Genomics of Mimicry in Neotropical Corydoradinae Catfishes

Emily Phelps

A thesis submitted for the degree of
Doctor of Philosophy

University of East Anglia, UK

School of Biological Sciences

September 2024

100311838

All paintings courtesy of Ian Fuller

Table of Contents

List of Figures

List of Tables

List of Supplementary Figures

Abstract

Protective mimicry is a phenomenon that has fascinated evolutionary ecologists since the inception of the field 162 years ago. There are two types of protective mimicry, Batesian and Müllerian. Batesian mimicry is characterised by palatable taxa evolving to resemble unpalatable taxa, whereas Müllerian mimicry occurs when unpalatable taxa evolve a resemblance. Despite the longstanding interest, relatively little is known about the genetics underpinning mimetic phenotypes. The Corydoradinae is a species rich and colourful subfamily of Neotropical catfish, with a propensity to form Müllerian mimetic communities. As such, it represents a valuable system through which we can expand our understanding of the evolution of mimetic systems. Among the Corydoradinae, there are 60 species participating in 27 mimicry rings, all defended by venom glands and sharp lockable spines.

In chapter 2, we generate novel genomic resources, including a scaffold level assembly for *Corydoras fulleri* as well as genome annotations for both *C. fulleri and Hoplisoma metae*. Utilizing these resources through a comparative genomic approach, we identified substantial differences in

genomic composition between the two closely related taxa, including strong evidence for whole genome duplication (WGD) and transposable element proliferation within the introns of *Hoplisoma.* The origin and timing of the WGD suggest *Hoplisoma* is a recent autopolyploid occurring only 4.86 mya. This is the youngest autopolyploid event identified within teleosts to date.

In chapter 3, we use the *C. fulleri* reference genome in conjunction with low coverage resequencing data to explore the nature of polymorphic mimicry. *Corydoras fulleri* exists in three morphs, which are found in sympatry. The morphs form a probable Müllerian mimicry ring with four additional (non-polymorphic) sympatric Corydoradinae *spp*, which resemble two of the polymorphic patterns found in *C. fulleri*. By utilising the variation within *C. fulleri*, we identify regions characterised by elevated genetic differentiation across the genome associated with the presence or absence of a melanic blotch on the flank, an essential characteristic allowing *C. fulleri* to share a resemblance with its co-mimics. The regions contain genes associated with melanin dispersion and density, indicating known pigment pathways are being targeted to produce phenotypic variation in this trait. The presence of these genes and associated loci across the genome indicates this trait is underpinned by a more complex architecture than those previously identified in other mimetic systems, highlighting the value of understanding the genetics of mimicry across a diverse spectrum of mimetic taxa.

In chapter 4, we investigated the interaction between mimetic Corydoradinae taxa by exploring the relative toxicity of common co-mimics *Hoplisoma* and *Corydoras* species. To do this we quantified the differences in potency between co-mimics using a brine shrimp cytotoxicity assay. Whilst we found a significant increase in mortality associated with crude venom compared to control muscle extract, we identified no significant difference in venom potency between the genera suggesting they are traditional Müllerian mimics. We then used RNA sequencing to identify candidate venom genes and venom housekeeping genes in the Corydoradinae, which allowed a comparison between the *Corydoras sp.* and *Hoplisoma sp.* venom gene expression. We did not identify large logfold changes between the two genera in venom gene expression which supports the venom potency assay. This further indicates Corydoradinae co-mimics form mutualistic Müllerian interactions.

In chapter 5, we explored the genetics underpinning a rare case of sex-limited Müllerian mimicry in *C. simulatus*, where males have a mimetic phenotype, sharing a resemblance with *H. metae* whilst females have a cryptic phenotype. We used a multi-omics approach, utilising both low coverage whole genome resequencing and RNA sequencing to identify four genes which are differentially expressed in differently pigmented regions within females as well as between males and females, representing novel pigment genes. Additionally, we identify two genes related to sex differentiation representing candidates for sex determination in the Corydoradinae. This study lays the

groundwork for the first vertebrate system in which to investigating how sexually antagonistic selection influences contrasting anti-predation strategies at the genetic level.

The work presented in this thesis expands our current understanding of the genetics of mimicry by exploring the molecular basis of both unpalatability and warning colouration in the Corydoradinae. As such, we provide much needed taxonomic breadth to the field.

Acknowledgements

I have often wondered what it would be like to finish this work. Now that I am here, I am just overwhelmed with pride and gratitude to all of those that accompanied me along the way. I can whole heartedly say that I have had a fantastic time doing this degree, from the research to the people to the place. Firstly, I want to thank Martin, for all the advice, guidance and patience over the past 4 years. I feel so lucky to have worked on a project with the freedom to explore my own research interests, allowing me to become a more independent researcher. I also want to thank Simone for the advice during my annual meetings, which often put me at ease and changed my work for the better. Additionally, I would like to thank Chris for his comradery and advice during my first few years in the Taylor lab. Similarly, I'd like to thank Ellen. I wish I had added you to my supervisory team earlier as our Friday meetings became invaluable, really helping me to gain confidence in my work and ability.

I'd also like to thank my fellow students (and postdocs) in Bio, specifically Claire, Mabel, Sarah, Johnny and George, who made the oddities of doing a PhD so much more enjoyable. I'd also like to thank Ian Fuller, who is a font of wisdom on all things Corydoradinae.

Thank you to all my friends, within Norwich and further afield, including Izzy, Rory, Dan, Charlotte, Hannah, Lyra, Jenny, Sophie, Joe, Johnny, Cat and Dee. I hope to see more of you all now that I have gotten over this hurdle. I'd like to extend a specific thank you to Judit. Meeting you was one of the best things to come out of this experience.

To Harry, I really don't have the words to describe how grateful I am to have met you. I'd like to say that in the last few months you've been a rock, but you have since day one. You are a source of laughter, wisdom and comfort. I hope to be only half the crew you have been for me when your turn to submit comes.

To my family, Mum, Dad and Zoe. I am so lucky to have you as my support system. You have always supported me, no matter how weird or niche my life decisions have turned out to be. This thesis is as much your achievement as it is mine. I really could not have gotten this far without you in my corner and ultimately would like to dedicate this work to you.

1 General Introduction

1.1 Background

Animal colouration is a charismatic phenotype with many diverse ecological functions and has subsequently, captured the interest of evolutionary biologists since the inception of the field (Endler, 1983; Hoekstra, 2006; Merrill *et al.*, 2015; Chen *et al.*, 2019). Colouration is also amenable and measurable, providing an excellent framework for investigating the link between phenotype and genotype (Hoekstra, 2006). Alfred Wallace, considered the father of animal colouration, was the first to believe colour pattern could serve a function for the bearer and consider colouration within the context of natural selection (Wallace, 1877; Caro, 2017). Wallace categorised the functions of animal colouration into 'sexual', 'typical', 'protective' and 'warning'. Sexual colours include those that evolve to gain mating opportunities, such as ornamentation and signals of fitness or aggression. For example, female guppies (*Poecilia reticulata*) prefer brightly coloured males (Endler, 1983), and in the West African cichlid *Pelvicachromis taeniatus,* males display bright yellow colouration to signal aggression to their competitors (John *et al.*, 2021). Conversely, typical colouration has subsequently been delimited to cover drivers of colour pattern evolution not known to early naturalists, including anti-parasite signals and intraspecific signalling (Caro and Allen, 2017). It is, however, the latter of Wallace's categories which are the topic of this thesis, warning colouration and protection, and will be expanded on in more detail here. Protective colouration strategies are employed by prey taxa to evade predation, e.g. crypsis and masquerade. Warning colouration strategies involve adopting conspicuous colouration to send a signal to predators, e.g. warning of poison, venom or weaponry.

A taxon's appearance plays a vital role in predator evasion, with predator-prey interaction being the focus of many studies of natural selection (Endler, 1978, 1983; Linnen *et al.*, 2013; Ruxton *et al.*, 2018). An encounter with a predator is followed by a sequence of events namely detection, recognition, and subjugation potentially followed by consumption (Ruxton *et al.*, 2018). Predator evasion strategies can be categorised through the stage of the predator-prey encounter at which they occur. Crypsis strategies aim to prevent predators from detecting prey entirely (Ruxton *et al.*, 2018). One type of crypsis is known as background matching, where taxa seek to blend in with their surroundings by matching the colour, lightness or pattern of its background (Stevens and Merilaita, 2011). In addition to background matching, cryptic prey will likely have disruptive elements to their colouration, which can help mask the outline of an individual (Ruxton *et al.*, 2018). Masquerade is another antipredation strategy where taxa resemble an object in their environment in order to deceive a predator (Font, 2019). A key distinction between masquerade and crypsis is that

masquerade targets the recognition stage of the predation encounter whereas crypsis focuses on failed detection (Ruxton *et al.*, 2018). Similarly to masquerade, mimicry is not associated with failed detection but is a signal of an organism's unpalatability (Ruxton *et al.*, 2018; Font, 2019).

A mimetic relationship is a resemblance in appearance, or behaviour, between a mimic and its model and includes three parties (the Mimicry Trinity): a model, a mimic and a receiver (Font, 2019). One species may act as both the model and receiver. For example, in cuckoo egg mimicry the host bird is both imitated (model) and duped (receiver) (Font, 2019; Anderson and de Jager, 2020). Cuckoo brood parasitism is an example of 'aggressive mimicry', where the mimic aims to gain access to a resource (Ruxton *et al.*, 2018). Another example of this is in the peacock blenny, where smaller males mimic females in order to parasitize fertilizations (Gonçalves *et al.*, 2003). Alternatively, 'protective mimicry' occurs when a resemblance is utilised to evade predation and includes both Batesian and Müllerian mimicry as well as masquerade (Ruxton *et al.*, 2018; Caro and Ruxton, 2019). Here, I focus on Batesian and Müllerian mimicry and will refer to these as protective mimicries. Protective mimetic colouration involves conspicuous colour patterns which function as an advertisement of unpalatability, regardless of the honesty of this signal (Ruxton *et al.*, 2018; Caro and Ruxton, 2019). Conspicuous colouration is thought to both aid predator learning, and reduce cheating (Lindström *et al.*, 1999; Sherratt and Beatty, 2003; Ruxton *et al.*, 2018). Empirical examples of protective mimicry have been observed in a range of taxa including fish, snakes, frogs, birds and various invertebrate taxa, such as insects, molluscs and arthropods (Symula, Schulte and Summers, 2003; Marek and Bond, 2009; Wright, 2011; Wilson *et al.*, 2012, 2018; Motyka, Kampova and Bocak, 2018; Garg *et al.*, 2019; Kikuchi *et al.*, 2020; Layton, Carvajal and Wilson, 2020; Motyka *et al.*, 2020, 2021).

In 1862, William Bates noted some palatable Amazonian butterflies species resembled the bright warning colouration of non-palatable species, and suggested this resemblance allowed the palatable species to evade protection (now known as Batesian mimicry) (Bates, 1862). Bates further noted some unpalatable species also shared a resemblance (Ruxton *et al.*, 2018). However, it was Johannes Friedrich Müller who proposed this phenomenon was the result of selection on unpalatable prey converging in appearance to benefit from the shared cost of predator learning, now known as Müllerian mimicry (Müller, 1878). Müllerian mimetic interactions are considered mutualistic, benefitting both co-mimics (Alexandrou *et al.*, 2011; Speed, 2014; Ruxton *et al.*, 2018). Empirical evidence supporting Müllerian mimicry is strong, including studies in *Heliconius* butterflies (now a model system for the study of mimicry) where in *Heliconius erato*, geographically polymorphic races were reciprocally transferred across a hybrid zone (Mallet and Barton, 1989). Fewer of the novel morph were recaptured, and of those which were recaptured, many carried beak marks suggesting

that predators had learned to recognise local patterns (Mallet and Barton, 1989; Langham, 2004). Additionally, there is evidence from recapture rates of butterflies to suggest there is strong selection for mimetic phenotypes (Mallet and Barton, 1989; Kapan, 2001). Batesian mimicry, conversely, is considered to be parasitic as the relationship between the unpalatable taxa (model) and its imitator (mimic) is antagonistic resulting in a co-evolutionary arms race between or among taxa (Nur, 1970). This is due to the mimic eroding the protection afforded to the model due to the dishonest signalling of the former (Balogh, Gamberale-Stille and Leimar, 2008). There are three 'laws' of Batesian mimicry, outlined by Alfred Wallace 1) resemblances are limited to those groups (models) which are abundant and 2) are defended 3) the species that mimic these models must be less abundant (Wallace, 1877; Caro, 2017). These rules generally hold true, and it is this final rule that outlines how Batesian mimetic interactions are maintained over time.

In reality, the difference between Müllerian and Batesian mimicry is less clear cut. Under Müller's original theory, all co-mimics were assumed to be equally unpalatable (Müller, 1878). If one co-mimic is less defended than the other, the relationship could shift from a mutualistic interaction to a parasitic one, known as quasi-Batesian mimicry (Speed, 1993). Unpalatability is a relative concept subject to the nature of the predator and prey. Prey that is inedible to one predator may be edible to another, rendering an interaction both Batesian and Müllerian under different scenarios (Ruxton *et al.*, 2018). The frequency and density of the less defended prey also has an impact on whether the interaction is parasitic or mutualistic in nature, with rare quasi-Batesian mimics imposing little evolutionary pressure on their co-mimic/models (Rowland *et al.*, 2010). Furthermore, the availability of palatable prey and predator starvation has also been shown to impact when a co-mimic is Müllerian or Batesian. For example, when the fat stores of starlings had been experimentally reduced, they were more likely to accept distinctly coloured quinine injected (distasteful) mealworms (Barnett, Bateson and Rowe, 2007). Currently, there are no empirical examples of quasi-Batesian mimicry, although diversity in toxic defences between congeners have been demonstrated in both *Heliconius* butterfly and *Ranitomeya* poison dart frog systems (Saporito *et al.*, 2010; Stuckert *et al.*, 2014; Marples, Speed and Thomas, 2018; Jeckel *et al.*, 2019). However, diversity in defences may not equate to differences in unpalatability and proof of differential selection between sympatric taxa is required to confirm a hypothesis of quasi-Batesian mimicry (Stuckert *et al.*, 2014). However, this does highlight a need for a clear framework to distinguish between quasi-Batesian mimicry and traditional Müllerian mimicry.

