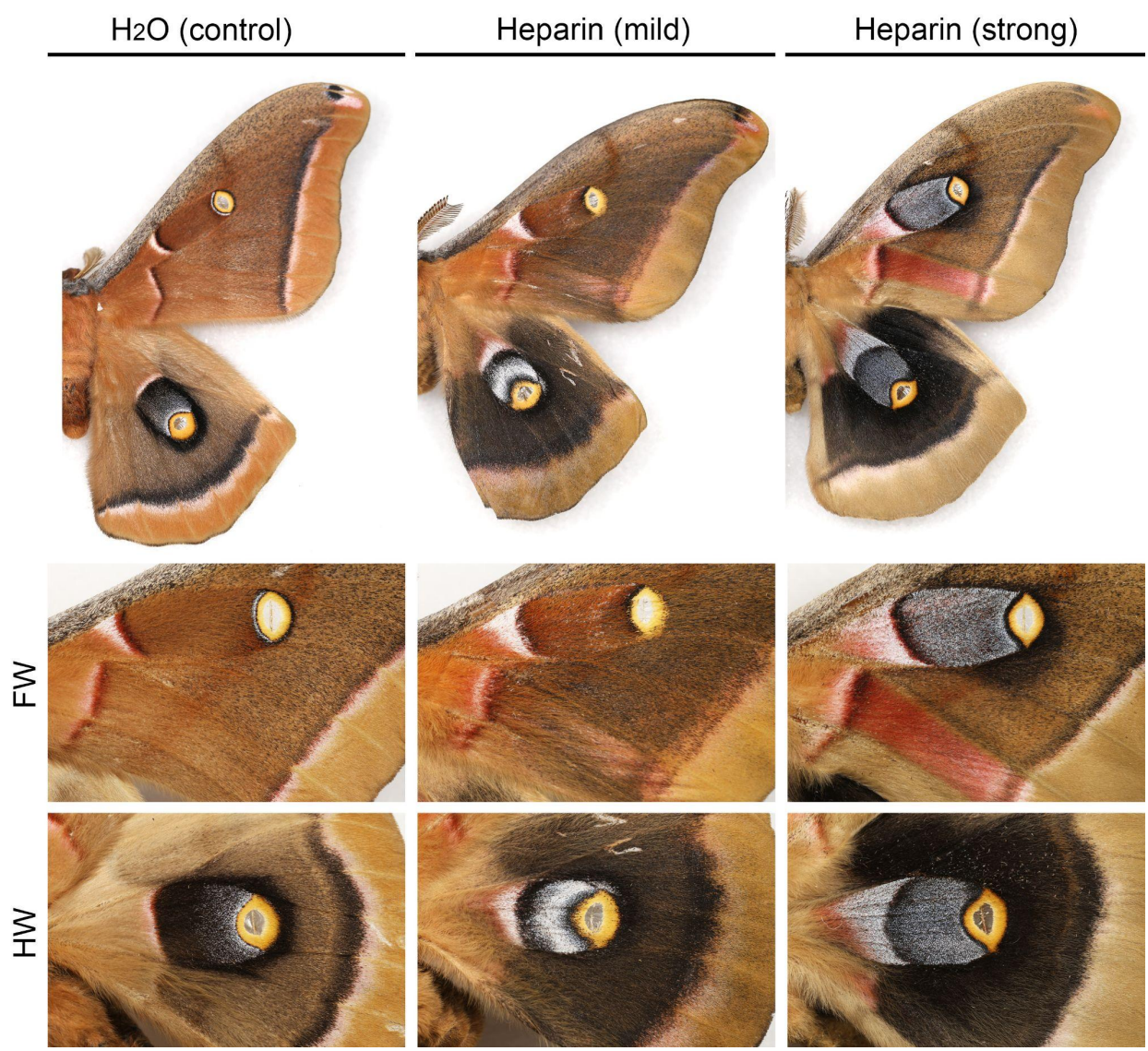


Fig S3.3. Additional phenotypes for control and heparin treatments in *Antheraea polyphemus*.

Control treatments using water produce wild-type wing patterns, whereas heparin treatments produce a range of mild to strong shifts in color patterns, likely due to misexpression of wnt signaling ligands during pupal wing development. What appear to be ‘mild’ effects of heparin treatment echo results from Sourakov & Shirai (2020); the grey and red zone of scale surrounding the window region expands and melanic scales expand on the hindwing.

Preface to Chapter 4

Back in 2014, I found myself in a unique position of working as a field biologist and marketing strategist at an ecotourism company in the heart of the Amazon rainforest. In addition to conducting biodiversity research, I worked closely with the marketing director of the company to generate global publicity of our work and conservation efforts through crafting viral social media content, coordinating expeditions with journalists, and hosting television productions.

During this time, I became intrigued by novel, portable scientific devices that provided new opportunities to 'take the lab into the field', such as origami-based paper microscopes (Foldscope Instruments, Inc) and miniaturized DNA sequencing platforms (the MinION, Oxford Nanopore Technologies, Inc). The prospect of decoding the DNA of unknown species in the middle of a rainforest – something that seemed limited to science fiction – inspired me to apply for funding from the National Geographic Society and apply for the MinION early access program.

Upon receiving funding and access to the early MinION, I was determined to expand my scientific background and skill sets further. I applied to graduate programs and was accepted to the University of California Berkeley in 2016. Subsequently, my collaborators and I put the MinION to the test under remote tropical field conditions, and successfully carried out molecular analyses of endemic fauna in one of the world's most imperiled biodiversity hotspots, the Ecuadorian Chocó (Pomerantz et al. 2018).
Since then, I have worked with numerous labs to assist them with nanopore sequencing experiments, ranging from highly multiplexed long-range amplicons, to whole genome sequencing, to full-length cDNA transcriptomics. Moreover, I am proud to have developed, raised seed funding for, and launched a new educational program for undergraduate students at UC Berkeley called 'Field Genomics', in which we teach students to how to conduct fieldwork and apply nanopore sequencing to address their scientific questions, including 16S metabarcoding and mitochondrial genome sequencing.

From what I have outlined, I hope you too will see how these experiences utilizing portable scientific equipment have been transformative for my development as a scientist and educator. While some studies to date have made use of nanopore-based amplicon sequencing, there are various options for portable genomics equipment, as well as DNA isolation, amplification, and bioinformatics strategies spread amongst the literature. As such, myself and collaborators have attempted to synthesize the current literature and report leading practices in the following protocol so that anyone, anywhere can carry out their own DNA amplicon sequencing projects using miniaturized laboratory equipment.

Chapter 4. A step-by-step protocol for DNA amplicon sequencing using miniaturized laboratory equipment for genetic biomonitoring and biodiversity exploration

Authors

Aaron Pomerantz[1,2]*, Kristoffer Sahlin[3]*, Nina Vasiljevic[4], Adeline Seah[5], Marisa Lim[5], Emily Humble[6], Susan Kennedy[7,8], Henrik Krehenwinkel[8], Sven Winter[9], Rob Ogden[6] ** and Stefan Prost[10,11,12]# **

Affiliations

[1]Department of Integrative Biology, University of California, Berkeley, CA-94720, USA
[2]Marine Biology Laboratory, Woods Hole, MA-02543, USA
[3]Department of Mathematics, Science for Life Laboratory, Stockholm University, 106 91 Stockholm, Sweden
[4]Zurich Institute of Forensic Medicine, University of Zurich, Winterthurerstrasse 190/52, 8057 Zurich, Switzerland
[5]Wildlife Conservation Society, Zoological Health Program, Bronx Zoo, 2300 Southern Blvd, Bronx, NY, 10460, USA
[6]Royal (Dick) School of Veterinary Studies and the Roslin Institute, University of Edinburgh, Easter Bush Campus, EH25 9RG, United Kingdom
[7]Biodiversity and Biocomplexity Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa, Japan
[8]Department of Biogeography, University of Trier, 54296 Trier, Germany
[9]Senckenberg Biodiversity and Climate Research Centre, 60325 Frankfurt am Main, Germany
[10]LOEWE-Centre for Translational Biodiversity Genomics, Senckenberg Museum, 60325 Frankfurt, Germany
[11]Natural History Museum, Vienna, Central Research Laboratories, Burgring 7, 1010 Vienna, Austria
[12]South African National Biodiversity Institute, National Zoological Garden, Pretoria 0002, South Africa
* Contributed equally
** Contributed equally
#Correspondence: stefanprost.research@protonmail.com

Abstract

Human-mediated environmental change is depleting biodiversity faster than it can be characterized, while invasive species cause agricultural damage, threaten human health, and disrupt native habitats. Consequently, the application of effective approaches for rapid surveillance and identification of biological samples is increasingly important to inform conservation efforts. Taxonomic assignments have been greatly advanced using sequence-based applications, such as DNA barcoding, a diagnostic

technique that utilizes polymerase chain reaction (PCR) and DNA sequence analysis of standardized genetic regions. However, in many biodiversity hotspots, endeavors are often hindered by a lack of genomic infrastructure and funding for biodiversity research and restrictions on the transport of biological samples. A promising development is the advent of low-cost, miniaturized scientific equipment. Such tools can be assembled into functional laboratories to carry out genetic analyses *in situ*, at local institutions, field stations, or classrooms. Here, we outline all the steps required to perform amplicon sequencing applications outside of a conventional laboratory environment using miniaturized scientific equipment.

Introduction

Biodiversity loss has increased rapidly during the past decades. Natural habitats are diminishing at an unprecedented rate and extinctions are predicted for many taxa across the tree of life[1,2]. Furthermore, the spread of invasive alien species into new environments can cause agricultural damage, jeopardize human health, and negatively impact native biodiversity[3,4]. Conservation and inspection-focused researchers and agencies are faced with the immense task of characterizing baseline biodiversity data, documenting how communities change, and detecting harmful invasive pests and pathogens. To do so, it is imperative that the tools and protocols are standardized and are time- and cost-efficient.

The advent of high-throughput sequencing (HTS) technologies, coupled with the development of standardized DNA marker systems, known as DNA barcodes[5], have greatly facilitated large-scale monitoring and community-level assessment of species diversity. Ever-growing reference databases (such as the Barcode of Life Data System (BOLD)[6] and the NCBI GenBank database[7]) and the ability to simultaneously analyze high numbers of samples[8] have further increased the utility of DNA sequencing as a valuable tool for biodiversity and biomonitoring exploration.

However, many biodiversity hotspots and inspection sites are located in regions that lack readily available access to genomic resources, such as HTS platforms. A traditional alternative, to transport samples abroad for analysis, has become increasingly restricted due to international conventions such as CITES (Convention on International Trade in Endangered Species of Wild Fauna and Flora) or the Convention on Biological Diversity's Nagoya Protocol on Access and Benefit Sharing. While these are important measures to control the shipment of natural and potentially protected wildlife, they can increase the time it takes to generate, analyze, and report on the biological information. The ability to sequence biological samples within the country of origin with portable, inexpensive laboratory equipment can bring significant benefits for biodiversity monitoring and explorations, while simultaneously creating opportunities for developing local scientific capacity (a key aspect of the Nagoya Protocol). Additionally, rapid characterization of invasive pests and pathogens at or near the site of detection via molecular analyses can be a useful means of mitigating their negative impacts on health, ecosystems, and economies[9]. Thus, these technologies offer a chance to empower local scientists and conservation agencies that currently rely on international research facilities. Miniaturized equipment, such as portable thermocyclers (miniPCR bio, MiniOne Systems) and nanopore-based nucleic acid sequencing devices (Oxford

Nanopore Technologies), have gained in popularity, as they are relatively inexpensive compared to traditional, bulky molecular laboratory equipment and allow for *in situ* processing of genetic material[10]. Portable genomics laboratories have been deployed around the world in a diverse set of ecosystems and settings, including the rainforests of Tanzania[11], Ecuador[12] and Madagascar[13], out at sea[14] and the Antarctic[15]; as well as being used to monitor disease outbreaks such as Ebola in West Africa[16] or Zika in Brazil[17], and for educational programs[18,19].