1.1.1 Evolution of Mimicry

The evolution of a novel mimetic resemblance among taxa involves traversing a fitness valley as a taxon will need to evolve away from an existing beneficial phenotype to a new phenotype (Merrill *et al.*, 2015). For example, the acquisition of a mimetic phenotype can involve switching from cryptic to the conspicuous aposematic phenotypes associated with protective mimicries (Wilson *et al.*, 2012; Ruxton *et al.*, 2018). In *Papilio polytes*, cryptic males are hypothesised to be representative of the ancestral form, from which the mimetic female phenotype diverged (Kunte *et al.*, 2014). There are two major theoretical mechanisms though which conspicuous colouration could evolve – the two step hypothesis (Nicholson, 1927; Sherratt, 2008; Merrill *et al.*, 2015) and the gradual evolution hypothesis (Fisher, 1927, 1931). In the two-step theory, crossing a fitness valley is thought to be achieved via a single mutation of major effect, followed by many mutations of smaller effect further refining the mimetic resemblance (Sherratt, 2008; Merrill *et al.*, 2015). The *Heliconius* radiation may provide an example of this phenomenon, where most of the wing patterning is controlled by only a few 'toolkit genetic loci', perhaps representing the large-effect mutations (Merrill *et al.*, 2015). Additionally, in *Heliconius erato*, several modifier loci of moderate effect have been identified, representing the second step in the two-step theory (Papa *et al.*, 2013). As mimetic phenotypes are multi-dimensional, it is not clear whether a mutation in a single trait is sufficient to overcome an adaptive valley. Franks and Sheratt (2007) found whilst mimicry in a single aspect of the phenotype is not sufficient to protect prey, it can lay the foundation upon which gradual evolution of multicomponent mimicry can act. In contrast, the gradual evolution hypothesis first suggested by Fisher in 1927 has received less support. Although prey with similar starting resemble could converge through small mutations (similar to the second step of the two-step hypothesis), prey with bigger phenotypic differences were considered less likely to converge through many small mutations (Ruxton *et al.*, 2018). However, Balogh and Leimar (2005) demonstrated that convergence of phenotype could occur through more gradual processes if a range of predators and with different ranges of generalisation were involved, supporting the gradual evolution hypothesis.

1.1.1.1 Advergence vs convergence

A mimetic resemblance can evolve through convergence- all taxa involved shift their phenotypes to a single shared phenotype, or advergence where one or more taxa change their phenotype to resemble a model (Ruxton *et al.*, 2018). In Batesian mimicry, the evolution of phenotypic resemblance requires the mimic to traverse a fitness valley to adverge on the model phenotype i.e. the model remains the same and the mimic changes phenotype. Such mimicry is believed to evolve when a species would benefit greatly from being visually conspicuous (and resembling a conspicuous unpalatable model) but cannot afford the cost of developing defensive mechanisms (Ruxton *et al.*, 2018). Under Müllerian

mimicry, identifying which of the taxa has crossed a valley is harder to decipher. In his original theory, Müller argued that both co-mimics would evolve a to share a colour pattern to look alike, known as convergence (Ruxton *et al.*, 2018). However, empirical evidence in the Müllerian mimic Appalachian millipedes, velvet ants (*Dasymutilla*), poison dart frogs (*Ranitomeya sp*) (but see Chouteau *et al.* (2011)) and in South Asian bumble bees (*Bombus breviceps* and *Bombus trifasciatus*) supports advergence (Symula, Schulte and Summers, 2001; Marek and Bond, 2009; Wilson *et al.*, 2012; Ruxton *et al.*, 2018; Muell *et al.*, 2022; Cui *et al.*, 2024). The prevalence of advergent mechanisms may be due to differences in the mutational space and intensity of selection (Balogh, Gamberale-Stille and Leimar, 2008; Balogh *et al.*, 2010; Ruxton *et al.*, 2018). It is also possible Müllerian mimicry involves an initial stage of advergence followed by convergence resulting in a shared resemblance (Ruxton *et al.*, 2018). Additionally, there is evidence Müllerian mimicry can result from conservation of the ancestral state. This is suggested to be the case in the *Synodontis* catfish, where the species within Lake Tanganyika have a stronger degree of resemblance than a nested sister species residing outside the lake (Wright, 2011). In this instance the origin of resemblance can be attributed to shared ancestry, but the ecological function of the resemblance is mimetic (Wright, 2011; Ruxton *et al.*, 2018). Ancestry is also thought to explain the resemblance between species of yellow-black solider beetles (*Chauliognathus*)(Machado *et al.*, 2004).

1.1.1.2 A role for introgression

The role of gene flow between co-mimics has increasingly been shown to be a key mechanism through which advergence of resemblance can cross species barriers. Adaptive introgression has been shown to produce novel mimetic phenotypes across the *Heliconius* system (Pardo-Diaz *et al.*, 2012; The Heliconius Genome Consortium, 2012; Jay *et al.*, 2018). This occurs through sharing of the mimetic colour alleles across species boundaries and has been shown for many of the large effect loci identified in the group (Pardo-Diaz *et al.*, 2012; Enciso-Romero *et al.*, 2017). For example, a region involved in wing pigment patterning (*cortex*) has been shown to be shared between *H. cydno* and *H. melpomene*, allowing both species to participate in novel mimicry rings (Enciso-Romero *et al.*, 2017). Additionally, signals of introgression are largely limited to regions of the genome associated with colour pattern across three *Heliconius* species (The Heliconius Genome Consortium, 2012). Prior work has largely focused on hybridization within a more recently diverged *Heliconius* clade, but more recent work has found signals of introgression are present across the genus, indicating the importance and prevalence of gene flow in the group (Kozak *et al.*, 2021). However, whilst introgression is clearly an important driver for the evolution of mimetic phenotypes in *Heliconius*, it is unclear how pervasive this pattern is across other mimetic communities. There is evidence for introgression in the evolution of the south variable pitohui *Pitohui uropygialis* from its co-mimic *Pitohui dichrous* (Garg *et al.*, 2019).

In this system, SNPs within introgressed regions are linked to genes related to melanin synthesis (Garg *et al.*, 2019). However, no follow up study has confirmed the link between these genes to mimicry in the system.

The role of introgression in *Heliconius* has additionally served as evidence for the shifting balance theory (Wright, 1932; Merrill *et al.*, 2015). Introgression is thought to occur in the latter stages of the shifting balance theory, which explains how populations can shift between adaptive peaks due to an interplay between evolutionary stochasticity and natural selection (Wright, 1932; Mallet and Joron, 1999; Mallet, 2010). Shifting balance theory states that genetic drift enables a locally adapted subpopulation to navigate an adaptive valley, reaching the base of a higher adaptive peak (Mallet, 2010). Natural selection then drives the subpopulation up this peak. Through introgression with migrants, the sub-optimal populations are able move to the higher fitness peaks, eventually resulting in all of the populations having a higher fitness due to this process (Mallet, 2010). Within *Heliconius sp.*, the introgression of wing patterning traits is thought to represent phase three of this model whilst the presence of hybrid zones is reminiscent of phase one (Mallet, 2010). However, the shifting balance model is controversial having garnered a lot of opposition, mainly a lack of empirical evidence of each phase occurring within a single population and epistatic interactions resulting in a complex landscape of adaptive peaks (Coyne, Barton and Turelli, 2000; Weinreich and Chao, 2005).

Shifting balance theory was introduced as one explanation for paradoxical observations of polymorphic Müllerian mimics. Under Müllers original model, all taxa should converge on the same phenotype to maximise shared predator learning (Müller, 1878; Ruxton *et al.*, 2018). However, there is diversity of mimicry phenotypes among microhabitats, regionally, and even locally as stable polymorphisms (Mallet and Joron, 1999). Phenotypic diversity of mimicry rings within same locality is observed with the Ithomiine butterflies and is thought to be the result of microhabitat segregation of models (Gompert, Willmott and Elias, 2011; Ruxton *et al.*, 2018). Additionally, *in-silico* evidence suggests the degree of habitat specialization can influence the diversity of mimicry rings present at a single site (Birskis-Barros, Freitas and Guimarães, 2021). Regional polymorphism can be maintained due to clinal changes in extrinsic environmental characteristics, intraspecific density-dependent regulatory mechanisms, or predator-driven frequency dependent selection (Svensson and Abbott, 2005; McLean and Stuart-Fox, 2014; Holmes, Grundler and Davis Rabosky, 2017; Doktorovová *et al.*, 2019).

1.1.1.3 Frequency dependent selection

Predator-driven negative frequency dependent selection occurs when common morphs experience a greater predation risk than their rarer counterparts (Holmes, Grundler and Davis Rabosky, 2017). This

type of selection is generated through the processes of neophobia and dietary conservatism, which lower the willingness of the predator to attack encountered prey, particularly novel prey (Marples and Mappes, 2011; Huang *et al.*, 2012; Holmes, Grundler and Davis Rabosky, 2017). Neophobia is an immediate response of fear of physical contact with a prey item, whereas dietary conservatism is a stable long-term process in which predators refuse to accept novel prey as part of their diet (Marples and Mappes, 2011). Even when only a proportion of the predator population display dietary conservatism this process can maintain a polymorphic population, as has been demonstrated using the great tits (*Parus major*) and three-spined sticklebacks (*Gasterosteus aculeatus*) as model predators (Thomas *et al.*, 2010; Marples and Mappes, 2011). Conversely, polymorphic mimicry is most paradoxical in systems under positive frequency dependent selection, when rare morphs are disproportionately consumed by predators, known as positive frequency dependent selection (Holmes, Grundler and Davis Rabosky, 2017). However, changes in the scale of predation (regional or local) have been shown to result in the maintenance of polymorphism in systems experiencing both positive and negative frequency dependent selection (Holmes, Grundler and Davis Rabosky, 2017).

1.2 Genetics of the Mimetic Phenotype

The genetic architecture of a trait includes the number and characteristics of the genetic elements, their effect size and interactions, which can all influence the evolution of an adaptive trait (Huber *et al.*, 2015). Knowledge of the molecular mechanisms underpinning the mimetic phenotype are vital to elucidate the evolutionary history of these systems including the role of advergence versus convergence, the role of introgression, and the maintenance of polymorphisms, including those that are locally stable. The mimetic phenotype can be divided into two components: unpalatability and the advertisement of this unpalatability, through warning colouration.

1.2.1 Warning Colouration

The evolution of warning colouration relating to mimicry has been most extensively studied in the *Heliconius* and *Papilio* systems. Within these systems there is a large degree of genetic conservatism (reuse of the same suite of genes), and most phenotypic variation is controlled by few regions of large effect across the genome (Kunte *et al.*, 2014; Jiggins, 2016). For example, red and orange patterning in many *Heliconius* sp. involves the same genomic region, known as the *optix* locus (Jiggins, 2016). This region contains both the *optix* gene, which encodes a transcription factor, and cis-regulatory elements which modify the gene's expression, leading to phenotypic variation (Reed *et al.*, 2011). Despite the shape of the red patterning differing between species, the *optix* region has been shown to be homologous between both *H. erato* and *H. melpomene* (Kronforst, Kapan and Gilbert, 2006; Jiggins, 2016). This region consists of several tightly linked loci rather than a single locus with diverse

phenotypic effects (Jiggins, 2016). Evidence for this comes from hybrid regions of polymorphic species, where partial phenotypes are inherited as well as established morphs displaying recombinant phenotypes (Mallet, 1989; Wallbank *et al.*, 2016). The development of shared mimetic colour patterns in *Heliconius* has been driven by both mutations and introgression. For example, *H. erato* and *H. melpomene* independently evolved different regulatory mutations around the *optix* gene, allowing them to participate in a Müllerian mimicry ring. In contrast, *H. timareta* has adverged on this pattern through the introgression of regulatory alleles from *H. melpomene* (Pardo-Diaz *et al.*, 2012; The Heliconius Genome Consortium, 2012). Another transcription factor, *aristaless1* has been shown to control the switch between yellow or white wing patterning in *H. cydno* (Westerman *et al.*, 2018). It has been hypothesised that features of these major effect loci (e.g., tier position in regulatory networks) make them ideal targets for repeated natural selection (Merrill *et al.*, 2015). For example, *optix* is positioned downstream of numerous genes whose spatial expression orientates wing development (Merrill *et al.*, 2015). Increasing evidence suggests the phenotypic variation seen across *Heliconius* mimicry rings is attributed to changes in the spatial and temporal expression of these major loci and the interactions between them (McMillan *et al.*, 2020). The *cortex* gene is thought to act as a master switch to determine the colour of scale cells, occurring earlier in development with *WntA,* and interacts with either *optix* or *aristaless1* to determine scale cell fate (McMillan *et al.*, 2020; Livraghi *et al.*, 2021).

Recombination of alleles associated with pigmentation in *Heliconius* can result in intermediate forms which have reduced fitness due to deviation from the optimum phenotype (Merrill *et al.*, 2015). This may occur when predation pressure is relaxed or in narrow hybrid zones where interbreeding monomorphic races meet, but could incur high costs for locally polymorphic species, such as *H. numata.* A chromosomal inversion within *H. numata* has resulted in very limited recombination of the *cortex* gene and its regulatory elements (Joron *et al.*, 2011; Saenko *et al.*, 2019). Due to this tight linkage, the region is known as a supergene, acting as a switch between the *H. numata* morphs and is thought to have evolved through introgression with the ancestor of *H. paradlinus* and *H. elevates,* and *H. numata* (Joron *et al.*, 2011; Jay *et al.*, 2018). Supergenes have also been shown to control polymorphism in Batesian Lepidoptera species. In *Papilio sp.,* males display a non-mimetic wing pattern whereas females are polymorphic displaying either a male-like pattern or one of several mimetic patterns resembling the unpalatable taxa from the genus *Pachliopta* (Kunte *et al.*, 2014). The gene *doublesex* has subsequently been identified as a supergene controlling both the female polymorphism and additional inter-sex variation (Kunte *et al.*, 2014; Komata, Lin and Fujiwara, 2022). In the *Papilio* system, the supergene architecture is characterised by a chromosomal inversion in some

species but not others, suggesting there are additional mechanisms allowing for reduced recombination (Iijima *et al.*, 2018).

Lepidopteran colour pattern mimicry is controlled by few regions of large effect, but it is not known how universal this architecture is across mimetic species. In the stone fly, *Zelandoperla fenestrate* (a harmless Batesian mimic), the insect melanin locus *ebony* is responsible for the majority of the variation in colour pattern (Foster *et al.*, 2023). Work on aposematism in the poison dart frogs (Dendrobatidae) is investigating whether the 'few genes of large effect' pattern holds in a vertebrate system. The polymorphic *Ranitomeya imitator* has undergone diversification in colour pattern, adverging on local *Ranitomeya* models (Symula, Schulte and Summers, 2001; Twomey *et al.*, 2016; Muell *et al.*, 2022). Comparisons between morphs of *R. imitator*, as well as its model species, *R. fantastica* and *R. variabilis* across different developmental stages, highlighted candidate genes associated with differences in pigment variation including those implicated in carotenoid metabolism or xanthophore production, the synthesis of pteridines, as well as genes related to melanophore development and melanin synthesis (Stuckert *et al.*, 2024). However, few genes were shared between intraspecific comparisons, indicating that convergent colour patterns likely evolved through species-specific mechanisms (Stuckert *et al.*, 2014). Similarly, the red/yellow polymorphism of *R. sirensis* is due to differences in the expression of the carotenoid ketolase (CYP3A80) and downregulation of BCO2 resulting in an accumulation of ketocarotenoids (Twomey *et al.*, 2020). The utilisation of known pathways in the *Ranitomeya* system indicates convergence at a pathway level, but not at a genic level. Additionally, whilst a large number of candidates have been identified through these and other expression studies, the lack of information about their location in the genome limits inferences about genetic architecture in this system or a role for introgression (Stuckert *et al.*, 2023, 2024; Rubio *et al.*, 2024).

1.2.1.1 Pigmentation in teleosts

Early studies exploring the genetic underpinnings of pigmentation in teleosts focused on exploring genetic linkage in zebrafish (*Danio rerio*) mutants (Streisinger *et al.*, 1986). However, fish pigmentation has since become an important trait for evolutionary geneticists, due to the diversity seen within and among species and the ecological interactions that shape this variability (Endler, 1978, 1983; Seehausen, Mayhew and Van Alphen, 1999; Santos *et al.*, 2014; Salis *et al.*, 2019; Parichy, 2021). Unlike mammals, which have only melanocytes, teleosts have six classes of pigment cell types (chromatophores), resulting in the diverse colour patterns observed across the group, making teleosts a valuable system to understand the full range of pigmentation in vertebrates (Irion and Nüsslein-Volhard, 2019; Parichy, 2021). These cell types include the include black melanophores, silver

iridophores, yellow orange xanthophores, red erythrophores, white leucophores and blue cyanophores. These different chromophores are distributed on the scales as well as in layers between the epidermis and muscle, with the position, shape and diversity of the chromatophores resulting in the overall colour (Salis *et al.*, 2019; Parichy, 2021).

The pigmentation of a teleost can be made up of multiple elements, each of which may be controlled by a few genes of large effect or be controlled by many genes of smaller effect sizes (Elkin *et al.*, 2023; Tapanes and Rennison, 2024). The Siamese fighting fish (*Betta splendens*) represents the former, where colouration and pattern is controlled by several loci of large effect spread across the genome (Zhang *et al.*, 2022). Additionally, in the three-spined stickleback, changes in *kitlig* expression were responsible for > 56% of variance in gill melanism (Miller *et al.*, 2007). Conversely, the majority of pigmentation traits explored in the sticklebacks have been due to small or moderate effect loci (Greenwood *et al.*, 2011; Greenwood, Cech and Peichel, 2012; Tapanes and Rennison, 2024). Additionally, in species where colouration is sex-limited, potentially due to sexual selection or differential natural selection pressures, pigmentation can have a more complex genetic architecture. In sexually dimorphic cichlid fish from Lake Malawi, the Orange Blotch (OB) locus is associated with a novel female sex-determiner, such that the OB phenotype is mainly seen in females (Roberts, Ser and Kocher, 2009). However, not all sex-specifics traits are linked to sex determining regions. In the Trinidadian guppy (*Poecilia reticulata*), male colouration was associated with genetic differentiation of an autosomal region, potentially due to epistatic interactions between the autosomal regions and polymorphisms on the male sex chromosome (Paris *et al.*, 2022).

In addition to examining the genetic architecture and linkage of a trait when drawing parallels with the genetics of lepidopteran mimics, it is also essential to consider the roles for convergent molecular evolution and genetic conservation. Across African Rift Valley Lakes, cichlid species with a striped phenotype have a lower expression of the gene *agrp2* than their non-striped counterparts (Kratochwil *et al.*, 2018). Phylogenetic analysis of the enhancer region of this gene revealed changes in expression occurred independently between Lakes and likely evolved from standing genetic variation (Kratochwil *et al.*, 2018; Urban *et al.*, 2021). Parallel genetic evolution has also occurred in some cave adapted teleosts. Albinism in the Mexican cave fish, *Astyanax mexicanus*, is underpinned by deletions in exons of the *oca2* gene resulting in a loss of pigment (Protas *et al.*, 2005; Klaassen *et al.*, 2018). This has occurred independently within two populations of *A. mexicanus*, in different exons. Similarly, a deletion in *oca2* is responsible for the amelanistic morph of the Lake Malawi cichlid *Melanochromis auratus*, where the deletion occurs in the second exon (Kratochwil, Urban and Meyer, 2019). In another species of cichlid, *Lamprologus lethops*, found in low light habitats of the lower Congo river

also lacks pigmentation, originating from a mutation disrupting the start codon of *oca2* (Aardema, Stiassny and Elizabeth Alter, 2020). Within this gene there is an accumulation of additional non-synonymous mutations compared to sister taxon, suggesting that selection is relaxed and no longer maintaining the gene function (Aardema, Stiassny and Elizabeth Alter, 2020). This evidence suggests that there is some genetic convergence in teleost pigmentation.

1.2.2 Unpalatability

Unpalatability, which underpins the evolution of mimicry, refers to a taxon being chemically defended and thus being noxious, either due to toxicity (ingestion) or envenomation (injection) (Marples, Speed and Thomas, 2018). Chemical defences can be sequestered from the environment or generated autonomously. The *Heliconius* butterflies both sequester toxic cyanogenic glucosides from *Passiflora* host plants during the larval stage and synthesize cyanogenic glucoside *de novo* (Cardoso and Gilbert, 2013; Arias *et al.*, 2016). However, the genetic basis for biosynthesis in this group is yet to be explored. Similarly, the poison dart frogs sequester alkaloids from an arthropod diet, which are stored in the skin granular glands before secretion (Saporito *et al.*, 2009; Stuckert *et al.*, 2014). Adaptations for resistance to the sequestered alkaloids and movement of toxins from the gut to the skin are required this group (Tarvin *et al.*, 2017; Caty *et al.*, 2019). Commonly, resistance from self-intoxication is encoded via changes in amino acids to toxin-binding sites (Tarvin *et al.*, 2017). Additionally, genes associated with sodium and ion pumps have been shown to be differentially expressed between wild caught aposematic *Oophaga sylvatica* poison arrow frogs and laboratory reared adults on alkaloid-free diet, representing candidate genes involved in self-resistance to alkaloids (Caty *et al.*, 2019).

In contrast to sequestered toxins, endogenous toxins are entirely synthesised within an organism's body. One example of these is venom, which is defined as a secretion containing molecules that disrupt normal physiological or biochemical functioning, produced in a specialized gland and delivered to a target individual through the infliction of a wound (Casewell *et al.*, 2013). Over 200,000 taxa are considered venomous, making it a widespread adaptive trait, having evolved independently over 100 times (Schendel *et al.*, 2019; Zancolli *et al.*, 2022). Venom serves multiple ecological functions, including in predation, defence and intraspecific conflict (Casewell *et al.*, 2013). Several mechanisms have been linked to the evolution of venom genes. For example, gene duplication via neofunctionalization, in which one gene copy retains its function but the other is released from selection and is able to accumulate mutations, can result in novel functions (Ohno, 1970). In venom gene evolution, an initial duplication event is often followed by subsequent duplications and venom gland limited expression, increasing the complexity of venom composition (Kordis and Gubensek, 2000; Casewell *et al.*, 2019). However, not all venom genes evolve via gene duplication, with evidence

for co-option of single-copy genes in parasitoid wasps and in the platypus (Wong and Belov, 2012; Wong *et al.*, 2012; Martinson *et al.*, 2017). Additionally, alternative splicing has been implicated in increasing venom complexity and diversity (Casewell *et al.*, 2013). In the common house spider (*Parasteatoda tepidariorum*) venom genes are characterised by increased transcript number (Haney *et al.*, 2019). The evolution of venom genes is primarily understood in the context of feeding, however, aposematic taxa employ venom as a form of defence, although the mechanisms of evolution are likely to be similar. For example, evidence from Hymenoptera demonstrated that venom evolved originally from the co-option of single copy genes, but diversity was subsequently increased through gene duplication (Koludarov *et al.*, 2023).

Over 2,900 fish (cartilaginous and ray-finned) fish species are thought to be venomous (Smith *et al.*, 2016). Despite this diversity, piscine systems are understudied, both with regard to venom evolution and venom composition (Smith and Wheeler, 2006; Harris and Jenner, 2019). Piscine venom is thought to have independently evolved 19 times, with the majority of taxa utilising venom for defence (Smith and Wheeler, 2006; Wright, 2009; Smith *et al.*, 2016; Casewell *et al.*, 2017; Harris and Jenner, 2019). Many fish species utilise spines for defence from predation, and of venomous species, the majority utilise spines as a delivery apparatus (but see Casewell *et al.* (2017))(Smith *et al.*, 2016). Additionally, the spine morphology of venomous fish has converged, often being characterized by anterolateral grooves and orange-yellow venomous tissue (Smith *et al.*, 2016). As such, the evolution of spine and spine morphology is intrinsically linked to the evolution of venom systems in piscine species (Harris and Jenner, 2019). Venom glands in fish are thought to have evolved from epidermal cells with the capacity to produce antiparasitic toxins near defensive spines (Cameron and Endean, 1973; Harris and Jenner, 2019). The skin mucus of fish contains ichthyocrinotoxins, with known antimicrobial and antiparasitic properties, which have been shown to play a role in predator avoidance (Ángeles Esteban, 2012; Gratzer *et al.*, 2015). Additionally, the stonustoxin present in the tropical stonefish venom (*Synanceia horrida*) was found to originate from a family of genes with immune related functions (Ellisdon *et al.*, 2015). This further indicates a role for the icthyocrinotoxins in the evolution of piscine venoms (Harris and Jenner, 2019). As such, the genetic mechanisms underpinning piscine venoms can provide further insights into the evolution of this trait. For example, Scorpaneoid venoms are thought to have evolved via gene duplication, as seen in some terrestrial systems (Chuang and Shiao, 2014; Ellisdon *et al.*, 2015). The current research is largely restricted to the Scorpeaniformes and Trachiniformes, focusing on a few genes or with the aim of identifying of novel proteins for bioprospecting (but see Casewell *et al.* (2017)) (Chuang and Shiao, 2014; Han *et al.*, 2017; Xie *et al.*, 2019; Lopes-Ferreira *et al.*, 2021). This highlights the need to better explore the evolutionary context in which piscine venoms have evolved and to understand their role in ecological interactions.

1.3 Corydoradinae catfishes

The Callichthyidae, comprising of the subfamilies Corydoradinae and Callichthyinae, is the third largest Siluriformes (catfish) family, with over 223 described members (Dias *et al.*, 2024; Fricke, Eschmeyer and Van der Laan, 2024). The majority of these species are within the subfamily Corydoradinae with 208 species, although many species remain undescribed (Fuller and Evers, 2005; Dias *et al.*, 2024). Fossil evidence suggests the group were already established in the mid/late Pleistocene (Bogan *et al.*, 2020). The Corydoradinae occupy a range of habitats, from small oxygen rich creeks to big rivers and flooded areas or swampy habitats where dissolved oxygen is virtually absent (Reis, 1998). Members of the family are characterised by longitudinal dermal plates as well as two or three (basal) barbels either side of the mouth (Nijssen, 1970). These plates are known as scutes and cover 46% of the body depth, which is higher than any other armoured fish species (Lowe, Summers, Walter *et al.* 2021).

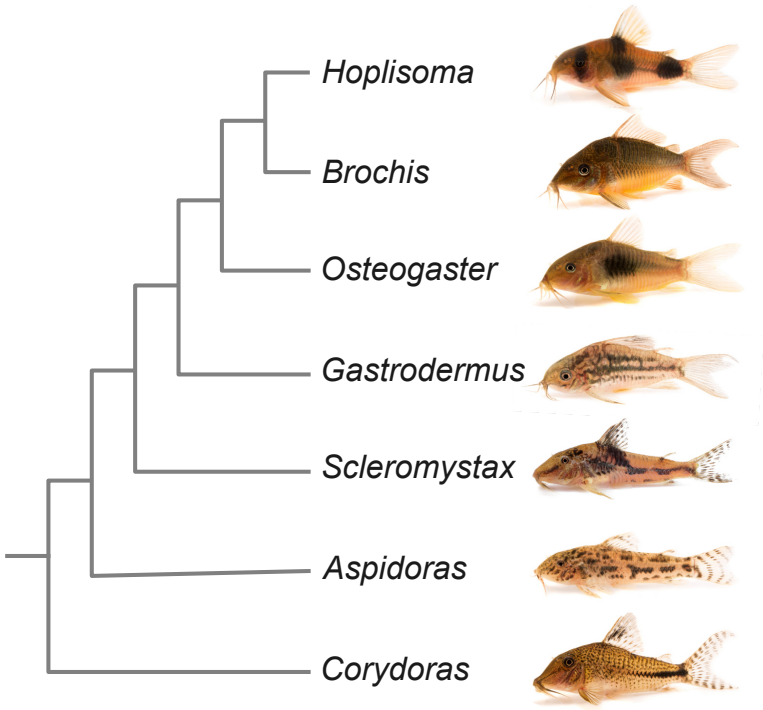


Figure 1.1 The evolutionary relatonship between Corydoradinae genera. Figure produced according to Dias *et al.* (2024).

A mtDNA phylogeny of the Corydoradinae revealed nine monophyletic lineages (Marburger *et al.*, 2018). However, there is some discordance between the placement of lineage 6 between phylogenies constructed with mtDNA and nuclear markers, where nuclear markers resolve lineage 6 and lineage 9 to be monophyletic (Marburger *et al.*, 2018). A more recent phylogeny using ultra conserved elements further corroborated monophyly of lineages 6 and 9 whilst also suggesting the monophyly of lineage 4 and 5 , although few species were sampled from lineage 4 (Dias *et al.*, 2024). As a result of this

analysis and morphological differences, the Corydoradinae have recently been divided into 7 genera (Figure 1.1).

A wide variety of colour and patterns are seen across the Corydoradinae, which vary geographically, include features such as eye stripes, large and small spots, fine dots, circles, body stripes, lines and blotches (Sands, 1994; Alexandrou *et al.*, 2011). These features can function as both warning or cryptic colouration, depending on the overall pattern and colouring (Nijssen, 1970; Alexandrou *et al.*, 2011). Within the group, background matching (a cryptic strategy) has been observed, with pigment of *Hoplisoma sp.* being directly associated with substrate (Sands, 1994). Specimens living over darker substrates had increased pigmentation intensity than those living over light-coloured sandy substrates (Nijssen, 1970). Additionally, disruptive colouration can be seen in the group, for example, the Corydoradinae species from the Rio Miua utilising dark dorsolateral stripes and eye masks which is high contrast to the pale flank (Sands, 1994; Ruxton *et al.*, 2018). During behaviour studies, *Corydoradinae* species were shown to freeze, which is also thought to be indicative of a crypsis strategy (Sands, 1994). Alternatively, these high contrasting patterns could be representative of conspicuous warning colouration. Many members of the Corydoradinae possess conspicuous

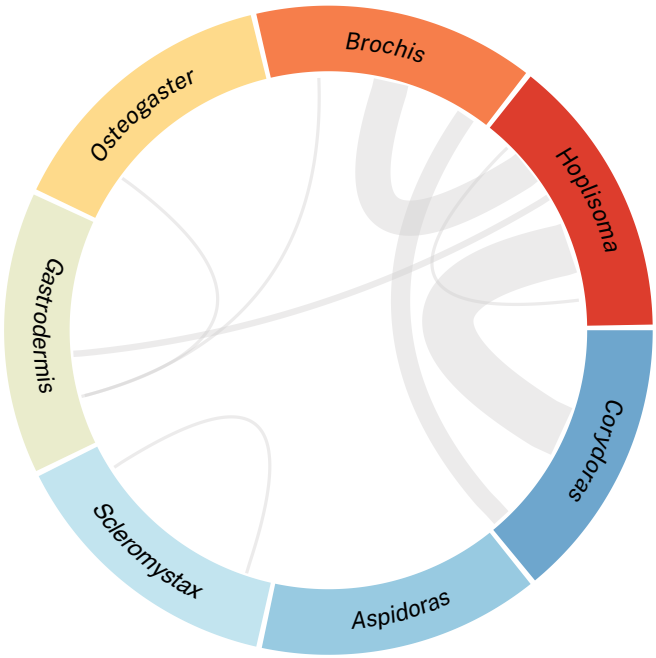


Figure 1.2 Mimetic relationships among Corydoradinae genera represented by linking arcs. The arc width represents the number of mimetic relationships between the two genera. Arcs within *Hoplisoma* and *Scleromystax* represent sympatric species from the same lineage that share colour patterns.

contrasting black and white stripes, orange and black patches as well as conspicuously coloured spines (Alexandrou *et al.*, 2011). Whilst crypsis and aposematism are viewed as opposing strategies, some colour patterns may utilise both. For example, narrow bands of contrasting colours can be visible from a short distance, but from a greater distance, they can blend, providing crypsis (Ruxton *et al.*, 2018). Therefore, the high contrast pigmentation of the Corydoradinae could be both cryptic and conspicuousness at different ranges.