While several studies have utilized nanopore-based amplicon sequencing, there are various options for portable genomics equipment, as well as DNA isolation, amplification, and bioinformatics strategies spread amongst the literature. As such, we have attempted to synthesize the current literature and report leading practices in this protocol so that anyone, anywhere can carry out their own DNA amplicon sequencing projects using miniaturized laboratory equipment.

Development of the protocol

Central to advancements in miniaturized genomics equipment is the small MinION sequencing device (Oxford Nanopore Technologies (ONT)), which uses 'nanopore sequencing', a process by which changes in the ionic current measured when a single-stranded DNA fragment is funneled through a biological pore in the device's membrane, can be converted into a nucleotide sequence[20]. ONT's MinION is a USB-powered, portable sequencing platform that was launched in 2014. Since that time, there have been substantial improvements made in sequencing yield and quality[20,21]. ONT also launched an in-expensive low throughput flow cell called 'Flongle', which can further reduce fixed costs of DNA barcoding projects on the MinION platform. Although raw sequence accuracy remains relatively low for the MinION compared to other HTS technologies[21,22], highly accurate consensus sequences can be generated for DNA amplicons[12,23,24]. Furthermore, the development of a new pore with two reader-heads (10.x) has drastically decreased the error rate and increased the consensus barcode accuracy[14]. Following several years of extensive experimentation and refinement of MinION methods by the scientific community, combined with the optimization of other commercially available field-deployable lab equipment, there is now an opportunity to consolidate a step-by-step protocol for DNA amplicon sequencing using miniaturized laboratory equipment. We assemble currently available best practices for performing amplicon sequencing experiments outside of conventional laboratory environments based on research groups who have processed samples and applied portable genomics tools under field conditions. Additionally, we provide cost-effective strategies for multiplexing high numbers of samples for each sequencing run through user-customized indexing of amplicons, and provide instructions for the loading of both the standard flow cell and the in-expensive lower throughput Flongle flow cell. Finally, as long-read platforms such as ONT result in higher raw-read error relative to other sequencing platforms and require various processing steps, we present a simplified downstream bioinformatics workflow for the demultiplexing, polishing and *de novo* assembly of raw data into accurate consensus amplicon sequences.

Advantages and limitations

We anticipate that this approach can be adapted for various projects that aim to perform real-time DNA amplicon sequencing, such as within-country biodiversity assessment efforts, on-site invasive species detection, and educational genomics programs, as users can tailor the protocol to assess any taxa of interest in a relatively time- and cost-efficient manner. As for limitations to this protocol, we note that advancements to nanopore technology, chemistry, and software are ever-changing and improving at a rapid pace, and as such, a user interested in these methods should also consult the most up-to-date literature and manufacturer protocols. We also note that with regard to miniaturized field-deployed tools, there can be a trade-off between quickly characterizing a small to medium number of biological samples and conducting more thorough biodiversity assessments that may involve several hundreds or thousands of samples, which could impact productivity and operation costs. As such, the user should take into account the goal of their project, the number of samples to be processed (e.g. processing capacity using miniaturized 16-32 well PCR machines compared to full-sized benchtop 96 well PCR machines), and the available storage options for samples and reagents required to carry out DNA amplicon sequencing experiments outside of a conventional laboratory environment.

Description of the methods

Here we present a protocol for designing and executing rapid, multiplexed amplicon sequencing using miniaturized laboratory equipment, including ONT's MinION sequencing platform (Figure 1). The approach can be used in typical molecular laboratories or non-conventional laboratory spaces, such as remote field stations, or classroom settings. We mark the use of relatively inexpensive mobile equipment and steps that reduce complexity using the FIELD flag throughout the protocol to enable users to carry out the molecular and bioinformatic processing in areas with limited infrastructure. Schematics of the laboratory processing and the bioinformatic analyses can be found in Figure 2 and Figure 3, respectively.

Transportation to the Field

Users should consider trade-offs between the time it may take to carry out experiments under field conditions versus carrying out experiments at nearby facilities that may have access to more stable storage and freezer options[12,15]. If equipment needs to be transported into the field, either to conduct on-site analyses or to perform analyses in a field-station with limited infrastructure, portable equipment can be transported in luggage to the site of interest, and for additional safeguarding, can be loaded into protective cases, such as a Pelican case (Pelican, Torrance, USA)[12,16]. Cold chain reagents and flow cells can be packed into polystyrene boxes with ice or cool packs and sealed for transportation[12,16]. Users can consider additional options for maintaining reagents at 2°C-8°C for several days with portable refrigeration systems (e.g. the Crēdo Cube (Pelican BioThermal)). Under cold environmental conditions, such as the Antarctic, the temperature of the MinION and other electronic equipment can be regulated using hand warmers and insulating materials[15.] Long-term storage of molecular biology reagents that require stable cold temperatures can still present challenges for nanopore sequencing projects under non-ideal environmental conditions.

Several field-friendly lyophilized reagents stable for long-term storage at ambient temperature have been reported, including lyophilized PCR reagents[11]. ONT offers a field sequencing kit (SQK-LRK001) for cold-chain free library preparation, but this option is not recommended for the presented DNA barcoding protocol, as the kit relies on transposase activity which cleaves template molecules.

Sample Collection

Sample collection permits should be authorized by the appropriate agencies and samples should be treated ethically in accordance with community guidelines. Whenever possible, samples that are collected and preserved should be properly vouchered, so that they can serve as verifiable and permanent records.

DNA extraction

The first step of the experiment requires the extraction of DNA from the samples of interest. There are numerous methods and commercial options to carry out DNA extraction, several of which have been applied under field conditions, such as QuickExtract™ (Lucigen)[14,24], spin column-based nucleotide isolation kits[12,13,15,23] and the HotShot extraction[25]. We encourage the user to seek out an extraction protocol that is most suitable for their sample type and price range. We have tested a variety of commercially available extraction kits or custom solutions and found them suitable for extraction of DNA from various sample types (such as tissue, fecal and soil samples) under field conditions, which produced sufficient yields and qualities for subsequent amplification of target regions (see reagents list and references[12,23,26]). These include standard kits such as the Qiagen QIAmp DNeasy Blood & Tissue kit and the Bio-Rad Chelex 100 resin extraction. The latter is more cost-efficient (approximately $0.17 USD per sample[26]) and can be carried out using only a portable thermocycler. Alternatively, to preserve intact specimens for morphological examination, QuickExtract™ (Lucigen) can be used for DNA extractions, which does not require centrifugation[10,19]. Buffer-based extraction methods such as the HotSHOT protocol[27] are also fast and easy to carry out using a thermocycler. Many of the aforementioned extraction kits and methods have been shown to work with non-invasively sampled DNA sources (e.g from scat, hair, feathers), which may contain PCR inhibitors or be degraded from environmental exposure[26].

Amplification, indexing and pooling

After isolating DNA from the samples, genetic regions of interest (DNA barcodes) are amplified via polymerase chain reaction (PCR), which can be efficiently performed using inexpensive mobile thermocyclers. For instance, researchers have reported success using the miniPCR (https://www.minipcr.com), MiniOne (https://theminione.com) and BentoLab (https://www.bento.bio) for thermocycling steps under field conditions (see e.g. [12,14,28,29]). These devices can also serve as miniature heating blocks, be programmed with a mobile phone or laptop, and run off external battery power in remote environments (see Figure 1). While standard amplicons for DNA barcoding are

generally ~300-900 bp in length, sequencing on the MinION also allows the user to generate long-range amplicons (thousands of bases long), which can increase phylogenetic resolution compared with shorter standard barcode sequences (see e.g. [23,30]). In order to pool amplicons from numerous samples on a single sequencing flow cell, indexes can be added to the amplicons. This can either be performed using a 1-Step or 2-Step PCR protocol (Figure 2). The 1-Step protocol may be preferable if a single primer set is used for the study, as it is faster than the 2-Step PCR protocol, requires fewer reagents, and reduces the chance of chimeric-amplicon formation (see e.g. [23,24]). In this case, unique indexes are added to the primer sequence directly. Alternatively, for studies that aim to use multiple primer sets (e.g. for multi-locus PCR amplification or the application of primer cocktails, which is common for universal COI amplification[31]) the 2-Step PCR protocol may be the preferred option. In the 2-Step PCR approach, universal tail sequences are added to the amplification primers. These universal tails then allow for the addition of index sequences to the ends of the DNA fragment in a second PCR amplification step. While the 2-Step PCR protocol increases the chance of chimeric-amplicon formation, these are only found in low abundance and do not present any issues for DNA barcoding experiments (on the contrary to metabarcoding applications[32]). Furthermore, they can be filtered out during the bioinformatic processing (see below). The 2-Step PCR protocol allows the user to share indexes between projects and research groups, thereby further reducing the cost of barcoding. Using different indexes on either end of the amplicon enables cost-effective sequencing for large-scale sampling in both indexing setups (the 1-Step and the 2-Step protocol)[24]. In the presented protocol we provide the option to incorporate ONT's standard universal tails in the primer sequences, which makes the primers compatible with inexpensive custom indexes and/or ONT's DNA barcoding kit (see reagents list). In case researchers plan to also sequence the amplicons on an Illumina HTS platform, Illumina's universal tails (see [33]) can be used for both Illumina and ONT sequencing.

Our group, along with other field-focused researchers, have transported a variety of polymerases into the field for DNA barcoding, and here we report the use of either DreamTaq™ Hot Start DNA Polymerase Master Mix (Thermo Fisher Scientific, Waltham, MA, USA) or Q5 Hot Start High-Fidelity 2X Master Mix (NEB, Ipswich, MA, USA) for carrying out PCR amplifications. We note that these polymerases are rather expensive and that a variety of commercial polymerases and PCR master mixes are available, and recommend the user should seek out the one most appropriate for their experiment and price range. Another interesting development that requires further testing is the use of lyophilized polymerases, which would reduce cold chain requirements. Assessment of DNA, either after amplification or during the sequencing library preparation, can be carried out using miniaturized gel-electrophoresis systems, such as the blueGel (miniPCR), MiniOne system (MinOne), or BentoLab (Bento Bioworks Ltd., London, United Kingdom), or quantification instruments such as the Qubit (Thermo Fisher Scientific) or the TapeStation (Agilent) in case these are available. To reduce time and resources running gels, users can check just a small subset of reactions on the gel electrophoresis to ensure that there is no widespread amplification failure or presence of amplicons in negative controls.

Sequencing library preparation

The presented sequencing library preparation protocol is based on ONT's library preparation manual for the Ligation Sequencing Kit (SQK-LSK109), experiences gained from working in the field (see [11,12,18,23,29]) and MinION-based DNA barcoding in the laboratory[26]. The library preparation can be carried out in approximately two hours in remote conditions, and requires pipettes, a small centrifuge, a miniature thermocycler and a small magnetic rack for bead clean-ups (Figure 1). We have included optional steps that require additional equipment in case the protocol is performed in a fully equipped laboratory.