The pigmentation of Corydoradinae species can vary spatially, with distantly related taxa in the same region sharing a single colour pattern, often being found in mixed shoals of up to five species (Nijssen, 1970; Alexandrou *et al.*, 2011). Co-occurring species avoid interspecific competition through fine scale trophic partitioning (Alexandrou *et al.*, 2011). As all Corydoradinae have lockable pectoral and dorsal spines with a venom gland positioned at the base, these are thought to represent Müllerian mimetic communities. Within the Corydoradinae, there are at least 27 Müllerian mimicry rings with over 60 species participating (Alexandrou *et al.*, 2011; Tencatt and Ohara, 2016; Tencatt *et al.*, 2021, 2024). Taxa participating in these mimicry rings are mostly from the genera *Corydoras*, *Brochis* and *Hoplisoma*, indicating a non-random frequency of co-occurrence between members of different Corydoradinae groups (Figure 1.2)(Alexandrou *et al.*, 2011). Despite the mimetic relationships within Corydoradinae being considered to be Müllerian there are cases of Batesian mimicry associated with the group. For example, non-toxic *Otocinclus mimulus* is thought to imitate *H. diphyes* (Axenrot and Kullander, 2003).

The Corydoradinae are one of nine Siluriformes families known to produce secretions from epidermal cells or venom apparatus, including pectoral or dorsal spines (Greven, Flasbeck and Passia, 2006; Wright, 2009). The axillary gland of *Osteogaster aeneus* contains secretory cells deriving from the epidermis, from which crinotoxins are release during a stress response (Greven, Flasbeck and Passia, 2006; Kiehl, Rieger and Greven, 2006). There is some evidence to suggest that venom in the Callichthyidae has evolved independently from other venomous Siluriformes families (Wright, 2009). However, discrepancies in the phylogenetic reconstruction of the Siluriformes has brought this into question (Smith *et al.*, 2016). Additionally, peptides of similar weights were found in both Callichthyidae and Siluroidei secretions, indicating either a shared origin or parallel evolution in venom proteins (Wright, 2011).

1.3.1 Changes in Genome Size

Early cytogenic analysis suggested the Corydoradinae had undergone drastic changes in genome size, which was supported by additional cytogenic and phylogenetic analysis (Oliveira *et al.*, 1992, 1993; Shimabukuro-Dias *et al.*, 2004; Shimabukuro-Dias, Oliveira and Foresti, 2004). More recently, the

evolution and mechanisms driving this shift were shown to be due to both whole genome duplication (WGD) and transposable element (TE) proliferation (Marburger *et al.*, 2018). Two WGD events are thought to have occurred within Corydoradinae, the first between 54 and 66 mya (base of *Aspidoras* and *Scleromystax*) and the second more recently, 20-30 mya (base of *Hoplisoma*) (Marburger *et al.*, 2018). These events occur when homologous chromosomes fail to separate during cell division resulting in the doubling of the genome (Bomblies, 2023). It has been suggested that WGD events play a role in diversification and tolerance to changing environments, partially due to the increase in genetic material and relaxed selection (Ohno, 1970; Meyer and Van de Peer, 2005; Mable, Alexandrou and Taylor, 2011; Robertson *et al.*, 2017; Bell *et al.*, 2020; Xu *et al.*, 2023). Within the *Corydoradinae,* diversity of gene number and genetic variation within Toll-like receptors is associated with reduced parasite loads in the putative polyploid *H. araguaiaensis*, highlighting that whole genome duplication could be increasing adaptive potential within the group (Bell *et al.*, 2020). Additionally, the diversification of *Hoplisoma* and its overrepresentation in Müllerian mimicry rings could be related to changes in genome size, including the WGD event (Alexandrou *et al.*, 2011)

The evolution of vertebrates is punctuated by WGD events. The 2R hypothesis states duplication played an instrumental role in the early evolution of deuterostomes, having occurred twice and predicting the vertebrate genome contains four paralogs for each proto-ortholog in cephalochordates, based on evidence from Hox clusters (Amores *et al.*, 1998; Meyer and Van de Peer, 2005). An additional WGD event is thought to has occurred early in the ray-finned fishes (Actinopterygii) prior to the divergence of most teleosts (Amores *et al.*, 1998; Meyer and Van de Peer, 2005; Kuraku and Meyer, 2009; Braasch and Postlethwait, 2012). Additional independent lineage-specific whole genome duplications have also been identified in some teleost lineages, for example within the salmonoids and cyprinids (Glasauer and Neuhauss, 2014; Ma *et al.*, 2014; Robertson *et al.*, 2017; Chen *et al.*, 2019; Xu *et al.*, 2019). The salmonoid (autopolyploid) WGD event was estimated to have occurred 88-103 mya and is suspected to have provided the genetic substrate for the evolution of migratory behaviour (Alexandrou *et al.*, 2013; Gundappa *et al.*, 2022). However, between the duplication event and the diversification of the salmonoid lineage, there is a lag of at least 40 my, suggesting the diversification of the group was not the result of WGD but in response to climate change (Glasauer and Neuhauss, 2014; Macqueen and Johnston, 2014). Recent evidence suggests that this lag is due to a delay in rediploidization and parts of the genome remaining functionally polyploid until after speciation had occurred (Robertson *et al.*, 2017; Gundappa *et al.*, 2022). Additionally, lags in diversification after whole genome duplication have been observed in other taxa, suggesting that diversification due to duplication could still occur long after the duplication event itself (Alexandrou *et al.*, 2013; Glasauer and Neuhauss, 2014).

A second driver of genome size in the Corydoradinae is transposable element proliferation (Marburger *et al.*, 2018). Transposable elements are DNA sequences with the ability to change position within a genome (Wells and Feschotte, 2020). Traditionally, transposable element insertion was viewed as deleterious but increasingly have been shown to play a role in phenotypic diversification (Schrader and Schmitz, 2019). There are various mechanisms through which transposable element expansion can drive phenotypic change. For example, insertion of a TE into the intron of a host can result in a TE exon being encompassed into a protein gene, known as exonization (Schrader and Schmitz, 2019). Additionally, proximity of a TE can affect the regulatory environment and subsequent transcription of a gene (Schrader and Schmitz, 2019). There is increasing evidence for the latter. For example, the black phenotype (*carbonairia*) of the peppered moth (*Biston betularia*) is caused by a TE insertion into first intron of the gene *cortex* (Van't Hof *et al.*, 2016). Similarly, the insertion into the first intron of the *goldentouch* gene in the Midas cichlid (*Amphilophus spp.*) results in a loss of melanin, underpinning a stable polymorphism in the species (Kratochwil *et al.*, 2022). Additionally, the insertion of a TE in the cis-regulatory region of *fhl2b* in the haplochromine cichlids is associated with the formation of egg-spots (Santos *et al.*, 2014). In the Corydoradinae*,* transposable elements proliferation coincides with the WGD events and is driven largely by an increase in a single element (Marburger *et al.*, 2018). Thus, changes in transposable element abundance, coupled with the diversification of phenotypes in Corydoradinae makes them a valuable resource for exploring the effect of transposon insertion on diversification, particularly with regards to pigmentation (Marburger *et al.*, 2018).

1.4 Investigating Adaptation

Investigating how diversity arises and is maintained is fundamental to evolutionary biology and can only be fully understood with knowledge of the molecular underpinnings of phenotypic traits (Kratochwil and Meyer, 2015; Pardo-Diaz, Salazar and Jiggins, 2015). Linking environmental factors with phenotypes and genotypes comes with the same issues that many statistical tests face- the ability to detect true effect over noise. Being able to identify adaptive loci requires ruling out allele frequencies are not changing as a result of genetic drift, and to show that changes in frequency are correlated with changes in a phenotype. This is difficult to achieve, particularly in traits with complex architecture (Wellenreuther and Hansson, 2016). However, with the improvement of sequencing technologies, we have more power to detect true cause and effect loci which is rapidly pushing forward the identification of adaptive loci (Wellenreuther and Hansson, 2016). The methods for detecting adaptive loci are often delimited into "forward" and "reverse" techniques. In forward approach (also known as "top-down"), a phenotype of interest is first identified, and subsequently the genes underpinning the phenotype are explored. Conversely, the "reverse" (also known as "bottom-

up") approach identifies genes showing signals of selection and then investigates the associated phenotype (Pardo-Diaz, Salazar and Jiggins, 2015; Bomblies and Peichel, 2022). Here we are interested in genes associated with mimetic phenotypes, so will focus on forward approaches.

1.4.1 Identifying adaptive loci

1.4.1.1 Quantitative Trait Loci Mapping

Recombination breaks down associations between loci on the same chromosome, leaving only genetic variation in physical linkage to a locus controlling a trait of interest, known as a quantitative trait locus (QTL)(Martin and Jiggins, 2013). Quantitative trait loci mapping approaches can be used to identify the location of these sites in the genome. This approach involves creating a pool of genetically variable, recombinant offspring which can be used to create a linkage map with genetic markers (Mackay, Stone and Ayroles, 2009). The location of the QTL can then be established and its distance from the marker can be inferred from how often recombination arises (Martin and Jiggins, 2013). A major limtation of QTL mapping approaches is the requirement of large number of recombinant offspring to obtain enough statistical power (Pardo-Diaz, Salazar and Jiggins, 2015). Additionally, spurious QTLs and overinflated effect sizes can occur when there are non-representative allele frequencies within the mapping popualtion (e.g. few founders), population stratification (family or poulation structuring) or low envrionmental variance (overestimation of genetic components) (Wellenreuther and Hansson, 2016).

1.4.1.2 Genome Wide Association Studies

Genome wide association studies (GWAS) utilise natural variation within a population, reflective of historical linkage, to detect loci of effect (Martin and Jiggins, 2013). By comparing markers across the genome between two groups with known phenotypic differences, markers associated with the trait of interest can be identified (Nadeau and Jiggins, 2010). However, the ability to detect true association between loci and an adaptive trait (power) depends on the phenotypic variance of the population explained by the loci (Pardo-Diaz, Salazar and Jiggins, 2015; Fuentes-Pardo and Ruzzante, 2017). Genome wide association studies rely on historical recombination, not recent events as in QTL mapping, thus, GWAS have higher power to detect small effect loci (Martin and Jiggins, 2013). However, this is governed by how strongly the alternative allelic variants differ in their phenotypic effects (effect size) and their frequency in the sample (Pardo-Diaz, Salazar and Jiggins, 2015). As a result, detection of small effect loci, as in polygenic traits, require large sample sizes (Wellenreuther and Hansson, 2016; Barghi, Hermisson and Schlötterer, 2020; Fagny and Austerlitz, 2021). For example, 45% of the variance for human height is explained by small effect loci, which were previously undetected due to a lack of power (Yang *et al.*, 2010). One method for overcoming this limitation is to

weight loci, according to an environmental variable (Barghi, Hermisson and Schlötterer, 2020). Power to detect associations can be further reduced as a result of background genetic variation, which can reduce patterns of historic linkage disequilibrium (Alhaddad *et al.*, 2013; Wellenreuther and Hansson, 2016). To counteract the effects of low linkage disequilibrium, large sample sizes are needed (Li and Merilä, 2011; Santure and Garant, 2018). Additionally, the size of the population can affect the power of a GWAS, with a large population having lower linkage disequilibrium (Li and Merilä, 2011; Santure and Garant, 2018). This was demonstrated in the Siberian Jay population, where there was strong population substructure and subsequently higher linkage disequilibrium, resulting in fewer markers being required to detect effect loci (Li and Merilä, 2011). This highlights another potential confounding factor of GWAS- relatedness. Underlying relatedness and substructure can result in false associations; therefore, scans must account for population and family structure (Malenfant *et al.*, 2018; Santure and Garant, 2018). However, this can largely be accounted for by including a principal component as a covariate (Price *et al.*, 2006; Chen *et al.*, 2016).

1.4.1.3 Outlier approaches

Outlier approaches (also known as genome scans) function by identifying regions with elevated genetic differentiation, commonly using Cockerham and Weir's F_{ST} (Weir and Cockerham, 1984; Willing, Dreyer and van Oosterhout, 2012; Hoban *et al.*, 2016). The statistic F_{ST} is a measure of the variance in allele frequency among populations and as such can give indication about the degree of resemblance among individuals within populations (Holsinger and Weir, 2009). By dividing the genome into windows and calculating per window F_{ST} between two populations (or subsets of population), smaller regions with elevated differentiation can be identified (Beaumont, 2005; Hoban *et al.*, 2016; Clucas *et al.*, 2019). Methods to determine the threshold above which a genomic window is considered an outlier can vary, including the use of an empirical cut off or using statistical significance (Beaumont, 2005; Narum and Hess, 2011). Whilst outlier F_{ST} methods are less susceptible to low power due to low sample size, they can be impacted by demographic processes (Willing, Dreyer and van Oosterhout, 2012; Hoban *et al.*, 2016). For example, expansion in ranges can cause false signals of selection due to allele surfing (Hoban *et al.*, 2016). This occurs when a population at the range edge is small and subsequently individuals from these populations contribute disproportionately to the expansion, so alleles are overrepresented (Hoban *et al.*, 2016).

1.4.2 Sequencing Design

The decreasing cost of sequencing and technological advances have allowed the use of a greater number of genetic markers, greater sample sizes and therefore has increased the accuracy and precision of detecting loci under selection (Wellenreuther and Hansson, 2016). Different sequencing

strategies differ in their compromises, including the number of markers sequenced, the depth at which to sequence and therefore confidence in genotyping, the number of samples and ultimately, the cost of sequencing (Buerkle and Gompert, 2013).

1.4.2.1 Reduced representation sequencing

Restriction-site associated DNA sequencing is a reduced representation approach to whole genome sequencing which utilises common enzymes to sheer DNA into smaller fragments to be sequenced (Peterson *et al.*, 2012). This captures many genomic sites, at mid depth coverage, without sequencing the whole genome and without the need for a reference genome (Peterson *et al.*, 2012). Whilst RAD sequencing has been utilised for QTL-analysis and GWAS, it is limited due small to the generation of short read lengths resulting in difficulty in assembly (Lowry *et al.*, 2017). Additionally, due to the reduced coverage across the genome, important loci may be missed by sequencing (Lowry *et al.*, 2017). This may be avoided via carefully designed experiments, however the information required for this design is not available for many non-model organisms (Catchen *et al.*, 2017; Fuentes-Pardo and Ruzzante, 2017).

1.4.2.2 Pooling samples

An alternative strategy, Poolseq, compromises individual genotyping information, but has a better coverage across the genome (Micheletti and Narum, 2018). In this technique, DNA is extracted from individuals and is pooled together by phenotype or population (Micheletti and Narum, 2018). This is a resequencing approach, where individuals with contrasting phenotypes are sequenced and aligned to a reference genome (Fuentes-Pardo and Ruzzante, 2017). In conjunction with F_{ST} outlier analysis, Poolseq successfully identified genes associated with adaptation to sulphidic environments in *Poecilia mexicana* (Pfenninger *et al.*, 2015). However, the need for a reference genome is a major limitation for the use of resequencing methods in non-model organisms (Fuentes-Pardo and Ruzzante, 2017). Another limitation of the Poolseq approach is that it is biased towards the detection of frequent and large effect alleles (Fuentes-Pardo and Ruzzante, 2017).

1.4.2.3 Low coverage sequencing

Another method to reduce the cost of whole genome resequencing is by reducing sequencing depth. Low coverage whole genome sequencing (lcWGS) reduces cost by reducing the depth at which each nucleotide is sequenced and thus allows a larger number of samples to be sequenced (Fuentes-Pardo and Ruzzante, 2017; Lou *et al.*, 2021). In lcWGS genotype likelihoods per site are calculated incorporating the uncertainty of data due to sequencing, alignment and SNP calling errors and is therefore less biased than Poolseq (Fuentes-Pardo and Ruzzante, 2017). Even at very low depth (e.g. <

1x) inferences can be made with confidence (Buerkle and Gompert, 2013). This is in part due to the ability to sequence a large number of individuals, allowing for the higher power needed for GWAS techniques (Buerkle and Gompert, 2013). However, at low depths and at low sample size, high quantities of missing data can generate noise, which can interfere with F_{ST} outlier approaches (Lou *et al.*, 2021).

1.5 Thesis Aims and Objectives

The Corydoradinae represent an important vertebrate system in which to study the evolution and maintenance of Müllerian mimetic communities. The overarching aim of this thesis is to uncover the molecular mechanisms underpinning both aspects of the Müllerian mimetic phenotype, unpalatability and the advertisement of unpalatability. Despite 162 years of interest in these systems, the majority of research has focussed on Lepidopteran mimicry rings. By expanding this knowledge to a novel vertebrate system, we are adding important taxonomic breadth to the current literature exploring the evolution of Müllerian mimicry. This will allow us to understand how universal the patterns observed in the butterflies are, providing a better understanding of mimetic interactions. The specific aims of this thesis are:

1. To understand differences in the genomic composition and the drivers of genome size differences between *Hoplisoma* and *Corydoras*. Not all Corydoradinae species participate in Müllerian mimicry rings, suggesting biological characteristics may underpin the propensity to form mimetic interactions. Specifically, *Hoplisoma* is overrepresented in mimetic communities, often forming interactions with *Brochis* or *Corydoras* species. One clear disparity between *Hoplisoma* and other Corydoradinae species is a difference in genome size. Using a comparative genomic approach, we investigate the mechanisms through which the genome expansion of *Hoplisoma* occurred, including whole genome duplication and transposable element proliferation. This utilises novel genetic resources generated for this chapter, making this a foundational study through which the genomics of mimicry can be better understood.

2. To understand the genetic basis for polymorphic mimicry in the *Corydoras.* Colour polymorphisms in Müllerian mimics are paradoxical as deviations from the shared mimetic resemblance should be selected against. However, they also offer a unique comparative framework in which the genetics underpinning mimetic colouration can be studied. Here we aim to utilise natural variation in *C. fulleri* and low coverage whole genome resequencing to i) confirm the polymorphic nature of the species as opposed to

two hybridising taxa undergoing range overlap ii) identify genomic regions associated with the presence or absence of the melanic coloration, iii) identify specific genes associated with different colour phenotypes.

3. To uncover the molecular composition of venom in the Corydoradinae and explore differences in unpalatability between common co-mimics. Unpalatability is a key tenet of protective mimicry systems but is often overlooked relative to warning colouration. Understanding the relative toxicity of mimetic taxa can elucidate the nature and evolution of mimetic interactions. Here we aim to explore differences in venom potency and composition between two genera whose species commonly form mimetic communities, *Hoplisoma* and *Corydoras.* We first aim to explore differences in venom potency between members of *Hoplisoma* and *Corydoras*, using a brine shrimp cytotoxicity assay. We then aim to identify candidate venom and venom housekeeping genes within a single *Corydoras* species (*C. simulatus*) before exploring differences in the putative venom gene expression between members of the *Corydoras* and *Hoplisoma*. Ultimately this will allow us to better understand the nature of the interaction between common Corydoradinae co-mimics.

4. To understand the genetic basis for male-limited mimicry in the *Corydoras.* Sexual dimorphism is the result of asymmetrical inter-sex selection, known as sexually antagonistic selection. *Corydoras simulatus* is a sexually dimorphic species where different predator evasion strategies have been adopted. Males form a mimetic interaction with another sympatric Corydoradinae species*, Hoplisoma metae*, whilst females adopt a cryptic phenotype. In this study, we utilise a multi-omics approach to explore the genetics underpinning male-limited Müllerian mimicry in this system. We first aim to use whole genome resequencing to identify regions of the genome associated with inter-sex genetic differentiation. In conjunction we will use RNA sequencing to identify differences in expression associated with differences in colour pattern, reflecting those genes that are being utilised to adopt these alternate strategies.

1.6 References

Aardema, M.L., Stiassny, M.L.J. and Elizabeth Alter, S. (2020) Genomic Analysis of the Only Blind Cichlid Reveals Extensive Inactivation in Eye and Pigment Formation Genes, *Genome Biology And Evolution* . https://doi.org/10.1093/gbe/evaa144.

Alexandrou, M.A., Oliveira, C., Maillard, M., *et al.* (2011) Competition and phylogeny determine community structure in Müllerian co-mimics, *Nature*, 469(7328), pp. 84–88. https://doi.org/10.1038/nature09660.

Alexandrou, M.A., Swartz, B.A., Matzke, N.J., *et al.* (2013) Genome duplication and multiple evolutionary origins of complex migratory behavior in Salmonidae, *Molecular Phylogenetics And Evolution*, 69(3), pp. 514–523. https://doi.org/10.1016/j.ympev.2013.07.026.

Alhaddad, H., Khan, R., Grahn, R.A., *et al.* (2013) Extent of Linkage Disequilibrium in the Domestic Cat, Felis silvestris catus, and Its Breeds, *PloS One*, 8(1). https://doi.org/10.1371/journal.pone.0053537.

Amores, A., Force, A., Yan, Y.L., *et al.* (1998) Zebrafish hox clusters and vertebrate genome evolution, *Science*, 282(5394), pp. 1711–1714. https://doi.org/10.1126/science.282.5394.1711.

Anderson, B. and de Jager, M.L. (2020) Natural selection in mimicry, *Biological Reviews Of The Cambridge Philosophical Society*, 95(2), pp. 291–304. https://doi.org/10.1111/brv.12564.

Ángeles Esteban, M. (2012) An overview of the immunological defenses in fish skin, *ISRN Immunology*, 2012(1), pp. 1–29. https://doi.org/10.5402/2012/853470.

Arias, M., Mappes, J., Théry, M., *et al.* (2016) Inter-species variation in unpalatability does not explain polymorphism in a mimetic species, *Evolutionary Ecology*, 30(3), pp. 419–433. https://doi.org/10.1007/s10682-015-9815-2.

Axenrot, T.E. and Kullander, S.O. (2003) Corydoras diphyes (Siluriformes: Callichthyidae) and Otocinclus mimulus (Siluriformes: Loricariidae), two new species of catfishes from Paraguay, a case of a case of mimetic association, *Ichthyological Exploration Of Freshwaters*, 14(7), pp. 249–272. http://dx.doi.org/

Balogh, A.C.V., Gamberale-Stille, G., Tullberg, B.S., *et al.* (2010) Feature theory and the two-step hypothesis of müllerian mimicry evolution, *Evolution*, 64(3), pp. 810–822. https://doi.org/10.1111/j.1558-5646.2009.00852.x.

Balogh, A.C.V., Gamberale-Stille, G. and Leimar, O. (2008) Learning and the mimicry spectrum: from quasi-Bates to super-Müller, *Animal Behaviour*, 76(5), pp. 1591–1599. https://doi.org/10.1016/j.anbehav.2008.07.017.

Balogh, A.C.V. and Leimar, O. (2005) Müllerian mimicry: an examination of Fisher's theory of gradual evolutionary change, *The Royal Society Proceedings Biological Sciences*, 272(1578), pp. 2269–2275. https://doi.org/10.1098/rspb.2005.3227.

Barghi, N., Hermisson, J. and Schlötterer, C. (2020) Polygenic adaptation: a unifying framework to understand positive selection, *Nature Reviews. Genetics*, 21(12), pp. 769–781. https://doi.org/10.1038/s41576-020-0250-z.

Barnett, C.A., Bateson, M. and Rowe, C. (2007) State-dependent decision making: educated predators strategically trade off the costs and benefits of consuming aposematic prey, *Behavioral Ecology*, 18(4), pp. 645–651. https://doi.org/10.1093/beheco/arm027.

Bates, H.W. (1862) Contributions to an insect fauna of the Amazon valley. Lepidoptera: heliconinae, *Journal Of The Proceedings Of The Linnean Society Of London Zoology*, 6(22), pp. 73–77. https://doi.org/10.1111/j.1096-3642.1862.tb00932.x.

Beaumont, M.A. (2005) Adaptation and speciation: what can Fst tell us?, *Trends In Ecology & Evolution*, 20(8), pp. 435–440. https://doi.org/10.1016/j.tree.2005.05.017.

Bell, E.A., Cable, J., Oliveira, C., *et al.* (2020) Help or hindrance? The evolutionary impact of whole-genome duplication on immunogenetic diversity and parasite load, *Ecology And Evolution*, 10(24), pp. 13949–13956. https://doi.org/10.1002/ece3.6987.

Birskis-Barros, I., Freitas, A.V.L. and Guimarães, P.R. (2021) Habitat generalist species constrain the diversity of mimicry rings in heterogeneous habitats, *Scientific Reports*, 11(1). https://doi.org/10.1038/s41598-021-83867-w.

Bogan, S., Agnolin, F.L., Cenizo, M., *et al.* (2020) A Pleistocene freshwater ichthyofaunal assemblage from central Argentina: What kind of fishes lived in the Pampean lagoons before the extinction of the megafauna?, *PloS One*, 15(7). https://doi.org/10.1371/journal.pone.0235196.

Bomblies, K. (2023) Learning to tango with four (or more): the molecular basis of adaptation to polyploid meiosis, *Plant Reproduction*, 36(1), pp. 107–124. https://doi.org/10.1007/s00497-022-00448-1.

Bomblies, K. and Peichel, C.L. (2022) Genetics of adaptation, *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 119(30). https://doi.org/10.1073/pnas.2122152119.

Braasch, I. and Postlethwait, J.H. (2012) Polyploidy in Fish and the Teleost Genome Duplication, in P.S. Soltis and D.E. Soltis (eds) *Polyploidy And Genome Evolution*. Berlin, Heidelberg: Springer, pp. 341–383. https://doi.org/10.1007/978-3-642-31442-1_17.

Buerkle, A.C. and Gompert, Z. (2013) Population genomics based on low coverage sequencing: how low should we go?, *Molecular Ecology*, 22(11), pp. 3028–3035. https://doi.org/10.1111/mec.12105.

Cameron, A.M. and Endean, R. (1973) Epidermal secretions and the evolution of venom glands in fishes, *Toxicon*, 11(5), pp. 401–410. https://doi.org/10.1016/0041-0101(73)90115-3.

Cardoso, M.Z. and Gilbert, L.E. (2013) Pollen feeding, resource allocation and the evolution of chemical defence in passion vine butterflies, *Journal Of Evolutionary Biology*, 26(6), pp. 1254–1260. https://doi.org/10.1111/jeb.12119.

Caro, T. (2017) Wallace on coloration: Contemporary perspective and unresolved insights, *Trends In Ecology & Evolution*, 32(1), pp. 23–30. https://doi.org/10.1016/j.tree.2016.10.003.

Caro, T. and Allen, W.L. (2017) Interspecific visual signalling in animals and plants: a functional classification, *Philosophical Transactions Of The Royal Society Of London. Series B*, 372(1724). https://doi.org/10.1098/rstb.2016.0344.

Caro, T. and Ruxton, G. (2019) Aposematism: Unpacking the defences, *Trends In Ecology & Evolution*, 34(7), pp. 595–604. https://doi.org/10.1016/j.tree.2019.02.015.

Casewell, N.R., Wüster, W., Vonk, F.J., *et al.* (2013) Complex cocktails: the evolutionary novelty of venoms, *Trends In Ecology & Evolution*, 28(4), pp. 219–229. https://doi.org/10.1016/j.tree.2012.10.020.

Casewell, N.R., Visser, J.C., Baumann, K., *et al.* (2017) The Evolution of Fangs, Venom, and Mimicry Systems in Blenny Fishes, *Current Biology: CB*, 27(8), pp. 1184–1191. https://doi.org/10.1016/j.cub.2017.02.067.

Casewell, N.R., Petras, D., Card, D.C., *et al.* (2019) Solenodon genome reveals convergent evolution of venom in eulipotyphlan mammals, *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 116(51), pp. 25745–25755. https://doi.org/10.1073/pnas.1906117116.

Catchen, J.M., Hohenlohe, P.A., Bernatchez, L., *et al.* (2017) Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations, *Molecular Ecology Resources*, pp. 362–365. https://doi.org/10.1111/1755-0998.12669.

Caty, S.N., Alvarez-Buylla, A., Byrd, G.D., *et al.* (2019) Molecular physiology of chemical defenses in a poison frog, *The Journal Of Experimental Biology*, 222(12). https://doi.org/10.1242/jeb.204149.

Chen, G.-B., Lee, S.H., Zhu, Z.-X., *et al.* (2016) EigenGWAS: finding loci under selection through genome-wide association studies of eigenvectors in structured populations, *Heredity*, 117(1), pp. 51–61. https://doi.org/10.1038/hdy.2016.25.

Chen, Z., Omori, Y., Koren, S., *et al.* (2019) De novo assembly of the goldfish (*Carassius auratus*) genome and the evolution of genes after whole-genome duplication, *Science Advances*, 5(6). https://doi.org/10.1126/sciadv.aav0547.

Chouteau, M., Summers, K., Morales, V., *et al.* (2011) Advergence in Müllerian mimicry: the case of the poison dart frogs of Northern Peru revisited, *Biology Letters*, 7(5), pp. 796–800. https://doi.org/10.1098/rsbl.2011.0039.