Sequencing

Running the MinION Mk1B sequencer can be performed using external battery power on either a laptop (with sufficient space and speed; see https://nanoporetech.com/community/lab-it-requirements) or on standalone devices such as ONT's Mk1C. Alternatively, DIY setups using mini-servers such as the Nvidia Jetson can be used as an inexpensive means to running the MinION and the basecalling (see e.g. https://github.com/sirselim/jetson_nanopore_sequencing). Sequencing using the Mk1B sequencer requires the use of the MinKNOW software, which does not require internet access if the proper offline version is installed. The MinION sequencing run generates data in real-time and depending on the number of samples and read output required, can be run for a few hours (for smaller runs) or in case of high-throughput multiplexing (e.g. hundreds to thousands of samples), the sequencer can be run for up to 72h. Additionally, nuclease-based washes can be applied to used flow cells for their re-use (Flow Cell Wash Kit: EXP-WSH004). In our experience with current versions, the MinION produces about 10,000 sequenced reads per minute for ~500bp long DNA barcodes. We recommend aiming for ~1,000x coverage per individual amplicon to obtain around 300-500x coverage after filtering, as recommended in [10,34]. This should provide the user with a good representation for each DNA barcode even in the presence of slightly uneven pooling of the different amplicons.

Bioinformatics

Recent developments in ONT's basecalling algorithms (the conversion of the MinION's ionic current profiles to fastq sequences) allow for live basecalling during sequencing using MinKNOW. We recommend using live basecalling for smaller scale sequencing runs of 1-3h. However, basecalling can also be performed at a later stage on a laptop or server if needed, especially for long sequencing runs producing several gigabases of data, or on the sequencing device directly when using ONT's Mk1C device. ONT offers several options for basecalling. In general, High-Accuracy models produce highly accurate basecalls at the expense of speed, while Fast models produce basecalls with lower accuracy, but are significantly faster than High-Accuracy models. The Guppy basecaller further offers the option to use graphics card chips (GPUs) instead of CPUs for basecalling, which increases the basecalling speed even further. Today, (small) computers with many GPUs can be purchased for relatively low prices. In the second step we show the use of two different demultiplexing tools, which work for ONT and

custom indexing (software tool: minibar[23]; https://github.com/calacademy-research/minibar), or in the case of Guppy (Oxford Nanopore Technologies, UK) only for ONT indexing kits (e.g. the PCR Barcoding Expansion 1-96 kit) and custom single indexing. These tools were specifically developed for error-prone nanopore sequencing data. In addition to demultiplexing of dual or single index libraries, minibar also offers the option to demultiplex samples based on primer sequences, in case different DNA barcodes were amplified in single reactions (multiplex PCR). There are other options available see e.g. [35]. We recommend demultiplexing the read data before quality assessment and filtering. This allows the user to pool barcodes of varying sizes in the same multiplex PCR amplification, and carry out filtering for each amplicon separately afterwards. NanoPlot (https://github.com/wdecoster/NanoPlot) outputs summary statistics and plots showing the quality of the sequenced reads. For downstream processing of raw reads, we use NGSpeciesID, a tool developed for the generation of highly accurate consensus sequences from third-generation long-read sequencing technologies[36]. We note that there are a variety of bioinformatic workflows that have been reported for generating accurate consensus amplicon sequences using the MinION[12,24,26,29,30,35]. Pipelines such as ONTrack[29] or Consension[30] have been shown to work more reliably than software used in the early stages of MinION-based DNA barcoding (such the workflow presented in [12]). NGSpeciesID has been benchmarked against current pipelines, and all have been found to be comparable for generating highly accurate consensus sequences. NGSpeciesID offers many options (including read filtering, subsampling, primer site removal) and is easy to run, install and scale up. Furthermore, it has recently been validated for its use in non-human forensic applications via thorough validation[34]. We thus decided to use it in this protocol. It first removes low quality reads and PCR chimeras from the data and subsequently, automatically carries out read clustering, consensus calling, polishing and optional primer removal (see description below). The final polished consensus sequences can subsequently be compared against different databases, such as the Barcode of Life Data System (BOLD)[6], the rRNA database project SILVA[37], or the National Center for Biotechnology Information (NCBI) GenBank nucleotide database[7]. BOLD offers a large compilation of Cytochrome c oxidase subunit 1 (COI) from many taxa, while GenBank includes a variety of different marker systems. We also outline different statistics used to interpret database hits with BLAST[7] and provide recommendations and guidelines on how to interpret the results.

NGSpeciesID

To simplify the bioinformatic processing, we provide commands for NGSpeciesID, a tool designed to generate highly reliable consensus sequences for amplicon sequencing reads (such as DNA barcodes) generated using long-read technologies[36]. This tool has been benchmarked against other published pipelines, such as ones that rely on multiple sequence alignment or operational taxonomic unit (OTU) clustering, and all have been shown to generate high accuracy consensus sequences[36]. NGSpeciesID allows the user to first filter the raw demultiplexed read data for (a) read quality using Phred scores, (b) amplicon lengths to filter out chimeras, and (c) carries out subsampling to obtain the preferred number of reads after the filtering. It then clusters the reads based

on expected sequence similarity using read-specific error rates with the isONclust algorithm[38]. A draft consensus is formed using spoa[39] for each cluster containing more reads than a user-defined abundance threshold (default: 10% of total sample read depth). Due to consensus calling based on multiple reads, the consensus sequences formed by spoa typically have significantly lower error rates than the individual reads. NGSpeciesID then detects and merges any consensus sequences classified as reverse complements to each other using pairwise alignment with parasail[40]. Two consensus sequences are merged if they have a sequence identity above a user-defined parameter given to NGSpeciesID (default 10%). Sequence identity is calculated as 1 - mismatches/alignment-length, where mismatches can be either indels or substitutions. If one or more of the consensus sequences are classified as reverse complements to each other, then the reads from all these clusters are merged into a single file. Finally, all draft consensus sequences passing this step are polished using the original reads and medaka (https://github.com/nanoporetech/medaka) or Racon[39]. These polished sequences are the final output of NGSpeciesID. Optionally, users can specify to automatically trim priming sequences from the consensus, in which case NGSpeciesID will also carry out an additional round of reverse complement detection and polishing.

Materials

REAGENTS
DNA extraction

- QIAmp DNeasy Blood and Tissue kit (Qiagen, cat. no. 69504)

  or:

  - HotSHOT method (after [27])
  - Chelex® 100 Resin (Bio-Rad, cat. no. 142-1253)
  - QuickExtract™ (Lucigen, cat. no. QE09050)
  - Quick-DNA Plant/Seed Miniprep Kit (Zymo Research, cat. no. D6020)

FIELD Qiagen's QIAmp DNeasy Blood and Tissue kit is a standard extraction kit, which works well for most tissue types. However, it is not the cheapest option. Inexpensive kits such as the Chelex® 100 Resin have been shown to work well for many tissue types[26]. Alternatively, buffer-based extractions, such as the HotSHOT method[27], are easy and quick to carry out and represent very cheap alternatives (see [25]). The user should take their sample type into consideration (e.g. plant, insect, scat, feather) to select the most appropriate DNA isolation method.

Amplicon generation

- DreamTaq™ Hot Start DNA Polymerase Master Mix (Thermo Fisher Scientific, cat. No K9011)

  or:

  - Q5 Hot Start High-Fidelity 2X Master Mix (NEB, cat. no. M0494S)
  - Assemble your own PCR mix (e.g. Taq DNA polymerase, buffer, dNTPs)

CRITICAL The primer annealing temperatures in our protocol are optimized for the two polymerase mixes listed above. While others may work, thermocycling conditions may need to be optimized.

FIELD Here, we present different options for hot start high-fidelity polymerases but note that the user should decide on the polymerase that is most cost-effective and appropriate for their experiment. Although hot-start master mixes tend to be more expensive than other commercial polymerases, we recommend these options here because (1) they are pre-mixed formulations, allowing PCR to be quickly carried out after the addition of DNA and primers, (2) they can be used to generate either short or long-range amplicons, and (3) amplification only starts after a short heating phase, which can increase shelf life as the enzymes will not be inadvertently activated in cases of unreliable freezing conditions. We recommend testing other polymerases before applying them in the field.

- PCR primers (that include ONT's universal tail sequences)
- Custom made indexing oligos

or:

- Optional: PCR Barcoding Expansion 1-96 (Oxford Nanopore Technologies, cat. No. EXP-PBC096). To avoid confusion with DNA barcoding, this kit will be referred to as the ONT indexing kit from here on.

FIELD: Custom barcoding using PCR primers that include ONT's universal tail sequences coupled with custom made oligos is a cost-effective alternative to purchasing ONT's PCR Barcoding Expansion kits.

- Agencourt AMPure XP (e.g. Beckman Coulter, cat. no. A63881)

or:

- Alternatively, home-made magnetic beads can be used to reduce costs (https://s3-us-west-2.amazonaws.com/oww-files-public/f/f8/SPRI_buffers_v2_2.pdf)

- Molecular Biology Grade Nuclease-free Water
- Ethanol, absolute (e.g Thermo Fisher Scientific, cat. no. BP28184
- Optional: Qubit dsDNA BR Assay Kit (Thermo Scientific, cat. no. Q32850))

Gel electrophoresis

- Gel Loading Dye, Purple (6×) (NEB, cat. no. B7024)
- 1 kb Plus DNA Ladder (NEB, cat. no. N3200S)
- Agarose or FIELD GelGreen® Agarose Tabs™ (MiniPCR™, cat no. RG-1500-10)
- 10× TBE Buffer
- SYBR Safe DNA Gel Stain (Thermo Fisher Scientific, cat. no. S33102)

ONT library preparation

- ONT Sequencing Kit (Oxford Nanopore Technologies, cat. no. SQK-LSK109 or SQK-LSK110)
- NEBNext® Companion Module for Oxford Nanopore Technologies® Ligation Sequencing (cat. no. E7180S), which contains all NEB reagents needed for use with the Ligation Sequencing Kit
- Optional: Alternatively, users can purchase individual NEBNext® products:
- NEB Next Ultra II End-repair/dA-tailing Module (NEB, cat. no. E7546)

- NEBNext Quick Ligation Module (NEB, cat. no. E6056)
- Freshly prepared 70% ethanol in nuclease-free water
- Nuclease-free water (e.g. ThermoFisher, cat # AM9937)
- Optional: D1000 ScreenTape and Reagents (Agilent, cat. no. 5067-5582, 5067-5583)
- Optional: Flow Cell Wash Kit (Oxford Nanopore Technologies, cat. no. EXP-WSH004)

Equipment

Standard equipment and consumables

- P1000, P200, P10 pipette and filtered pipette tips
- 1.5 ml Eppendorf DNA LoBind tubes (e.g. Eppendorf, cat. no. 022431021)
- 0.2 ml strip thin-walled PCR tubes (e.g. Thermo Fisher Scientific, cat. no. AB0451)
- Standard thermocycler

or:

- FIELD 16 or 32-well mobile thermocycler (e.g. miniPCR™ Thermal Cyclers, cat. no. QP-1016-01; or MiniOne PCR System, cat. no. M4000, or BentoLab)
- Benchtop microcentrifuge (e.g. Thermo Fisher Scientific mySPIN 6, cat. no. 75004061).