Chuang, P.-S. and Shiao, J.-C. (2014) Toxin gene determination and evolution in scorpaenoid fish, *Toxicon*, 88, pp. 21–33. https://doi.org/10.1016/j.toxicon.2014.06.013.

Clucas, G.V., Lou, R.N., Therkildsen, N.O., *et al.* (2019) Novel signals of adaptive genetic variation in northwestern Atlantic cod revealed by whole-genome sequencing, *Evolutionary Applications*, 12(10), pp. 1971–1987. https://doi.org/10.1111/eva.12861.

Coyne, J.A., Barton, N.H. and Turelli, M. (2000) Is Wright's shifting balance process important in evolution?, *Evolution*, 54(1), pp. 306–317. https://doi.org/10.1111/j.0014-3820.2000.tb00033.x.

Cui, J., Chen, Y., Hines, H.M., *et al.* (2024) Does coevolution in refugia drive mimicry in bumble bees? Insights from a South Asian mimicry group, *Science Advances*, 10(24). https://doi.org/10.1126/sciadv.adl2286.

Dias, A.C., Tencatt, L.F.C., Roxo, F.F., *et al.* (2024) Phylogenomic analyses in the complex Neotropical subfamily Corydoradinae (Siluriformes: Callichthyidae) with a new classification based on morphological and molecular data, *Zoological Journal Of The Linnean Society* .

Doktorovová, L., Exnerová, A., Hotová Svádová, K., *et al.* (2019) Differential bird responses to colour morphs of an aposematic leaf beetle may affect variation in morph frequencies in polymorphic prey populations, *Evolutionary Biology*, 46(1), pp. 35–46. https://doi.org/10.1007/s11692-018-9465-8.

Elkin, J., Martin, A., Courtier-Orgogozo, V., *et al.* (2023) Analysis of the genetic loci of pigment pattern evolution in vertebrates, *Biological Reviews Of The Cambridge Philosophical Society* . https://doi.org/10.1111/brv.12952.

Ellisdon, A.M., Reboul, C.F., Panjikar, S., *et al.* (2015) Stonefish toxin defines an ancient branch of the perforin-like superfamily, *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 112(50), pp. 15360–15365. https://doi.org/10.1073/pnas.1507622112.

Enciso-Romero, J., Pardo-Díaz, C., Martin, S.H., *et al.* (2017) Evolution of novel mimicry rings facilitated by adaptive introgression in tropical butterflies, *Molecular Ecology*, 26(19), pp. 5160–5172. https://doi.org/10.1111/mec.14277.

Endler, J.A. (1978) A Predator's View of Animal Color Patterns, in M.K. Hecht, W.C. Steere, and B. Wallace (eds) *Evolutionary Biology*. Boston, MA: Springer US, pp. 319–364. https://doi.org/10.1007/978-1-4615-6956-5_5.

Endler, J.A. (1983) Natural and sexual selection on color patterns in poeciliid fishes, *Environmental Biology Of Fishes*, 9(2), pp. 173–190. https://doi.org/10.1007/BF00690861.

Fagny, M. and Austerlitz, F. (2021) Polygenic Adaptation: Integrating Population Genetics and Gene Regulatory Networks, *Trends In Genetics*, 37(7), pp. 631–638. https://doi.org/10.1016/j.tig.2021.03.005.

Fisher, R.A. (1927) On some objections to mimicry theory; Statistical and genetic, *Transactions Of The Royal Entomological Society Of London*, 75(2), pp. 269–278. https://doi.org/10.1111/j.1365-2311.1927.tb00074.x.

Fisher, R.A. (1931) The genetical theory of natural selection, *Journal Of The Royal Statistical Society*, 94(1). https://doi.org/10.2307/2341823.

Font, E. (2019) Mimicry, Camouflage and Perceptual Exploitation: the Evolution of Deception in Nature, *Biosemiotics*, 12(1), pp. 7–24. https://doi.org/10.1007/s12304-018-9339-6.

Foster, B.J., McCulloch, G.A., Foster, Y., *et al.* (2023) ebony underpins Batesian mimicry in melanic stoneflies, *Molecular Ecology*, 32(18), pp. 4986–4998. https://doi.org/10.1111/mec.17085.

Franks, D.W. and Sherratt, T.N. (2007) The evolution of multicomponent mimicry, *Journal Of Theoretical Biology*, 244(4), pp. 631–639. https://doi.org/10.1016/j.jtbi.2006.09.019.

Fricke, R., Eschmeyer, W.N. and Van der Laan, R. (2024) *Eschmeyer's Catalog of Fishes*. Available at: https://researcharchive.calacademy.org/research/ichthyology/catalog/fishcatmain.asp (Accessed: 19 September 2024).

Fuentes-Pardo, A.P. and Ruzzante, D.E. (2017) Whole-genome sequencing approaches for conservation biology: Advantages, limitations and practical recommendations, *Molecular Ecology*, 26(20), pp. 5369–5406. https://doi.org/10.1111/mec.14264.

Fuller, I.A.M. and Evers, H.E. (2005) *Identifying Corydoradinae Catfishe*. Kidderminster, England: Ian Fuller Enterprises.

Garg, K.M., Sam, K., Chattopadhyay, B., *et al.* (2019) Gene flow in the müllerian mimicry ring of a poisonous Papuan songbird clade (Pitohui; Aves), *Genome Biology And Evolution*, 11(8), pp. 2332–2343. https://doi.org/10.1093/gbe/evz168.

Glasauer, S.M.K. and Neuhauss, S.C.F. (2014) Whole-genome duplication in teleost fishes and its evolutionary consequences, *Molecular Genetics And Genomics*, 289(6), pp. 1045–1060. https://doi.org/10.1007/s00438-014-0889-2.

Gompert, Z., Willmott, K. and Elias, M. (2011) Heterogeneity in predator micro-habitat use and the maintenance of Müllerian mimetic diversity, *Journal Of Theoretical Biology*, 281(1), pp. 39–46. https://doi.org/10.1016/j.jtbi.2011.04.024.

Gonçalves, D., Oliveira, R.F., Körner, K., *et al.* (2003) Intersexual copying by sneaker males of the peacock blenny, *Animal Behaviour*, 65(2), pp. 355–361. https://doi.org/10.1006/anbe.2003.2065.

Gratzer, B., Millesi, E., Walzl, M., *et al.* (2015) Skin toxins in coral-associated Gobiodon species (Teleostei: Gobiidae) affect predator preference and prey survival, *Marine Ecology*, 36(1), pp. 67–76. https://doi.org/10.1111/maec.12117.

Greenwood, A.K., Jones, F.C., Chan, Y.F., *et al.* (2011) The genetic basis of divergent pigment patterns in juvenile threespine sticklebacks, *Heredity*, 107(2), pp. 155–166. https://doi.org/10.1038/hdy.2011.1.

Greenwood, A.K., Cech, J.N. and Peichel, C.L. (2012) Molecular and developmental contributions to divergent pigment patterns in marine and freshwater sticklebacks, *Evolution & Development*, 14(4), pp. 351–362. https://doi.org/10.1111/j.1525-142X.2012.00553.x.

Greven, H., Flasbeck, T. and Passia, D. (2006) Axillary glands in the armoured catfish Corydoras aeneus (Callichthyidae, Siluriformes), *Verhandlungen Der Gesellschaft Für Ichthyologie*, 5, pp. 65–69. https://doi.org/10.1590/1809-4392202101350

Gundappa, M.K., To, T.-H., Grønvold, L., *et al.* (2022) Genome-wide reconstruction of rediploidization following autopolyploidization across one hundred million years of Salmonid evolution, *Molecular Biology And Evolution*, 39(1). https://doi.org/10.1093/molbev/msab310.

Han, H., Baumann, K., Casewell, N.R., *et al.* (2017) The Cardiovascular and Neurotoxic Effects of the Venoms of Six Bony and Cartilaginous Fish Species, *Toxins*, 9(2). https://doi.org/10.3390/toxins9020067.

Haney, R.A., Matte, T., Forsyth, F.S., *et al.* (2019) Alternative transcription at venom genes and its role as a complementary mechanism for the generation of venom complexity in the common house spider, *Frontiers In Ecology And Evolution*, 7. https://doi.org/10.3389/fevo.2019.00085.

Harris, R.J. and Jenner, R.A. (2019) Evolutionary Ecology of Fish Venom: Adaptations and Consequences of Evolving a Venom System, *Toxins*, 11(2). https://doi.org/10.3390/toxins11020060.

Hoban, S., Kelley, J.L., Lotterhos, K.E., *et al.* (2016) Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions, *The American Naturalist*, 188(4), pp. 379–397. https://doi.org/10.1086/688018.

Hoekstra, H.E. (2006) Genetics, development and evolution of adaptive pigmentation in vertebrates, *Heredity*, 97(3), pp. 222–234. https://doi.org/10.1038/sj.hdy.6800861.

Holmes, I.A., Grundler, M.R. and Davis Rabosky, A.R. (2017) Predator Perspective Drives Geographic Variation in Frequency-Dependent Polymorphism, *The American Naturalist*, 190(4), pp. 78–93. https://doi.org/10.1086/693159.

Holsinger, K.E. and Weir, B.S. (2009) Genetics in geographically structured populations: defining, estimating and interpreting F(ST), *Nature Reviews. Genetics*, 10(9), pp. 639–650. https://doi.org/10.1038/nrg2611.

Huang, W., Haubold, B., Hauert, C., *et al.* (2012) Emergence of stable polymorphisms driven by evolutionary games between mutants, *Nature Communications*, 3(1), p. 919. https://doi.org/10.1038/ncomms1930.

Huber, B., Whibley, A., Poul, Y.L., *et al.* (2015) Conservatism and novelty in the genetic architecture of adaptation in Heliconius butterflies, *Heredity*, 114(5), pp. 515–524. https://doi.org/10.1038/hdy.2015.22.

Iijima, T., Kajitani, R., Komata, S., *et al.* (2018) Parallel evolution of Batesian mimicry supergene in two Papilio butterflies, P. polytes and P. memnon, *Science Advances* . https://doi.org/10.1126/sciadv.aao5416.

Irion, U. and Nüsslein-Volhard, C. (2019) The identification of genes involved in the evolution of color patterns in fish, *Current Opinion In Genetics & Development*, 57, pp. 31–38. https://doi.org/10.1016/j.gde.2019.07.002.

Jay, P., Whibley, A., Frézal, L., *et al.* (2018) Supergene evolution triggered by the introgression of a chromosomal inversion, *Current Biology*, 28(11), pp. 1839-1845.e3. https://doi.org/10.1016/j.cub.2018.04.072.

Jeckel, A.M., Kocheff, S., Saporito, R.A., *et al.* (2019) Geographically separated orange and blue populations of the Amazonian poison frog Adelphobates galactonotus (Anura, Dendrobatidae) do not differ in alkaloid composition or palatability, *Chemoecology*, 29(5–6), pp. 225–234. https://doi.org/10.1007/s00049-019-00291-3.

Jiggins, C.D. (2016) Genes on the wing: colour pattern genetics, in *The Ecology And Evolution Of Heliconius Butterflies*. Oxford University Press, pp. 112–137. https://doi.org/10.1093/acprof:oso/9780199566570.003.0008.

John, L., Rick, I.P., Vitt, S., *et al.* (2021) Body coloration as a dynamic signal during intrasexual communication in a cichlid fish, *BMC Zoology*, 6(1). https://doi.org/10.1186/s40850-021-00075-9.

Joron, M., Frezal, L., Jones, R.T., *et al.* (2011) Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry, *Nature*, 477(7363), pp. 203–206. https://doi.org/10.1038/nature10341.

Kapan, D.D. (2001) Three-butterfly system provides a field test of müllerian mimicry, *Nature*, 409(6818), pp. 338–340. https://doi.org/10.1038/35053066.

Kiehl, E., Rieger, C. and Greven, H. (2006) Axillary gland secretions contribute to the stress- induced discharge of a bactericidal substance in Corydoras sterbai (Callichthyidae, Siluriformes), *Verhandlungen Der Gesellschaft For Ichthyologie*, 5.

Kikuchi, D.W., Waldron, S.J., Valkonen, J.K., *et al.* (2020) Biased predation could promote convergence yet maintain diversity within Müllerian mimicry rings of Oreina leaf beetles, *Journal Of Evolutionary Biology*, 33(7), pp. 887–898. https://doi.org/10.1111/jeb.13620.

Klaassen, H., Wang, Y., Adamski, K., *et al.* (2018) CRISPR mutagenesis confirms the role of oca2 in melanin pigmentation in Astyanax mexicanus, *Developmental Biology*, 441(2), pp. 313–318. https://doi.org/10.1016/j.ydbio.2018.03.014.

Koludarov, I., Velasque, M., Senoner, T., *et al.* (2023) Prevalent bee venom genes evolved before the aculeate stinger and eusociality, *BMC Biology*, 21(1). https://doi.org/10.1186/s12915-023-01656-5.

Komata, S., Lin, C.-P. and Fujiwara, H. (2022) doublesex Controls Both Hindwing and Abdominal Mimicry Traits in the Female-Limited Batesian Mimicry of Papilio memnon, *Frontiers In Insect Science*, 2. https://doi.org/10.3389/finsc.2022.929518.

Kordis, D. and Gubensek, F. (2000) Adaptive evolution of animal toxin multigene families, *Gene*, 261(1), pp. 43–52. https://doi.org/10.1016/s0378-1119(00)00490-x.

Kozak, K.M., Joron, M., McMillan, W.O., *et al.* (2021) Rampant genome-wide admixture across the Heliconius radiation, *Genome Biology And Evolution*, 13(7). https://doi.org/10.1093/gbe/evab099.

Kratochwil, C.F., Liang, Y., Gerwin, J., *et al.* (2018) Agouti-related peptide 2 facilitates convergent evolution of stripe patterns across cichlid fish radiations, *Science*, 362(6413), pp. 457–460. https://doi.org/10.1126/science.aao6809.

Kratochwil, C.F., Kautt, A.F., Nater, A., *et al.* (2022) An intronic transposon insertion associates with a trans-species color polymorphism in Midas cichlid fishes, *Nature Communications*, 13(1). https://doi.org/10.1038/s41467-021-27685-8.

Kratochwil, C.F. and Meyer, A. (2015) Closing the genotype-phenotype gap: emerging technologies for evolutionary genetics in ecological model vertebrate systems, *BioEssays: News And Reviews In Molecular, Cellular And Developmental Biology*, 37(2), pp. 213–226. https://doi.org/10.1002/bies.201400142.

Kratochwil, C.F., Urban, S. and Meyer, A. (2019) Genome of the Malawi golden cichlid fish (Melanochromis auratus) reveals exon loss of oca2 in an amelanistic morph, *Pigment Cell & Melanoma Research*, 32(5), pp. 719–723. https://doi.org/10.1111/pcmr.12799.

Kronforst, M.R., Kapan, D.D. and Gilbert, L.E. (2006) Parallel genetic architecture of parallel adaptive radiations in mimetic Heliconius butterflies, *Genetics*, 174(1), pp. 535–539. https://doi.org/10.1534/genetics.106.059527.