FIELD We have used microcentrifuge devices for this protocol which spin at 6,000 RPM, which we have found to be sufficient for spin column-based DNA isolation methods.

- Optional: 3D-print a hand-powered centrifuge device which can be used for spin column-based DNA isolations[41].

- FIELD external battery packs to run miniature PCR machines and microcentrifuge. We have used the RAVPower (model: RP-PB055) or Poweradd (model: Pilot Pro).
- Magnetic rack (e.g. Thermo Fisher Scientific DynaMag-2, cat. no. 12321D)
- Cell phone with camera
- Ice or frozen cool packs
- Optional: Qubit™ 4 Fluorometer (Thermo Fisher Scientific, Waltham, MA, USA)

CRITICAL The protocol can be run without measuring exact DNA concentrations, but pooling amplicons in equimolar ratios will help to generate even sequencing coverage and recovery rate for highly multiplexed libraries.

- Optional: TapeStation 2200 (Agilent)
- Optional: Hula mixer (gentle rotator mixer); FIELD alternatively the mixing can be performed by hand
- Optional: Scale. Not needed when agarose tabs are used or if agarose is individually packed in the correct quantities beforehand.
- Optional: Portable vortex shaker (see e.g. https://gistgear.com/industrial/lab-equipment/lab-vortex-shakers)

Gel electrophoresis

- Electrophoresis chamber such as blueGel™ electrophoresis (MiniPCR™, cat no. RG-1500-01)

or:

- MiniOne System (MiniOne®, cat. no. M1000)
- BentoLab (Bento Bioworks Ltd., London, United Kingdom)

MinION sequencing

- MinION (Oxford Nanopore Technologies, cat. no. MinION Mk1B)
- Laptop that meets ONT host computer specifications (including solid-state disk drive and sufficient memory and storage) for running MinION sequencing (https://nanoporetech.com/community/lab-it-requirements).

or:

- MinION Mk1C is a standalone device for running MinION sequencing.
  - Custom built mini-server (such as the Nvidia Jetson (https://developer.nvidia.com/buy-jetson?product=all&location=US))
- MinION Flow Cell Options:
  - Standard MinION flow cell R9.4 (Oxford Nanopore Technologies, cat. no. FLO-MIN106D)

    or:
    - R10.x (Oxford Nanopore Technologies, cat. no. FLO-MIN111)

  or:
  - Flongle Adapter plus flow cells (Oxford Nanopore Technologies, cat. no. FLGIntSP)

  FIELD We recommend the use of the Flongle for smaller scale DNA barcoding projects, as it is a more cost-effective (but lower throughput) flow cell compared to the classic MinION flow cell.

Software

Users going into remote field conditions should consider downloading relevant software and reference databases ahead of time.

Index design

- Barcode_Generator (https://github.com/lcomai/barcode_generator)

MinION sequencing

- MinKNOW (Oxford Nanopore Technologies, UK)

FIELD CRITICAL Make sure you acquire the offline version of MinKNOW for sequencing in remote areas with limited to no internet access from ONT. Ensure obtaining the country permits from ONT for use of the MinION in the respective country.

Basecalling

- Optional: Guppy (Oxford Nanopore Technologies, UK); can be used instead of live basecalling with MinKNOW.

Quality Control

- NanoPlot (https://github.com/wdecoster/NanoPlot)

Read De-multiplexing

· Guppy (Oxford Nanopore Technologies, UK)
CRITICAL Guppy only works for demultiplexing of ONT's indexing kits or custom single indexing.
or:
· minibar (https://github.com/calacademy-research/minibar)

DNA barcode read filtering and consensus generation
- NGSpeciesID (https://github.com/ksahlin/NGSpeciesID)

Comparison of the sequences against a database
- BLAST (ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/)

DNA barcode databases
- BOLD (www.boldsystems.org)
- NCBI nucleotide (https://www.ncbi.nlm.nih.gov/nuccore) or NCBI via BLAST (https://blast.ncbi.nlm.nih.gov/Blast.cgi)

Reagent Setup

Home-made Size-selection Beads
In order to prepare inexpensive home-made size selection beads, follow the protocol at: https://s3-us-west-2.amazonaws.com/oww-files-public/f/f8/SPRI_buffers_v2_2.pdf

70% Ethanol
In order to prepare 70% Ethanol, mix 70 ml of absolute Ethanol with 30 ml of nuclease-free water. CRITICAL Make fresh before use.

Primer and index dilutions and aliquots
To make 1:10 dilution of primer stocks (100 µM), mix 1 µl of primer stock solution with 9 µl of nuclease-free water to obtain a final concentration of 10 µM. CRITICAL We strongly recommend to aliquoting primer and index dilutions to avoid contaminating the stock solutions.

Procedure

Primer and index design
Adapting amplification primers (2-Step PCR protocol) ● TIMING 5min
In order to carry out multiplexing of a high number of amplicons on a single flow cell sequencing run, indexes need to be added to the amplicon in one of two ways: 1) using custom indexes designed by the user or 2) using ONT's PCR Barcoding Expansion kits (1-12, cat. no. EXPPBC001 or 1-96, cat. no. EXP-PBC096). First add ONT's universal tail sequences to your primer sequences. We recommend a custom indexing approach, as it is cost-effective and double index combinations allow the user to pool more than 96 samples on a single flow cell.

P1| Add ONT's universal tail (capitalized nucleotides) to your primer sequences before ordering the oligos:

5' TTTCTGTTGGTGCTGATATTGC-[project-specific forward primer sequence] 3'
5' ACTTGCCTGTCGCTCTATCTTC-[project-specific reverse primer sequence] 3'

CRITICAL For the 1-Step PCR protocol replace ONT's universal tail sequence here with the index sequences generated in P2| (e.g. 5'-ctgtagacaaatcaaggcctccag-[project-specific forward primer sequence]-3').

Designing custom indexes for multiplexing ● TIMING 5min
We recommend barcode_generator (https://github.com/lcomai/barcode_generator) to generate the custom index design. In order to maximize the number of indexes while minimizing costs, we recommend dual indexing (different indexes on the 5' and the 3' end of the DNA fragment). We describe how to generate custom indexes in P2| below. CRITICAL Be sure to take into account 5' - 3' strand orientation when appending custom indexes on forward and reverse primer sequences.

P2| Generate custom indexes using barcode_generator:

python3 barcode_generator_3.4.py none 24 20 20 40 50

This tool allows the user to specify a range of settings for the generation of indexes (also referred to as sequencing barcodes). For MinION-based sequencing, we recommend designing indexes with a length of 24bp to account for sequencing error rate. Due to improving error rates, e.g. with the new R10.3 pore, users can also test shorter indexes if desired (see [24]). Furthermore, we recommend specifying a hamming distance of at least one third to half the index length, e.g. 8, to avoid cross-contamination due to inaccurate demultiplexing of reads with high errors in the index sequences. The tool also allows the user to set a range of GC content for the indexes. We recommend generating indexes that have a similar GC content and annealing temperature to the universal tails used for efficient amplification (in the example we used a GC content range from 40 to 50, similar to ONT's indexes). The tool automatically removes indexes that show homopolymers >4bp, which could otherwise cause issues due to the MinION's error profile. This number can be in- or decreased as needed. Barcode generator will automatically attach ONT's universal tails to the indexes (forward: TTTCTGTTGGTGCTGATATTGC and reverse: ACTTGCCTGTCGCTCTATCTTC), which enables the addition of index sequences to DNA barcode amplicons using a second PCR amplification. Each forward and reverse index can be used in different combinations to increase the amount of experimental multiplexing. For example, 20 unique forward indexes and 20 unique reverse indexes can yield 400 sample combinations.

Example forward index (lower case) appended to ONT forward universal tail (upper case):

5’-ctgtagacaaatcaaggcctccagTTTCTGTTGGTGCTGATATTGC-3’

Example reverse index (lower case) appended to ONT reverse universal tail (upper case):

5’-tgtcgttagtagctcgttctacctACTTGCCTGTCGCTCTATCTTC-3’

DNA extraction and purification from collected sample TIMING 30min-2h
For the first step, collect and cut tissue into small pieces to ensure rapid lysis and high yields. CRITICAL Sterilize tools for manipulating and lysing tissues with a flame or bleach in between processing samples. Residual ethanol can affect downstream applications, so ensure the tools are dry.

Protocol for DNeasy Blood & Tissue Kit (Procedure for Tissue Sample)

E1| Add 180 µl Buffer ATL and 20 µl of proteinase K (25 mg/mL) to each sample. Incubate at 55°C for 1 hr.

E2| Mix well for 15 s. Add 200 µl Buffer AL to the sample, and mix thoroughly. Then add 200 µl ethanol (96–100%), and mix again thoroughly.

E3| Pipette the mixture into a DNeasy Mini spin column placed in a 2 ml collection tube and centrifuge for 1 min. Discard flow-through and collection tube. FIELD we have used microcentrifuges for this protocol that spin at 6,000 RPM, which we have found to be sufficient for spin-column-based DNA isolation.

E4| Place the DNeasy Mini spin column in a new 2 ml collection tube, add 500 µl Buffer AW1, and centrifuge for 1 min. Discard flow-through and collection tube.

E5| Place the DNeasy Mini spin column in a new 2 ml collection tube, add 500 µl Buffer AW2, and centrifuge for 3 min to dry the DNeasy membrane. Discard flow-through and collection tube.

E6| Place the DNeasy Mini spin column in a clean 1.5 ml or 2 ml microcentrifuge tube, and pipet 50-100 µl Buffer AE directly onto the DNeasy membrane. Incubate at room temperature for 1 min, and then centrifuge for 1 min to elute.

FIELD Protocol for alkaline lysis buffer-based HotSHOT DNA extraction method:

E1| Submerge small amount of tissue or small specimen in 50 µl of the alkaline lysis HotSHOT solution (25 mM NaOH, 0.2 mM disodium EDTA, pH of 12, see [27]).

E2| Incubate at 95°C for 20 minutes (some tissue types may require longer incubation) and then cool to 4°C

E3| Add an equal volume of neutralization reagent (40 mM Tris-HCl); mix well.

FIELD Protocol for the Chelex 100 Resin:

E1| Add 190 µl of 5% Chelex reagent (dissolved in DNAse-free distilled water) and 10 µl of proteinase K (25 mg/mL) to each sample. Up to 20% Chelex can be used depending on sample.

E2| Incubate at 55°C for 1 hr.

E3| Incubate at 100°C for 20 minutes.

PAUSE POINT Genomic DNA can be stored at 4° or -20°C until amplification and library preparation.

PCR amplification and indexing

PCR Amplification of the target region (same for the 1-Step and 2-Step PCR setups) TIMING 2.5 – 4h

1| In a 0.2 ml tube, set up a barcoding PCR reaction as follows:

Reagent	Volume (µl) per sample
DreamTaq Hot Start master mix	6.25
Forward primer (10µM)	2
Reverse primer (10µM)	2
Template DNA	1.25 (or up to 1 µg)
Water, nuclease-free	up to 12.5
Total	12.5

Place in thermocycler and run the following program:

PCR Program	Temperature (°C)	Duration	No. of cycles
Step 1 (Initial denaturation)	95	3 min	1
Step 2 (Denaturation)	95	30 sec	34

Step 3 (Annealing)	Variable, based on annealing temperature of the primer	30 sec	34
Step 4 (optional)	72	ramp	34
Step 5 (Extension)	72	30 sec	34
Step 6 (Final extension)	72	5 min	1
Step 7 (Hold)	4	∞	

CRITICAL When closing the lid (either strip or single cap) on the PCR tube, be careful not to touch the inside of the lid while handling.
FIELD CRITICAL The PCR settings may need to be adjusted depending on your amplicon length and choice of polymerase. For amplification with the DreamTaq™ Hot Start DNA Polymerase, elongation for 30 sec is sufficient for amplicons up to 1 kb. For longer amplicons, increase the elongation time by 1min/1kb. We recommend the use of the NEB Tm Calculator (https://tmcalculator.neb.com/) for calculating the annealing temperature of primers.
CRITICAL In the optional step 4 of the PCR program, the temperature is ramped from the primer annealing temperature to 72°C. This step was shown to improve the PCR yield. However, many field-friendly PCR machines do not have this option, in which case just continue straight from Step 3| to 5| in the PCR program.

2| Determine the concentration of the amplified barcodes either using the Agilent Bioanalyser, the TapeStation or the Qubit with the broad-range assay per the manufacturer's instructions. FIELD Alternatively, DNA concentrations can be roughly estimated based on intensity of gel electrophoresis bands in comparison to a DNA ladder.

PAUSE POINT Cleaned-up PCR products can be stored at 4°C until library preparation or up to a year at -20°C.

3| For the 2-step PCR protocol, CRITICAL The final concentration of amplified barcodes (from Step 2|) per 50 µl indexing reaction should be ~0.5 ng/µl (e.g., add 1 µl of a barcode solution with 25 ng/µl). Fill in the right amount in the table below (x). For barcodes >2,000 bp, the concentration and number of PCR cycles might need to be increased.

Reagent	Volume (µl) per sample
PCR Index forward (10µM)	1
PCR Index reverse (10µM)	1
Amplified barcode	x
DreamTaq Hot Start master mix	25
Nuclease-free water	fill up to 50
Total	50

Place in thermocycler and run the following program:

PCR Program	Temperature (°C)	Duration	No. of cycles
Step 1 (Initial denaturation)	95	3 min	1
Step 2 (Denaturation)	95	15 sec	12-15
Step 3 (Annealing)	55	15 sec	12-15
Step 4 (Extension)	72	15 sec	12-15
Step 5 (Final extension)	72	1 min	1
Step 6 (Hold)	4	∞	

CRITICAL Temperatures here were determined for the DreamTaq Hot Start master mix. Adjust accordingly for different lengths of amplicons and the type of polymerase being used.

4| Determine the concentration of the amplified barcodes either using the Agilent Bioanalyser, the TapeStation or the Qubit with the broad-range assay per the manufacturer's instructions. FIELD Alternatively, DNA concentrations can be roughly estimated based on intensity of gel electrophoresis bands in comparison to a DNA ladder.

CRITICAL Make sure that the addition of the indexes was successful by checking the amplicon size on the gel. You might see two bands if the index addition was not very efficient.

TROUBLESHOOTING

PAUSE POINT PCR products can be stored at 4°C or -20°C until library preparation.

Normalization and pooling of indexed barcode amplicons

The library preparation is the same for the standard MinION and the Flongle flow cell apart from the amount of starting DNA (see 5|) and the elution volume (13l|).

5| Pool all barcoded amplicons in the desired ratios to prepare 1 µg of pooled barcoded amplicons. Fill up with nuclease-free water to obtain a total of 53.5 µl if needed. If the volume of your pool exceeds 53.5 µl proceed directly to step 6|.

SUGGESTION Start the protocol for MinION flow cells with ~ 1 µg of DNA or <100-200 fmol. Be careful to properly pool amplicons in equimolar ratios for best downstream sequencing results. For use with the Flongle flow cell, ~500 ng is sufficient.

Cleanup and quantification of amplicons

6| Purify the amplified barcode regions using AMPure XP or homemade size-selection beads. SUGGESTION For amplicons below 1,000 bp, we recommend using a ratio of 0.6-1.0x of bead volume to PCR product volume. For amplicons >1,000 bp, use a ratio of 0.6x-0.8x. Take the beads out of the fridge and use them at room temperature.

(i) Mix the beads well so the liquid appears homogeneous and consistent in colour.
(ii) Take a new 1.5 ml tube, add the PCR product first, then the beads, and mix well by pipetting up and down (at least 10 times).
(iii) Incubate for 5 min at room temperature on a rotator mixer. FIELD Alternatively, carefully mix by inverting the tube by hand a couple of times.
(iv) Put the tubes in a magnetic rack and wait for solution to clear (3-5 min).
(v) Keeping the tubes in the magnetic rack, discard the cleared solution (supernatant) from the tubes without touching the beads.
(vi) Wash Step 1: keep the tubes in the magnetic rack and add 200µl 70% ethanol.
(vii) Let the tubes sit for 1 min to allow beads to settle; then remove the ethanol.
(viii) Wash Step 2: keep the tubes in the magnetic rack and add 200µl 70% ethanol.
(ix) Let the tubes sit for 1 min to allow beads to settle; then remove the ethanol.
(x) Seal the tubes and spin down briefly.
(xi) Return to the magnetic rack and wait for 1 min.
(xii) Remove any remaining ethanol, being careful not to touch the bead pellet.
(xiii) Leave the tubes open until the beads are dry (usually 30 sec to 2 min). Do not over-dry, as this might decrease the yield.

(xiv) Add 53.5 µl nuclease-free water directly to the bead pellet. Mix well by pipetting up and down at least 10 times.
(xv) Incubate for 3 min at room temperature.
(xvi) Centrifuge briefly, then put back in the magnetic rack for 2-3 min or until the solution is clear.
(xvii) Remove 53.5 µl of the supernatant and transfer to a clean 0.2 ml PCR tube.

Library preparation and sequencing

Library Preparation (using the SQK–LSK109 sequencing kit) TIMING ~2,5-3h
7| End-prep in a 0.2 ml PCR tube:

Reagent	Volume (µl) per sample
Ultra II End Prep Reaction Buffer	3.5
Ultra II End Prep Enzyme Mix	3
Pooled Barcoded Amplicons (from Step 6|)	53.5
Total	60

8| Incubate at room temperature (20°C) for 10 min.
SUGGESTION During this step, take out the AMX and BLUNT/TA ligase master mix and place on ice.
9| Incubate at 65°C for 10 min in the Thermocycler.
10| Place on ice to cool down for 30 sec.

FIELD CRITICAL ONT recommends that the end-prepped DNA sample be subjected to a bead clean-up (see step 6|). This clean-up can be omitted for simplicity and to reduce library preparation time. However, it has been observed that omission of this clean-up can reduce subsequent adapter ligation efficiency, increase the prevalence of chimeric reads, and lead to an increase in pores being unavailable for sequencing. If omitting the clean-up step, proceed directly to step 11|.

Adapter ligation and clean-up

11| In a 1.5 ml Eppendorf DNA LoBind tube, mix in the following order:

CRITICAL Perform this step on ice or ice packs.

Reagent	Volume (µl) per sample

DNA sample from step 10|	60
NEBNext Quick T4 DNA Ligase	10
Ligation Buffer (LNB)	25
Adapter Mix (AMX)	5
Total	100

12| Mix gently, incubate at room temperature for 10-20 minutes
13| Purify the library with a bead cleanup:
(i) Add 40 µl of resuspended beads and mix by flicking tube.
(ii) Incubate at room temperature for 10 min.
(iii) Spin briefly and place in the magnetic rack until solutions clears (2-3 min).
(iv) Remove the supernatant, avoiding the pellet.
(v) Wash Step 1: add 250 µl Short Fragment Buffer (SFB) at room temperature and resuspend by gently flicking.
(vi) Spin briefly and place in the magnetic rack until solution clears.
(vii) Remove the supernatant, avoiding the pellet.
(viii) Wash Step 2: add 250 µl SFB at room temperature and resuspend by flicking.
(ix) Spin briefly and place in the magnetic rack until solution clears.
(x) Remove the supernatant, avoiding the pellet.
(xi) Spin down and remove all residual supernatant, allow to dry for ~30 seconds.
(xii) Remove tube from magnetic rack and add 15 µl (standard MinION flow cell) or 7 µl (Flongle flow cell) of Elution Buffer (EB) and resuspend beads by flicking.
(xiii) Incubate at room temperature for 10 min.
(xiv) Spin down briefly and pellet the beads on a magnet until the eluate is clear and colorless.
(xv) Transfer 15 µl (standard MinION flow cell) or 7 µl (Flongle flow cell) of eluate into a clean 1.5 ml Eppendorf DNA LoBind tube. The prepared library is used for loading into the flow cell.
(xvi) Optional: Quantify 1µl of eluted sample using a Qubit fluorometer
PAUSE POINT Store the library on ice until ready to load.

CRITICAL ONT recommends loading 5–50 fmol of this final prepared library onto R9.4.1 flow cells. For R10.3 flow cells, ONT recommends loading 25-75 fmol. Loading more than 50 fmol of DNA can have a detrimental effect on throughput. Dilute the library in Elution Buffer if required. If you are using the Flongle for sample prep development, ONT recommends loading 3-20 fmol instead. To calculate the right fmol we recommend to use the Promega online tool: at https://promega.com/resources/tools/biomath.

14| Preparing flow cell for sequencing

(i) Thaw the Sequencing Buffer (SQB), Loading Beads (LB), Flush Tether (FLT) and one tube of Flush Buffer (FLB) at room temperature before placing the tubes on ice as soon as thawing is complete.

(ii) Mix the Sequencing Buffer (SQB) and Flush Buffer (FLB) tubes by mixing, spin down and return to ice.

(iii) Spin down the Flush Tether (FLT) tube, mix by pipetting, and return to ice.

CRITICAL Follow the specific steps for each flow cell type below. We also recommend to check ONT's current protocol for updates and changes (https://community.nanoporetech.com/protocols).

Standard MinION flow cell (Step 14| (iv) to 19|)

(iv) Open the lid of the nanopore sequencing device and slide the flow cell's priming port cover clockwise so that the priming port is visible.

Priming and loading (standard MinION flow cell):(v) Set a P1000 pipette to 200 µl.

(vi) Insert the tip into the priming port.

CRITICAL Care must be taken when drawing back buffer from the flow cell. The array of pores must be covered by buffer at all times. Removing more than 20-30 µl risks damaging the pores in the array.

CRITICAL After opening the priming port, check for a small bubble under the cover. Draw back a small volume to remove any bubble (20-30 µl).

(vii) Turn the wheel until the dial shows 220-230 µl, or until you can see a small volume of buffer entering the pipette tip.

CRITICAL Visually check that there is continuous buffer from the priming port across the sensor array.

TROUBLESHOOTING

(viii) Prepare the flow cell priming mix: add 30 µl of thawed and mixed Flush Tether (FLT) directly to the tube of thawed and mixed Flush Buffer (FLB), and mix by pipetting up and down.

(ix) Load 800 µl of the priming mix into the flow cell via the priming port, avoiding the introduction of air bubbles. Wait for 5 minutes.

15| Library dilution for sequencing (standard MinION flow cell)

(i) Thoroughly mix the contents of the Loading Beads (LB) tube by pipetting.