Kunte, K., Zhang, W., Tenger-Trolander, A., *et al.* (2014) Doublesex is a mimicry supergene, *Nature*, 507(7491). https://doi.org/10.1038/nature13112.

Kuraku, S. and Meyer, A. (2009) The evolution and maintenance of Hox gene clusters in vertebrates and the teleost-specific genome duplication, *The International Journal Of Developmental Biology*, 53(5–6), pp. 765–773. https://doi.org/10.1387/ijdb.072533km.

Langham, G.M. (2004) Specialized avian predators repeatedly attack novel color morphs of Heliconius butterflies, *Evolution*, 58(12), pp. 2783–2787. https://doi.org/10.1111/j.0014-3820.2004.tb01629.x.

Layton, K.K.S., Carvajal, J.I. and Wilson, N.G. (2020) Mimicry and mitonuclear discordance in nudibranchs: New insights from exon capture phylogenomics, *Ecology And Evolution*, 10(21), pp. 11966–11982. https://doi.org/10.1002/ece3.6727.

Li, M.-H. and Merilä, J. (2011) Population differences in levels of linkage disequilibrium in the wild, *Molecular Ecology*, 20(14), pp. 2916–2928. https://doi.org/10.1111/j.1365-294X.2011.05154.x.

Lindström, L., Alatalo, R.V., Mappes, J., *et al.* (1999) Can aposematic signals evolve by gradual change?, *Nature*, 397(6716), pp. 249–251. https://doi.org/10.1038/16692.

Linnen, C.R., Poh, Y.-P., Peterson, B.K., *et al.* (2013) Adaptive evolution of multiple traits through multiple mutations at a single gene, *Science*, 339(6125), pp. 1312–1316. https://doi.org/10.1126/science.1233213.

Livraghi, L., Hanly, J.J., Van Bellghem, S.M., *et al.* (2021) Cortex cis-regulatory switches establish scale colour identity and pattern diversity in Heliconius, *ELife*, 10. https://doi.org/10.7554/eLife.68549.

Lopes-Ferreira, M., Sosa-Rosales, I., Silva Junior, P.I., *et al.* (2021) Molecular characterization and functional analysis of the nattectin-like toxin from the venomous fish Thalassophryne maculosa, *Toxins*, 14(1). https://doi.org/10.3390/toxins14010002.

Lou, R.N., Jacobs, A., Wilder, A.P., *et al.* (2021) A beginner's guide to low-coverage whole genome sequencing for population genomics, *Molecular Ecology*, 30(23), pp. 5966–5993. https://doi.org/10.1111/mec.16077.

Lowry, D.B., Hoban, S., Kelley, J.L., *et al.* (2017) Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation, *Molecular Ecology Resources*, 17(2), pp. 142–152. https://doi.org/10.1111/1755-0998.12635.

Ma, W., Zhu, Z.-H., Bi, X.-Y., *et al.* (2014) Allopolyploidization is not so simple: evidence from the origin of the tribe Cyprinini (Teleostei: Cypriniformes), *Current Molecular Medicine*, 14(10), pp. 1331–1338. https://doi.org/10.2174/1566524014666141203101543.

Mable, B.K., Alexandrou, M.A. and Taylor, M.I. (2011) Genome duplication in amphibians and fish: an extended synthesis, *Journal Of Zoology* , 284(3), pp. 151–182. https://doi.org/10.1111/j.1469-7998.2011.00829.x.

Machado, V., Araujo, A.M., Serrano, J., *et al.* (2004) Phylogenetic relationships and the evolution of mimicry in the Chauliognathus yellow-black species complex (Coleoptera: Cantharidae) inferred from mitochondrial COI sequences, *Genetics And Molecular Biology*, 27(1), pp. 55–60. https://doi.org/10.1590/S1415-47572004000100010.

Mackay, T.F.C., Stone, E.A. and Ayroles, J.F. (2009) The genetics of quantitative traits: challenges and prospects, *Nature Reviews Genetics*, 10(8), pp. 565–577. https://doi.org/10.1038/nrg2612.

Macqueen, D.J. and Johnston, I.A. (2014) A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification, *Proceedings Of The Royal Society B*, 281(1778). https://doi.org/10.1098/rspb.2013.2881.

Malenfant, R.M., Davis, C.S., Richardson, E.S., *et al.* (2018) Heritability of body size in the polar bears of Western Hudson Bay, *Molecular Ecology Resources*, 18(4), pp. 854–866. https://doi.org/10.1111/1755-0998.12889.

Mallet, J. (1989) The genetics of warning colour in Peruvian hybrid zones of Heliconius erato and H. melpomene, *Proceedings Of The Royal Society Of London*, 236(1283), pp. 163–185. https://doi.org/10.1098/rspb.1989.0019.

Mallet, J. (2010) Shift happens! Shifting balance and the evolution of diversity in warning colour and mimicry, *Ecological Entomology*, 35, pp. 90–104. https://doi.org/10.1111/j.1365-2311.2009.01137.x.

Mallet, J. and Barton, N.H. (1989) Strong natural selection in a warning-color hybrid zone, *Evolution*, 43(2). https://doi.org/10.2307/2409217.

Mallet, J. and Joron, M. (1999) Evolution of Diversity in Warning Color and Mimicry: Polymorphisms, Shifting Balance, and Speciation, *Annual Review Of Ecology And Systematics*, 30(1), pp. 201–233. https://doi.org/10.1146/annurev.ecolsys.30.1.201.

Marburger, S., Alexandrou, M.A., Taggart, J.B., *et al.* (2018) Whole genome duplication and transposable element proliferation drive genome expansion in Corydoradinae catfishes, *Proceedings Of The Royal Society B*, 285(1872). https://doi.org/10.1098/rspb.2017.2732.

Marek, P.E. and Bond, J.E. (2009) A Müllerian mimicry ring in Appalachian millipedes, *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 106(24), pp. 9755–9760. https://doi.org/10.1073/pnas.0810408106.

Marples, N.M. and Mappes, J. (2011) Can the dietary conservatism of predators compensate for positive frequency dependent selection against rare, conspicuous prey?, *Evolutionary Ecology*, 25(4), pp. 737–749. https://doi.org/10.1007/s10682-010-9434-x.

Marples, N.M., Speed, M.P. and Thomas, R.J. (2018) An individual-based profitability spectrum for understanding interactions between predators and their prey, *Biological Journal Of The Linnean Society. Linnean Society Of London*, 125(1), pp. 1–13. https://doi.org/10.1093/biolinnean/bly088.

Martin, S.H. and Jiggins, C.D. (2013) Genomic studies of adaptation in natural populations, *ELS* . https://doi.org/10.1002/9780470015902.a0024613.

Martinson, E.O., Mrinalini, Kelkar, Y.D., *et al.* (2017) The evolution of venom by co-option of single-copy genes, *Current Biology: CB*, 27(13), pp. 2007-2013.e8. https://doi.org/10.1016/j.cub.2017.05.032.

McLean, C.A. and Stuart-Fox, D. (2014) Geographic variation in animal colour polymorphisms and its role in speciation, *Biological Reviews Of The Cambridge Philosophical Society*, 89(4), pp. 860–873. https://doi.org/10.1111/brv.12083.

McMillan, W.O., Livraghi, L., Concha, C., *et al.* (2020) From patterning genes to process: Unraveling the gene regulatory networks that pattern Heliconius wings, *Frontiers In Ecology And Evolution*, 8. https://doi.org/10.3389/fevo.2020.00221.

Merrill, R.M., Dasmahapatra, K.K., Davey, J.W., *et al.* (2015) The diversification of Heliconius butterflies: what have we learned in 150 years?, *Journal Of Evolutionary Biology*, 28(8), pp. 1417–1438. https://doi.org/10.1111/jeb.12672.

Meyer, A. and Van de Peer, Y. (2005) From 2R to 3R: evidence for a fish-specific genome duplication (FSGD), *BioEssays*, 27(9), pp. 937–945. https://doi.org/10.1002/bies.20293.

Micheletti, S.J. and Narum, S.R. (2018) Utility of pooled sequencing for association mapping in nonmodel organisms, *Molecular Ecology Resources*, 18(4), pp. 825–837. https://doi.org/10.1111/1755-0998.12784.

Miller, C.T., Beleza, S., Pollen, A.A., *et al.* (2007) cis-Regulatory Changes in Kit Ligand Expression and Parallel Evolution of Pigmentation in Sticklebacks and Humans, *Cell*, 131(6), pp. 1179–1189. https://doi.org/10.1016/j.cell.2007.10.055.

Motyka, M., Bocek, M., Kusy, D., *et al.* (2020) Interactions in multi-pattern Müllerian communities support origins of new patterns, false structures, imperfect resemblance and mimetic sexual dimorphism, *Scientific Reports*, 10(1). https://doi.org/10.1038/s41598-020-68027-w.

Motyka, M., Kusy, D., Masek, M., *et al.* (2021) Conspicuousness, phylogenetic structure, and origins of Müllerian mimicry in 4000 lycid beetles from all zoogeographic regions, *Scientific Reports*, 11(1). https://doi.org/10.1038/s41598-021-85567-x.

Motyka, M., Kampova, L. and Bocak, L. (2018) Phylogeny and evolution of Müllerian mimicry in aposematic Dilophotes: evidence for advergence and size-constraints in evolution of mimetic sexual dimorphism, *Scientific Reports*, 8(1), pp. 1–10. https://doi.org/10.1038/s41598-018-22155-6.

Muell, M.R., Chávez, G., Prates, I., *et al.* (2022) Phylogenomic analysis of evolutionary relationships in Ranitomeya poison frogs (Family Dendrobatidae) using ultraconserved elements, *Molecular Phylogenetics And Evolution*, 168(107389). https://doi.org/10.1016/j.ympev.2022.107389.

Müller, F. (1878) Ueber Die Vortheile Der Mimicry Bei Schmetterlingen, *Zool. Anzeiger.*

Nadeau, N.J. and Jiggins, C.D. (2010) A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations, *Trends In Genetics*, 26(11), pp. 484–492. https://doi.org/10.1016/j.tig.2010.08.004.

Narum, S.R. and Hess, J.E. (2011) Comparison of F(ST) outlier tests for SNP loci under selection, *Molecular Ecology Resources*, 11 Suppl 1(s1), pp. 184–194. https://doi.org/10.1111/j.1755-0998.2011.02987.x.

Nicholson, A.J. (1927) A new theory of mimicry in insects, *The Australian Zoologist*, 5, pp. 10–104.

Nijssen, H. (1970) Records of the catfish genus Corydoras from Brazil and french Guiana with descriptions of eight new species (Pisces, Siluriformes, callichthyidae), *Netherlands Journal Of Zoology*, 21(4), pp. 412–433. https://doi.org/10.1163/002829671x00078.

Nur, U. (1970) Evolutionary rates of models and mimics in Batesian mimicry, *The American Naturalist*, 104(939), pp. 477–486. https://doi.org/10.1086/282682.

Ohno, S. (1970) *Evolution by gene duplication*. New York: Springer.

Oliveira, C., Almeida-Toledo, L.F., Mori, L., *et al.* (1992) Extensive chromosomal rearrangements and nuclear DNA content changes in the evolution of the armoured catfishes genus Corydoras (Pisces, Siluriformes, Callichthyidae), *Journal Of Fish Biology*, 40(3), pp. 419–431. https://doi.org/10.1111/j.1095-8649.1992.tb02587.x.

Oliveira, C., Almeida-Toledo, L.F., Mori, L., *et al.* (1993) Cytogenetic and DNA content in six genera of the family Callichthyidae (Pisces, Siluriformes), *Caryologia*, 46(2–3), pp. 171–188. https://doi.org/10.1080/00087114.1993.10797258.

Papa, R., Kapan, D.D., Counterman, B.A., *et al.* (2013) Multi-allelic major effect genes interact with minor effect QTLs to control adaptive color pattern variation in Heliconius erato, *PloS One*, 8(3). https://doi.org/10.1371/journal.pone.0057033.

Pardo-Diaz, C., Salazar, C., Baxter, S.W., *et al.* (2012) Adaptive introgression across species boundaries in Heliconius butterflies, *PLoS Genetics*, 8(6), p. e1002752. https://doi.org/10.1371/journal.pgen.1002752.

Pardo-Diaz, C., Salazar, C. and Jiggins, C.D. (2015) Towards the identification of the loci of adaptive evolution, *Methods In Ecology And Evolution*, 6(4), pp. 445–464. https://doi.org/10.1111/2041-210X.12324.

Parichy, D.M. (2021) Evolution of pigment cells and patterns: recent insights from teleost fishes, *Current Opinion In Genetics & Development*, 69, pp. 88–96. https://doi.org/10.1016/j.gde.2021.02.006.

Paris, J.R., Whiting, J.R., Daniel, M.J., *et al.* (2022) A large and diverse autosomal haplotype is associated with sex-linked colour polymorphism in the guppy, *Nature Communications*, 13(1). https://doi.org/10.1038/s41467-022-28895-4.

Peterson, B.K., Weber, J.N., Kay, E.H., *et al.* (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species, *PloS One*, 7(5). https://doi.org/10.1371/journal.pone.0037135.

Pfenninger, M., Patel, S., Arias-Rodriguez, L., *et al.* (2015) Unique evolutionary trajectories in repeated adaptation to hydrogen sulphide-toxic habitats of a neotropical fish (Poecilia mexicana), *Molecular Ecology*, 24(21), pp. 5446–5459. https://doi.org/10.1111/mec.13397.

Price, A.L., Patterson, N.J., Plenge, R.M., *et al.* (2006) Principal components analysis corrects for stratification in genome-wide association studies, *Nature Genetics*, 38(8), pp. 904–909. https://doi.org/10.1038/ng1847.

Protas, M.E., Hersey, C., Kochanek, D., *et al.* (2005) Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism, *Nature Genetics*, 38(1), pp. 107–111. https://doi.org/10.1038/ng1700.

Reed, R.D., Papa, R., Martin, A., *et al.* (2011) Optix drives the repeated convergent evolution of butterfly wing pattern mimicry, *Science*, 333(6046), pp. 1137–1141. https://doi.org/10.1126/science.1208227.

Reis, R.E. (1998) Anatomy and phylogenetic analysis of the neotropical callichthyid catfishes (Ostariophysi, Siluriformes), *Zoological Journal Of The Linnean Society*, 124(2), pp. 105–168. https://doi.org/10.1111/j.1096-3642.1998.tb00571.x.

Roberts, R.B., Ser, J.R. and Kocher, T.D. (2009) Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes, *Science*, 326(5955), pp. 998–1001. https://doi.org/10.1126/science.1174705.

Robertson, F.M., Gundappa, M.K., Grammes, F., *et al.* (2017) Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification, *Genome Biology*, 18(1). https://doi.org/10.1186/s13059-017-1241-z.

Rowland, H.M., Mappes, J., Ruxton, G.D., *et al.* (2010) Mimicry between unequally defended prey can be parasitic: evidence for quasi-Batesian mimicry, *Ecology Letters*, 13(12), pp. 1494–1502. https://doi.org/10.1111/j.1461-0248.2010.01539.x.