CRITICAL The Loading Beads (LB) tube contains a suspension of beads. These beads settle very quickly. It is vital that they are mixed immediately before use.

(ii) Thoroughly mix the contents of SQB (ONT Ligation Kit) and LB (ONT library loading bead kit) tubes by pipetting.

Reagent	Volume (µl) per sample
Sequencing Buffer SQB	37.5
Loading Beads II (LBII), mixed immediately before	25.5
Library	12
Total	75

16| Complete the flow cell priming:
(i) Gently lift the SpotON sample port cover to make the SpotON sample port accessible.
(ii) Load 200 µl of the priming mix into the flow cell via the priming port (not the SpotON sample port), avoiding the introduction of air bubbles again by reverse pipetting.
17| Mix the prepared library gently by pipetting up and down just prior to loading.
18| Add 75 µl of sample to the flow cell via the SpotON sample port in a dropwise fashion. Ensure each drop flows into the port before adding the next.
CRITICAL The library is loaded dropwise without putting the pipette tip firmly into the port. Take care to avoid introducing any air during pipetting.
19| Gently replace the SpotON sample port cover, making sure the bung enters the SpotON port, then close the priming port and the MinION lid.
TROUBLESHOOTING
PAUSE POINT

Flongle flow cell (Step 14| (iv) to 20|)

(iv) Peel back the seal tab until the sample port is exposed.

Priming and loading (Flongle flow cell):

CRITICAL The library is loaded by putting the pipette tip into the port. Take care to avoid introducing any air during pipetting.

(v) Prepare the flow cell priming mix in a new tube: add 117 µl of mixed Flush Buffer (FLB) and 3 µl of mixed Flush Tether (FLT), and mix by pipetting up and down.
(vi) Load 120 µl of the priming mix into the sample port, avoiding the introduction of air bubbles by reverse pipetting. Wait for 5 minutes.

TROUBLESHOOTING

15| Library dilution for sequencing (Flongle flow cell)
(i) Thoroughly mix the contents of the LB tube by pipetting.
CRITICAL The Loading Beads (LB) tube contains a suspension of beads. These beads settle very quickly. It is vital that they are mixed immediately before use.
(ii) Thoroughly mix the contents of SQB (ONT Ligation Kit box) and LB (ONT library loading bead kit box) tubes by pipetting

Reagent	Volume (µl) per sample
SQB	15
LB	10
Library	5
Total	30

16| Mix the prepared library gently by pipetting up and down just prior to loading.
17| Add 30 µl of sample to the flow cell via the sample port by by dialing down the pipetting volume.
18| Gently reseal the sample port with the seal tab, making sure the sample port is sealed properly.
19| Bring the top (wheel icon section) to its original position and close the MinION lid.
TROUBLESHOOTING

From here on, the steps are the same for the standard MinION Mk1B flow cells and the Flongle flow cell using MinKNOW on a computer.

Getting ready for Sequencing TIMING 0.5-1h

20| Set up the MinKNOW software:
(i) Plug in your MinION device.
(ii) Open MinKNOW on the computer.
(iii) Select the sequencing device connected to the computer, then select the ‘Start Sequencing’ option on the Start homepage.
(iv) Enter information such as experiment name, sample ID and flow cell type, and select kit type.
(v) Optional: Turn basecalling OFF if you want to basecall the data later on your laptop or a server, or if you plan to sequence for longer than a day.

(vi) Run options: 24h or 72h depending on the maximum run time required, output file fasta5, and save to the local drive.

Sequencing TIMING 1-3h (up to 3d depending on the required amount of data)
21| Start sequencing by pressing the START button.
The run time depends on the amount of coverage desired per individual DNA barcode (amplicon). As a rule of thumb, use a rate of 10,000 sequenced reads per minute for the total run time calculations. CRITICAL We recommend aiming for >1,000x coverage per individual amplicon. This should give you a good representation for each DNA barcode even in the presence of slightly uneven pooling of the different amplicons.
TROUBLESHOOTING

22| Optional: To recover flow cell pores and re-use the flow cell after a run, a nuclease-flush protocol can be performed (see ONT flow cell wash kit).

DNA sequence processing workflow

(up to 1d for large amounts of data if the basecalling still needs to be performed, depending on the server/computer capacity)

Here, we outline a bioinformatics protocol, which can be scaled up to analyze a large number of pooled amplicon samples. The bioinformatic processing can be carried out on UNIX platforms, on Windows with an Ubuntu terminal or on Mac platforms using the terminal (command line). In the example steps, we provide names for the files and file directories for clarity; however, these should be modified to match your file system when running. We also provide example parameters, which should be adapted to fit the individual DNA barcode preferences. The example commands and related data can be found https://github.com/ksahlin/NGSpeciesID (see EXAMPLE WORKFLOW section of the readme documentation).

B1| Basecalling and quality check

In this step the raw current profiles are converted into fastq formatted sequences. We recommend using the live basecalling included in MinKNOW for shorter sequencing runs (1-3 hours). Alternatively, for longer runs or if the data are being re-basecalled, use the following command:

guppy_basecaller --input_path minKNOW_input/ --save_path basecalled_fastqs/ -c dna_r9.4.1_450bps_fast.cfg --recursive --disable_pings

This will carry out basecalling using Guppy for data sequenced on flow cells using the R9.4 pore (standard MinION flow cell or Flongle). For the R10.3 based standard MinION flow cell, use -c dna_r10.3_450bps_fast.cfg. Furthermore, in addition to Fast models, Guppy also offers High-accuracy models, which have a much longer runtime, but produce more accurate basecalls. Information on these models and the current model

names can be found in the Guppy documentation. All the fastq files will be stored in the folder: *basecalled_fastqs*. We usually filter the reads after the basecalling, but if you want to filter at this point, use: *--min_qscore 7*, which roughly corresponds to a basecall accuracy of 85%. The tool can be parallelized using the command *--num_callers*. The option *--disable_pings* disables both the transmission of telemetry pings and automatic upload of crash reports, which require internet connection.

CRITICAL ONT uses different basecalling model names in different Medaka versions. Check the Medaka manual for the model name that corresponds to your version of Medaka.

B2| Go to the folder with the fastq files generated by Guppy.

B3| Concatenate all the read files into one large file:

*cat *.fastq > sequencing_reads.fastq*

B4| Check raw read quality/stats with NanoPlot:

NanoPlot --fastq_rich sequencing_reads.fastq -o sequencing_run -p sequencing_run

This will create a html based report in the folder sequencing_run (option *-o*). The prefix of the output files can be specified using the option *-p*.

B5| Demultiplexing of the sequencing data

Here we provide commands for demultiplexing using Guppy (B5a|) or minibar (B5b|). Minibar allows for demultiplexing of custom indexed libraries, but requires more preparation as it will need a list of all the index and primer combinations. Guppy is easy to run, but only works for ONT indexing kits. Example files can be found in: Supplementary Data 2 (a file containing 3,000 reads in fastq format) and Supplementary Data 3 (the index file used for demultiplexing with minibar), and on https://github.com/ksahlin/NGSpeciesID.

B5a| Demultiplexing using Guppy

guppy_barcoder -i sequencing_reads.fastq -s demultiplex_folder --trim_barcodes --disable_pings

The demultiplexing stringency can be adjusted with the *--min_score* flag. The default is 60. Increasing this threshold results in fewer demultiplexed reads, but more accurate read assignment. The option *-s* sets the output directory for the demultiplexed read files. The indexes will be automatically trimmed with the flag *--trim_barcodes*. The demultiplexing can be parallelized using *--worker_threads*. See https://community.nanoporetech.com/protocols/Guppy-protocol/v/gpb_2003_v1_revv_14dec2018/barcoding-demultiplexing for information on how to demultiplex custom indexes using Guppy.

or:

B5b| Demultiplexing using minibar

python minibar.py indexes.txt sequencing_reads.fastq -T -F -e 3 -E 11

This will carry out demultiplexing of sequencing reads (*sequencing_reads.fastq*: Supplementary Data 1) using a set of pre-specified index and primer combinations (*indexes.txt*Supplementary Data 2). When using the option *-T*, minibar will trim the index and primer sequences on both ends. The option *-F* will tell minibar to output an individual fastq sequencing file for each index and primer combination. The stringency can be adjusted using the options *-e* and *-E*, which specify the edit distance allowed between indexes (*-e*) and primers (*-E*). Note that the edit distance is higher for primer regions, which is due to possible ambiguities in primer sequences. Lower edit distances will result in fewer reads, but more accurate read assignment. An example index file for minibar can be found in Supplementary Data 2.

B6| Read filtering, clustering, consensus generation and polishing

NGSpeciesID --ont --consensus --sample_size 500 --m 800 --s 100 --medaka --primer_file primers.txt --fastq barcode0.fastq --outfolder barcode0_consensus

NGSpeciesID uses fastq or fasta files as read input. The output folder can be specified using *--outfolder*. We suggest filtering the sequencing reads for Phred quality score and length. This allows the user to remove reads with lower qualities, as well as many chimeric reads that result from polymerase jumping during the amplification or non-target reads. We recommend filtering for Phred score higher than 10. This can be lowered if needed (e.g. to increase the number of available reads) or increased if the data shows a high average quality (>13). To avoid chimeric and non-target reads, adjust the intended target length (*--m*) and maximum deviation from target length (*--s*) values for each amplicon. For example, if the target amplicon has a length of 800bp, we recommend removing reads < 600 bp and > 1,000 bp. As too many reads can add noise to the clustering and polishing, we recommend subsampling the read data to 300 to 1,000 reads per sample using the option: *--sample_size* (see [23,34]). Use the *--ont* flag for MinION read data. The tool automatically generates consensus sequences for all read clusters that include more than 10% of all the reads. This value can be changed if the output is expected to include several consensus sequences (e.g. for species pools). This technically allows for the creation of individual consensus sequences for mixed samples. However, we have to caution that this might not work for distinguishing closely related species (i.e., within the same genera) due to the MinION's high error rate of raw sequence reads. The tool automatically carries out consensus sequence polishing using all the reads that make up the cluster from which the individual consensus sequence was created when the flags *--medaka* or *--racon* are used. The *--primer_file* flag defines the fasta file including the primer sequences to be removed from the consensus sequence (optional step). A primer example fasta file can be found in Supplementary Data 3 and on https://github.com/ksahlin/NGSpeciesID.

SUGGESTION We suggest using a bash script for rapid processing of multiple files. For example, create a text file called consensus.sh. In this file, save:

```
for file in *.fastq; do
    bn=`basename $file .fastq`
        NGSpeciesID --ont --consensus --sample_size 500 --m 800 --
        s 100 --medaka --primer_file Supplementary_File4_primer.txt
        --fastq $file --outfolder ${bn}
done
```

An example file can be found on https://github.com/ksahlin/NGSpeciesID. Then execute the file in your UNIX or Mac terminal using the command:

bash consensus.sh

B7| In the last step, the sequences are compared against a database (e.g. NCBI nucleotide or BOLD for Eukaryotes) using BLAST. This can be carried out online, or locally on a laptop using command line BLAST. If conducting BLAST offline, remember to download or curate a reference database locally on a laptop in fasta format.

(i) First the database.fasta file has to be converted into a BLAST readable format. CRITICAL This only needs to be done for each database file once.

makeblastdb -in database.fasta -dbtype nucl -out database

This will create a nucleotide database of the fasta file with the name database. You can create targeted databases or download the respective taxon groups from NCBI nucleotide or BOLD.

(ii) BLAST the consensus sequences against the database.

```
blastn -db database -query barcode0_consensus.fasta -
outfmt 6
-out barcode0_consensus_blast.out
```

In this example, we compare the consensus sequence of barcodes against the respective database. BLASTn can also be parallelized to save time when more than one thread is available (option *-num_threads*). Furthermore, it provides different output formats, which can be specified with *-outfmt* (6 is commonly used).

BLAST provides several output statistics characterizing the quality of a match:

Bit score: A measure of the quality of the alignment between the reference and query. Generally, the higher the bit score, the better the quality of the hit.

Query Coverage: A measure for how much of the query (here our barcode consensus sequence) matches the reference. Coverage should be close to 100% for DNA barcodes.

E-value: A measure for the likelihood that a given sequence match is purely by chance (depends on the database size). The smaller the E-value, the better the match.

Percent Identity: A measure for sequence similarity between the query and the reference database match. The percent identity for species assignment can differ between taxa.

CRITICAL Even though the percent identity is the most intuitive measure of quality of a database match, we recommend not relying on it as the sole measure for taxonomic assignment. Furthermore, note that depending on the taxon studied, the available sequence database can be very poorly covered and thus the best BLAST hit in the database used might not be the actual species or closest relative. In addition, some DNA barcodes are more suited than others for species delineation for certain taxa, and so even a 100% match might not be a reliable species assignment (particularly in very conserved markers). We recommend looking at several BLAST hits to check whether other species match with equally good or poor quality and incorporating phylogenetic analyses, before deciding on a final taxonomic assignment.

Troubleshooting

Troubleshooting guidance can be found in Table 1.

Table 1| Troubleshooting table

Step	Problem	Possible reason	Solution

2- 4	Low or no PCR product yield	Not enough template	Assess quantity of initial DNA extraction. Increase concentration of DNA template or perform DNA extraction of sample again Use electrophoresis to check DNA quality
		Template is degraded	Check expiration date of components
		Reaction mix components are compromised	Aliquot biological components of reaction mixture and avoid multiple freeze-thaw cycles
		Reaction is missing polymerase or other reaction component	Make sure each component was added to PCR reaction
		Over dried beads	Do not dry for more than 1 min
		There is residual ethanol and DNA does not elute well	Make sure to remove any remaining ethanol using pipette P20
		Low amplification efficiency	Try increasing the amount of primers used or a more efficient polymerase such as the KAPA HiFi HotStart ReadyMix Consider performing an additional bead clean-up step, as described in 6|.

15	A small bubble may remain even after you have removed more than 20-30µl		Very slowly add the flow cell priming mix and monitor the bubble. If it becomes dislodged, immediately stop and try drawing back enough volume to remove the bubble
18	The library is not flowing into the SpotON port after each drop	The SpotON port is blocked with the beads	Carefully draw back the library and try again. More priming buffer may have to be added
21	MinKNOW fails to start script	Problem with MinKNOW installation	Reinstall MinKNOW from scratch and restart the script

Anticipated Results

The described protocol can be carried out under a variety of settings, including in the field, in standard laboratories, or in classroom environments. It was developed to require minimal equipment and funding to process samples on-site, which makes the protocol particularly useful for hands-on genomics teaching programs or for conducting genetic assessments of biological samples in areas with limited research funding or infrastructure. All the steps can be performed on battery power, which enables the user to run experiments even if constant electricity is lacking. The bioinformatic processing can be carried out on a laptop with minimal computational experience. The outlined pipeline generates polished consensus sequences for each DNA barcode amplified. It will output individual consensus files (in fasta format), which can be used for downstream analyses such as taxonomic assignments or phylogenetic analyses. Furthermore, this nanopore-based protocol is highly customizable for multiplexing any type of amplicon that the user is interested in, including long-range amplicons, or amplicon mixes generated using multiplex PCR.

References

1. Barnosky, A. D. *et al.* Approaching a state shift in Earth's biosphere. *Nature* 486, 52–58 (2012).
2. Dirzo, R. *et al.* Defaunation in the Anthropocene. *Science* 345, 401–406 (2014).
3. Diagne, C. *et al.* High and rising economic costs of biological invasions worldwide. *Nature* 592, 571–576 (2021).

4. Seebens, H. *et al.* Global rise in emerging alien species results from increased accessibility of new source pools. *PNAS* 115, E2264–E2273 (2018).

5. Hebert, P. D. N., Ratnasingham, S. & de Waard, J. R. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270, S96–S99 (2003).

6. Ratnasingham, S. & Hebert, P. bold: The Barcode of Life Data System. *Mol. Ecol. Notes* 7, 355–364 (2007).

7. Mizrachi, I. *Chapter 1: GenBank: The Nucleotide Sequence Database*.

8. Shokralla, S. *et al.* Massively parallel multiplex DNA sequencing for specimen identification using an Illumina MiSeq platform. *Sci Rep* 5, 1–7 (2015).

9. Martinez, B. *et al.* Technology innovation: advancing capacities for the early detection of and rapid response to invasive species. *Biol Invasions* 22, 75–100 (2020).

10. Krehenwinkel, H., Pomerantz, A. & Prost, S. Genetic Biomonitoring and Biodiversity Assessment Using Portable Sequencing Technologies: Current Uses and Future Directions. *Genes* 10, 858 (2019).

11. Menegon, M. *et al.* On site DNA barcoding by nanopore sequencing. *PLoS One* 12, (2017).

12. Pomerantz, A. *et al.* Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid biodiversity assessments and local capacity building. *Gigascience* 7, (2018).

13. Blanco, M. B. *et al.* Next-generation technologies applied to age-old challenges in Madagascar. *Conserv Genet* 21, 785–793 (2020).

14. Chang, J. J. M., Ip, Y. C. A., Ng, C. S. L. & Huang, D. Takeaways from Mobile DNA Barcoding with BentoLab and MinION. *Genes* 11, 1121 (2020).

15. Johnson, S. S., Zaikova, E., Goerlitz, D. S., Bai, Y. & Tighe, S. W. Real-Time DNA Sequencing in the Antarctic Dry Valleys Using the Oxford Nanopore Sequencer. *J Biomol Tech* 28, 2–7 (2017).

16. Quick, J. *et al.* Real-time, portable genome sequencing for Ebola surveillance. *Nature* 530, 228–232 (2016).

17. Faria, N. R. *et al.* Mobile real-time surveillance of Zika virus in Brazil. *Genome Medicine* 8, 97 (2016).

18. Watsa, M., Erkenswick, G. A., Pomerantz, A. & Prost, S. Portable sequencing as a teaching tool in conservation and biodiversity research. *PLOS Biology* 18, e3000667 (2020).

19. Salazar, A. N. *et al.* An educational guide for nanopore sequencing in the classroom. *PLOS Computational Biology* 16, e1007314 (2020).

20. Jain, M., Olsen, H. E., Paten, B. & Akeson, M. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. *Genome Biol* 17, 239 (2016).

21. Weirather, J. L. *et al.* Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Res* 6, (2017).

22. Wick, R. R., Judd, L. M. & Holt, K. E. Deepbinner: Demultiplexing barcoded Oxford Nanopore reads with deep convolutional neural networks. *PLoS Comput Biol* 14, (2018).

23. Krehenwinkel, H. *et al.* Nanopore sequencing of long ribosomal DNA amplicons enables portable and simple biodiversity assessments with high phylogenetic resolution across broad taxonomic scale. *Gigascience* 8, (2019).

24. Srivathsan, A. *et al.* Rapid, large-scale species discovery in hyperdiverse taxa using 1D MinION sequencing. *BMC Biology* 17, 96 (2019).

25. Labrador, K., Agmata, A., Palermo, J. D., Follante, J. & Pante, Ma. J. Authentication of processed Philippine sardine products using Hotshot DNA extraction and minibarcode amplification. *Food Control* 98, 150–155 (2019).

26. Seah, A., Lim, M. C. W., McAloose, D., Prost, S. & Seimon, T. A. MinION-Based DNA Barcoding of Preserved and Non-Invasively Collected Wildlife Samples. *Genes* 11, 445 (2020).

27. Truett, G. e. *et al.* Preparation of PCR-Quality Mouse Genomic DNA with Hot Sodium Hydroxide and Tris (HotSHOT). *BioTechniques* 29, 52–54 (2000).

28. Knot, I. E., Zouganelis, G. D., Weedall, G. D., Wich, S. A. & Rae, R. DNA Barcoding of Nematodes Using the MinION. *Front. Ecol. Evol.* 8, (2020).

29. Maestri, S. *et al.* A Rapid and Accurate MinION-Based Workflow for Tracking Species Biodiversity in the Field. *Genes* 10, 468 (2019).

30. Wurzbacher, C. *et al.* Introducing ribosomal tandem repeat barcoding for fungi. *Molecular Ecology Resources* 19, 118–127 (2019).

31. Kress, W. J. & Erickson, D. L. DNA Barcodes: Methods and Protocols. in *DNA Barcodes: Methods and Protocols* (eds. Kress, W. J. & Erickson, D. L.) 3–8 (Humana Press, 2012). doi:10.1007/978-1-61779-591-6_1.

32. Zizka, V. M. A., Elbrecht, V., Macher, J.-N. & Leese, F. Assessing the influence of sample tagging and library preparation on DNA metabarcoding. *Molecular Ecology Resources* 19, 893–899 (2019).

33. Lange, V. *et al.* Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing. *BMC Genomics* 15, 63 (2014).

34. Vasiljevic, N. *et al.* Developmental validation of Oxford Nanopore Technology MinION sequence data and the NGSpeciesID bioinformatic pipeline for forensic genetic species identification. *Forensic Science International: Genetics* 53, 102493 (2021).

35. Srivathsan, A. *et al.* A MinION™-based pipeline for fast and cost-effective DNA barcoding. *Molecular Ecology Resources* 18, 1035–1049 (2018).

36. Sahlin, K., Lim, M. C. W. & Prost, S. NGSpeciesID: DNA barcode and amplicon consensus generation from long-read sequencing data. *Ecology and Evolution* 11, 1392–1398 (2021).

37. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research* 41, D590–D596 (2013).

38. Sahlin, K. & Medvedev, P. De Novo Clustering of Long-Read Transcriptome Data Using a Greedy, Quality-Value Based Algorithm. in *Research in Computational Molecular Biology* (ed. Cowen, L. J.) 227–242 (Springer International Publishing, 2019). doi:10.1007/978-3-030-17083-7_14.

39. Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 27, 737–746 (2017).

40. Daily, J. Parasail: SIMD C library for global, semi-global, and local pairwise sequence alignments. *BMC Bioinformatics* 17, 81 (2016).

41. Byagathvalli, G., Pomerantz, A., Sinha, S., Standeven, J. & Bhamla, M. S. A 3D-printed hand-powered centrifuge for molecular biology. *PLOS Biology* 17, e3000251 (2019).

Acknowledgements

We thank Luca Comai for the adjustment of the barcode_generator tool to work for our MinION indexing protocol. We thank Oxford Nanopore Technologies for providing technical support, and for making the offline MinKNOW software available to us. We also thank Hitomi Asahara and the UC Berkeley DNA Sequencing Facility for technical support. We thank Tropical Herping, Rainforest Expeditions, Field Projects International and the Inkaterra Guides Field Station for support during fieldwork.

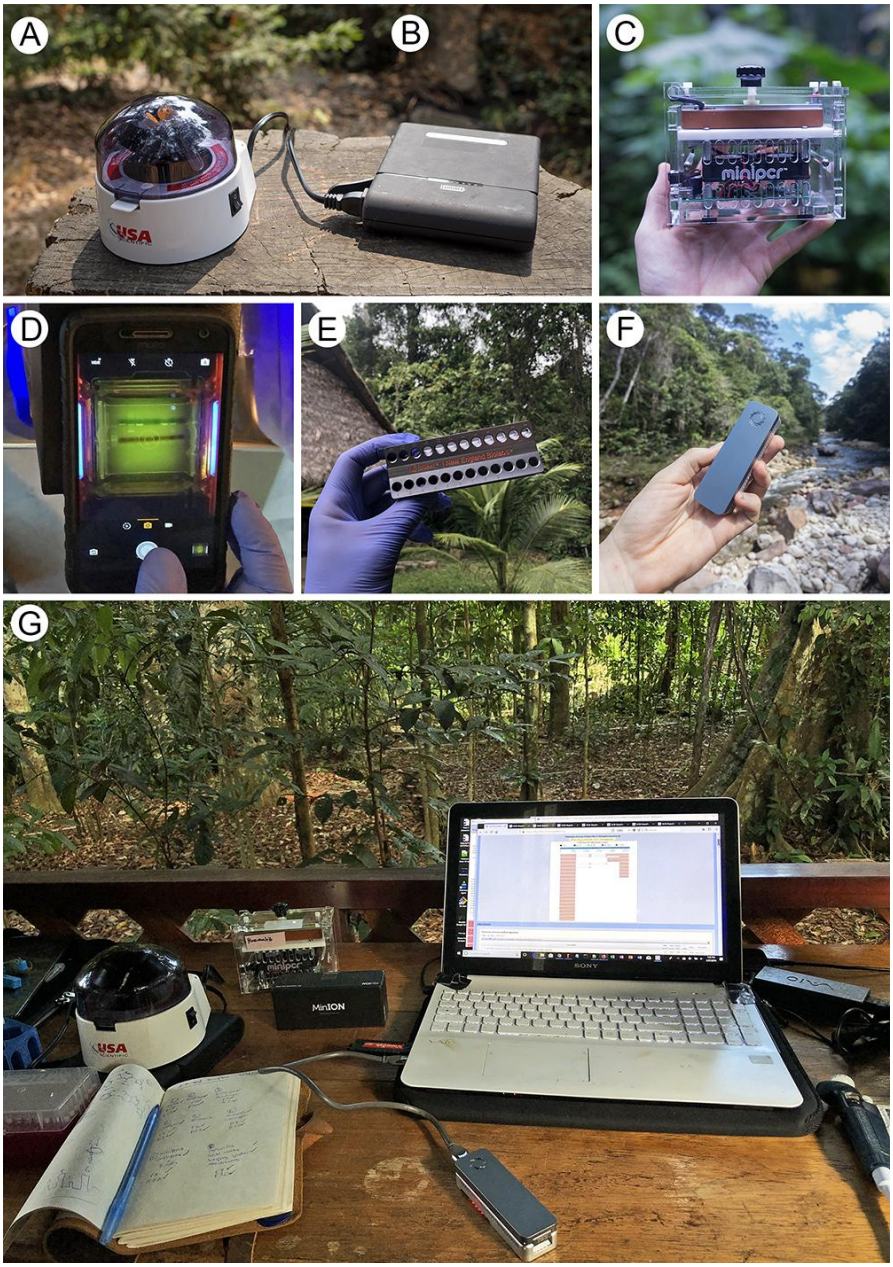


Figure 4.1: Examples of equipment used to carry out field-deployed DNA amplicon sequencing.

(A) Benchtop centrifuge connected to (B) external power source (RAVPower) to carry out DNA extraction steps. (C) Small thermocycler (miniPCR) to carry out polymerase chain reaction (PCR) amplification or heat block steps. (D) Miniaturized gel electrophoresis system (MiniOne) and mobile phone to visualize PCR amplification products. (E) Small magnetic rack to perform bead cleanups of PCR product and during ONT library preparation steps. (F) The portable MinION sequencer (Oxford Nanopore Technologies). (G) Example setup of portable genomics tools sequencing DNA amplicons in the Amazon rainforest.

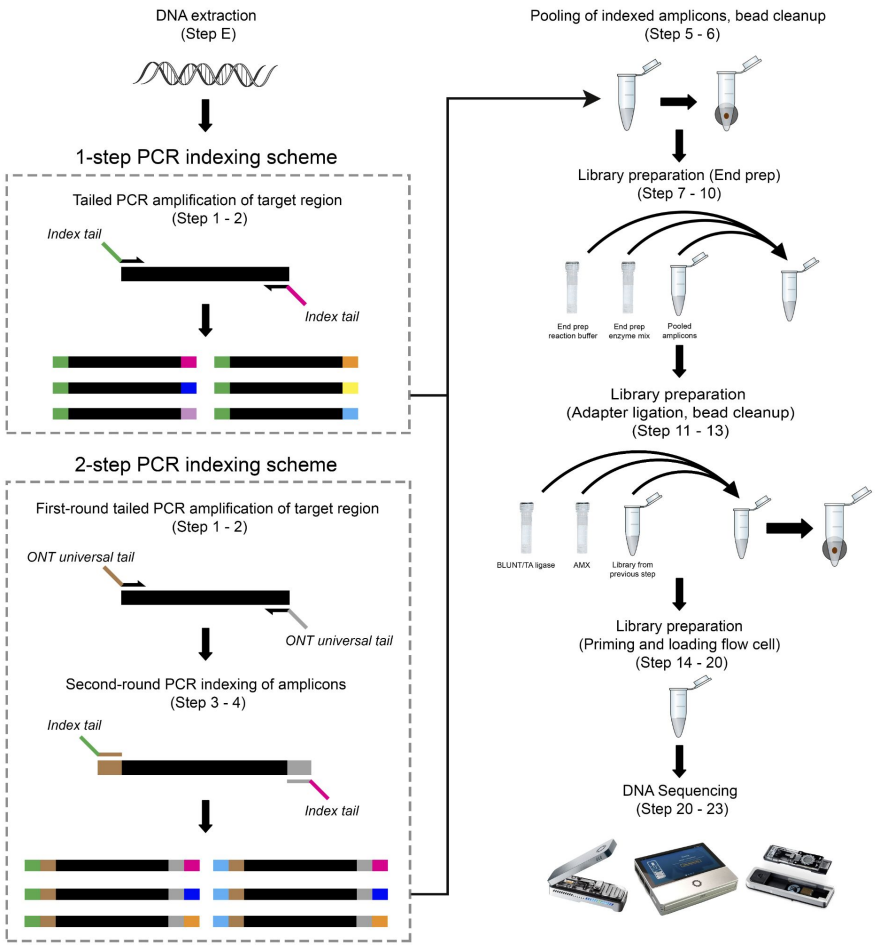


Figure 4.2: Illustration of the steps involved in the laboratory part of the protocol.
After isolating DNA from the samples, genetic regions of interest are amplified via polymerase chain reaction (PCR). Indexes can be added to the amplicons using a 1-Step or 2-Step PCR protocol in order to pool and sequencing high numbers of amplicons from numerous samples. After pooling, library preparation is carried out and can be sequenced using the portable MinION Mk1B (connected to a laptop or mini-server) or Mk1C (standalone) platforms.

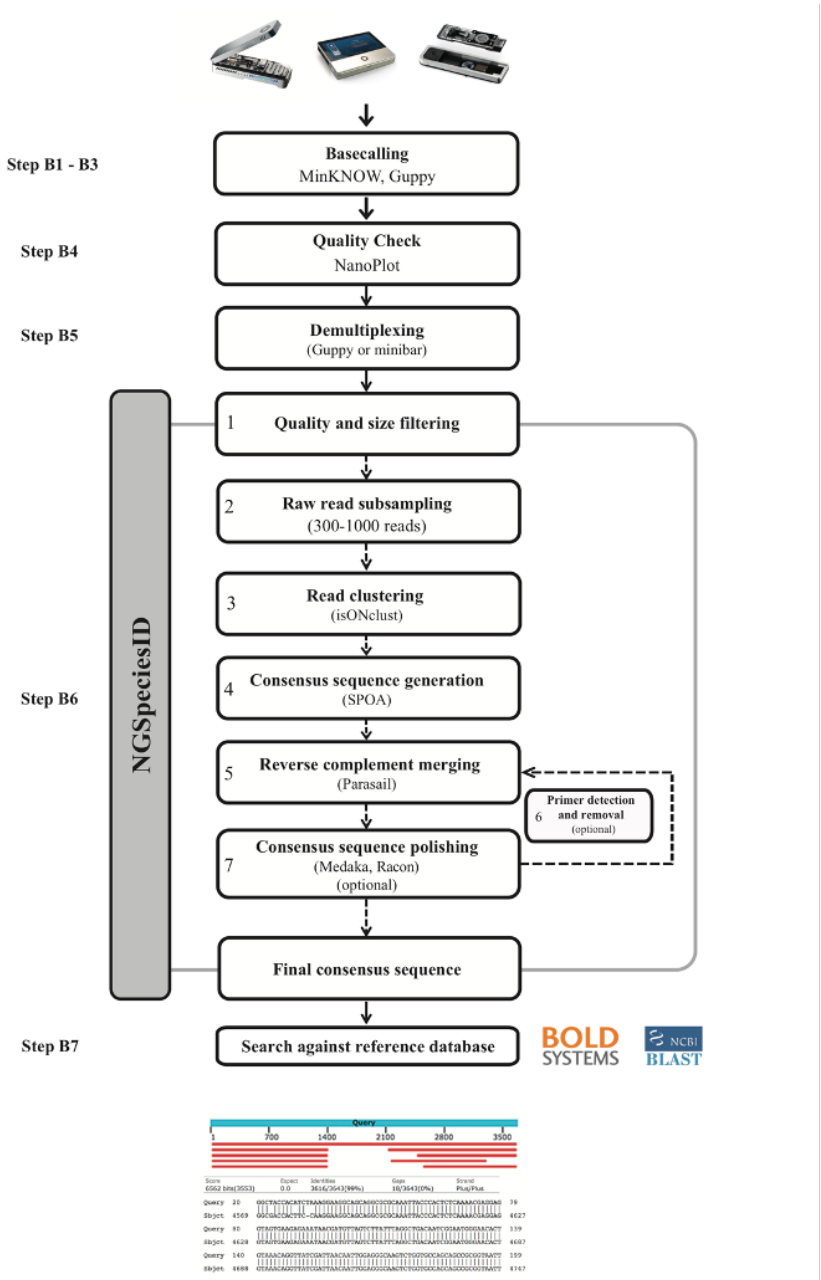


Figure 4.3: Illustration of the steps involved in the bioinformatic part of the protocol.

After the sequencing data have been generated, DNA amplicons are basecalled, quality checked, and demultiplexed. The pro-gram NGSpeciesID can then be used to perform downstream processing in one step, including quality and size filtering, subsampling, read clustering, and final consensus generation for each of the DNA amplicons. The final polished consensus sequences can subsequently be compared against reference databases, such as the NCBI or BOLD.