Rubio, A.O., Stuckert, A.M.M., Geralds, B., *et al.* (2024) What makes a mimic? Orange, red, and black color production in the mimic poison frog (Ranitomeya imitator), *Genome Biology And Evolution*, 16(7). https://doi.org/10.1093/gbe/evae123.

Ruxton, G.D., Allen, W.L., Sherratt, T.N., *et al.* (2018) Avoiding Attack: The Evolutionary Ecology of Crypsis, Aposematism, and Mimicry, in *Avoiding Attack*. 2nd edn. Oxford University Press. https://doi.org/10.1093/oso/9780199688678.001.0001.

Saenko, S.V., Chouteau, M., Piron-Prunier, F., *et al.* (2019) Unravelling the genes forming the wing pattern supergene in the polymorphic butterfly Heliconius numata, *EvoDevo*, 10. https://doi.org/10.1186/s13227-019-0129-2.

Salis, P., Lorin, T., Laudet, V., *et al.* (2019) Magic Traits in Magic Fish: Understanding Color Pattern Evolution Using Reef Fish, *Trends In Genetics*, 35(4), pp. 265–278. https://doi.org/10.1016/j.tig.2019.01.006.

Sands, D. (1994) *The behaviour and evolutionary ecology of corydoras adolfoi and corydoras imitator: Studies on two species of sympatric catfish from the upper rio negro, brazil*. PhD. Liverpool University.

Santos, M.E., Braasch, I., Boileau, N., *et al.* (2014) The evolution of cichlid fish egg-spots is linked with a cis-regulatory change, *Nature Communications*, 5(1), pp. 1–11. https://doi.org/10.1038/ncomms6149.

Santure, A.W. and Garant, D. (2018) Wild GWAS—association mapping in natural populations, *Molecular Ecology Resources*, 18(4), pp. 729–738. https://doi.org/10.1111/1755-0998.12901.

Saporito, R.A., Spande, T.F., Martin Garraffo, H., *et al.* (2009) Arthropod alkaloids in poison frogs: A review of the 'Dietary Hypothesis', *Heterocycles*, 79(1). https://doi.org/10.3987/REV-08-SR(D)11.

Saporito, R.A., Donnelly, M.A., Madden, A.A., *et al.* (2010) Sex-related differences in alkaloid chemical defenses of the dendrobatid frog Oophaga pumilio from Cayo Nancy, Bocas del Toro, Panama, *Journal Of Natural Products*, 73(3), pp. 317–321. https://doi.org/10.1021/np900702d.

Schendel, V., Rash, L.D., Jenner, R.A., *et al.* (2019) The diversity of venom: The importance of behavior and venom system morphology in understanding its ecology and evolution, *Toxins*, 11(11). https://doi.org/10.3390/toxins11110666.

Schrader, L. and Schmitz, J. (2019) The impact of transposable elements in adaptive evolution, *Molecular Ecology*, 28(6), pp. 1537–1549. https://doi.org/10.1111/mec.14794.

Seehausen, O., Mayhew, P.J. and Van Alphen, J.J.M. (1999) Evolution of colour patterns in east African cichlid fish, *Journal Of Evolutionary Biology*, 12, pp. 514–534.

Sherratt, T.N. (2008) The evolution of Müllerian mimicry, *The Science Of Nature*, 95(8), pp. 681–695. https://doi.org/10.1007/s00114-008-0403-y.

Sherratt, T.N. and Beatty, C.D. (2003) The evolution of warning signals as reliable indicators of prey defense, *The American Naturalist*, 162(4), pp. 377–389. https://doi.org/10.1086/378047.

Shimabukuro-Dias, C.K., Oliveira, C., Reis, R.E., *et al.* (2004) Molecular phylogeny of the armored catfish family Callichthyidae (Ostariophysi, Siluriformes), *Molecular Phylogenetics And Evolution*, 32(1), pp. 152–163. https://doi.org/10.1016/j.ympev.2003.11.010.

Shimabukuro-Dias, C.K., Oliveira, C. and Foresti, F. (2004) Cytogenetic analysis of five species of the subfamily Corydoradinae (Teleostei: Siluriformes: Callichthyidae), *Genetics And Molecular Biology*, 27(4), pp. 549–554. https://doi.org/10.1590/S1415-47572004000400014.

Smith, W.L., Stern, J.H., Girard, M.G., *et al.* (2016) Evolution of Venomous Cartilaginous and Ray-Finned Fishes, *Integrative And Comparative Biology*, 56(5), pp. 950–961. https://doi.org/10.1093/icb/icw070.

Smith, W.L. and Wheeler, W.C. (2006) Venom evolution widespread in fishes: a phylogenetic road map for the bioprospecting of piscine venoms, *The Journal Of Heredity*, 97(3), pp. 206–217. https://doi.org/10.1093/jhered/esj034.

Speed, M. (2014) Mimicry. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9780470015902.a0001790.pub3.

Speed, M.P. (1993) Müllerian mimicry and the psychology of predation, *Animal Behaviour*, 45(3), pp. 571–580. https://doi.org/10.1006/anbe.1993.1067.

Stevens, M. and Merilaita, S. (2011) *Animal Camouflage*. Cambridge, England: Cambridge University Press. https://doi.org/10.1017/cbo9780511852053.

Streisinger, G.F., Singer, F., Walker, C., *et al.* (1986) Segregation Analyses and Gene-Centromere Distances in Zebrafish, *Genetics*, 112(2), pp. 311–319. https://doi.org/10.1093/genetics/112.2.311.

Stuckert, A.M.M., Saporito, R.A., Venegas, P.J., *et al.* (2014) Alkaloid defenses of co-mimics in a putative Müllerian mimetic radiation, *BMC Evolutionary Biology*, 14(1), pp. 1–8. https://doi.org/10.1186/1471-2148-14-76.

Stuckert, A.M.M., Freeborn, L., Howell, K.A., *et al.* (2023) Transcriptomic analyses during development reveal mechanisms of integument structuring and color production, *Evolutionary Ecology* . https://doi.org/10.1007/s10682-023-10256-2.

Stuckert, A.M.M., Chouteau, M., McClure, M., *et al.* (2024) The genomics of mimicry: Gene expression throughout development provides insights into convergent and divergent phenotypes in a Müllerian mimicry system, *Molecular Ecology*, 33(14). https://doi.org/10.1111/mec.17438.

Svensson, E.I. and Abbott, J. (2005) Evolutionary dynamics and population biology of a polymorphic insect: Polymorphism and evolutionary dynamics, *Journal Of Evolutionary Biology*, 18(6), pp. 1503–1514. https://doi.org/10.1111/j.1420-9101.2005.00946.x.

Symula, R., Schulte, R. and Summers, K. (2001) Molecular phylogenetic evidence for a mimetic radiation in Peruvian poison frogs supports a Müllerian mimicry hypothesis, *Proceedings Of The Royal Society Of London. Series B*, 268(1484), pp. 2415–2421. https://doi.org/10.1098/rspb.2001.1812.

Symula, R., Schulte, R. and Summers, K. (2003) Molecular systematics and phylogeography of Amazonian poison frogs of the genus Dendrobates, *Molecular Phylogenetics And Evolution*, 26(3), pp. 452–475. https://doi.org/10.1016/s1055-7903(02)00367-6.

Tapanes, E. and Rennison, D.J. (2024) The genetic basis of divergent melanic pigmentation in benthic and limnetic threespine stickleback, *Heredity*, pp. 1–9. https://doi.org/10.1038/s41437-024-00706-0.

Tarvin, R.D., Borghese, C.M., Sachs, W., *et al.* (2017) Interacting amino acid replacements allow poison frogs to evolve epibatidine resistance, *Science*, 357(6357), pp. 1261–1266. https://doi.org/10.1126/science.aan5061.

Tencatt, L.F.C., Dos Santos, S.A., Evers, H.-G., *et al.* (2021) Corydoras fulleri (Siluriformes: Callichthyidae), a new catfish species from the rio Madeira basin, Peru, *Journal Of Fish Biology*, 99(2), pp. 614–628. https://doi.org/10.1111/jfb.14750.

Tencatt, L.F.C., do Couto, O.L.P., Santos, S.A., *et al.* (2024) A new long-snouted Corydoras (Siluriformes: Callichthyidae) from the rio Xingu and rio Tapajós basins, Brazilian Amazon, *Neotropical Ichthyology: Official Journal Of The Sociedade Brasileira De Ictiologia* . https://doi.org/10.1590/1982-0224-2023-0112.

Tencatt, L.F.C. and Ohara, W.M. (2016) Two new species of Corydoras Lacépède, 1803 (Siluriformes: Callichthyidae) from the rio Madeira basin, Brazil, *Neotropical Ichthyology: Official Journal Of The Sociedade Brasileira De Ictiologia*, 14(1). https://doi.org/10.1590/1982-0224-20150063.

The Heliconius Genome Consortium (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species, *Nature*, 487(7405), pp. 94–98. https://doi.org/10.1038/nature11041.

Thomas, R.J., King, T.A., Forshaw, H.E., *et al.* (2010) The response of fish to novel prey: evidence that dietary conservatism is not restricted to birds, *Behavioral Ecology: Official Journal Of The International Society For Behavioral Ecology*, 21(4), pp. 669–675. https://doi.org/10.1093/beheco/arq037.

Twomey, E., Vestergaard, J.S., Venegas, P.J., *et al.* (2016) Mimetic divergence and the speciation continuum in the mimic poison frog Ranitomeya imitator, *The American Naturalist*, 187(2), pp. 205–224. https://doi.org/10.1086/684439.

Twomey, E., Johnson, J.D., Castroviejo-Fisher, S., *et al.* (2020) A ketocarotenoid-based colour polymorphism in the Sira poison frog Ranitomeya sirensis indicates novel gene interactions underlying aposematic signal variation, *Molecular Ecology*, 29(11), pp. 2004–2015. https://doi.org/10.1111/mec.15466.

Urban, S., Nater, A., Meyer, A., *et al.* (2021) Different sources of allelic variation drove repeated color pattern divergence in cichlid fishes, *Molecular Biology And Evolution*, 38(2), pp. 465–477. https://doi.org/10.1093/molbev/msaa237.

Van't Hof, A.E., Campagne, P., Rigden, D.J., *et al.* (2016) The industrial melanism mutation in British peppered moths is a transposable element, *Nature*, 534(7605), pp. 102–105. https://doi.org/10.1038/nature17951.

Wallace, A.R. (1877) The colors of animals and plants, *The American Naturalist*, 11(11), pp. 641–662. https://doi.org/10.1086/271979.

Wallbank, R.W.R., Baxter, S.W., Pardo-Diaz, C., *et al.* (2016) Evolutionary novelty in a butterfly wing pattern through enhancer shuffling, *PLoS Biology*, 14(1). https://doi.org/10.1371/journal.pbio.1002353.

Weinreich, D.M. and Chao, L. (2005) Rapid evolutionary escape by large populations from local fitness peaks is likely in nature, *Evolution*, 59(6), pp. 1175–1182. https://doi.org/10.1111/j.0014-3820.2005.tb01769.x.

Weir, B.S. and Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure, *Evolution*, 38(6), pp. 1358–1370. https://doi.org/10.2307/2408641.

Wellenreuther, M. and Hansson, B. (2016) Detecting polygenic evolution: Problems, pitfalls, and promises, *Trends In Genetics*, 32(3), pp. 155–164. https://doi.org/10.1016/j.tig.2015.12.004.

Wells, J.N. and Feschotte, C. (2020) A field guide to eukaryotic transposable elements, *Annual Review Of Genetics*, 54, pp. 539–561. https://doi.org/10.1146/annurev-genet-040620-022145.

Westerman, E.L., VanKuren, N.W., Massardo, D., *et al.* (2018) Aristaless controls butterfly wing color variation used in mimicry and mate choice, *Current Biology*, 28(21), pp. 3469-3474 https://doi.org/10.1016/j.cub.2018.08.051.

Willing, E.-M., Dreyer, C. and van Oosterhout, C. (2012) Estimates of genetic differentiation measured by FST do not necessarily require large sample sizes when using many SNP markers, *PloS One*, 7(8). https://doi.org/10.1371/journal.pone.0042649.

Wilson, J.S., Williams, K.A., Forister, M.L., *et al.* (2012) Repeated evolution in overlapping mimicry rings among North American velvet ants, *Nature Communications*, 3(1), pp. 1–7. https://doi.org/10.1038/ncomms2275.

Wilson, J.S., Pan, A.D., Limb, E.S., *et al.* (2018) Comparison of African and North American velvet ant mimicry complexes: Another example of Africa as the 'odd man out', *PloS One*, 13(1). https://doi.org/10.1371/journal.pone.0189482.

Wong, E.S.W., Papenfuss, A.T., Whittington, C.M., *et al.* (2012) A limited role for gene duplications in the evolution of platypus venom, *Molecular Biology And Evolution*, 29(1), pp. 167–177. https://doi.org/10.1093/molbev/msr180.

Wong, E.S.W. and Belov, K. (2012) Venom evolution through gene duplications, *Gene*, 496(1), pp. 1–7. https://doi.org/10.1016/j.gene.2012.01.009.

Wright, J.J. (2009) Diversity, phylogenetic distribution, and origins of venomous catfishes, *BMC Evolutionary Biology*, 9. https://doi.org/10.1186/1471-2148-9-282.

Wright, J.J. (2011) Conservative coevolution of müllerian mimicry in a group of rift lake catfish, *Evolution*, 65(2), pp. 395–407. https://doi.org/10.1111/j.1558-5646.2010.01149.x.

Wright, S. (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution, *Proceedings Of The XI International Congress Of Genetics*, 1, pp. 356–366.

Xie, B., Yu, H., Kerkkamp, H., *et al.* (2019) Comparative transcriptome analyses of venom glands from three scorpionfishes, *Genomics*, 111(3), pp. 231–241. https://doi.org/10.1016/j.ygeno.2018.11.012.

Xu, P., Xu, J., Liu, G., *et al.* (2019) The allotetraploid origin and asymmetrical genome evolution of the common carp Cyprinus carpio, *Nature Communications*, 10(1). https://doi.org/10.1038/s41467-019-12644-1.

Xu, S., Guo, Z., Feng, X., *et al.* (2023) Where whole-genome duplication is most beneficial: Adaptation of mangroves to a wide salinity range between land and sea, *Molecular Ecology*, 32(2), pp. 460–475. https://doi.org/10.1111/mec.16320.

Yang, J., Benyamin, B., McEvoy, B.P., *et al.* (2010) Common SNPs explain a large proportion of the heritability for human height, *Nature Genetics*, 42(7), pp. 565–569. https://doi.org/10.1038/ng.608.

Zancolli, G., Reijnders, M., Waterhouse, R.M., *et al.* (2022) Convergent evolution of venom gland transcriptomes across Metazoa, *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 119(1). https://doi.org/10.1073/pnas.2111392119.

Zhang, W., Wang, H., Brandt, D.Y.C., *et al.* (2022) The genetic architecture of phenotypic diversity in the Betta fish (*Betta splendens*), *Science Advances*, 8(38). https://doi.org/10.1126/sciadv.abm4955.

2 Genome Evolution in the *Corydoradinae*

2.1 Abstract

Understanding the drivers of rapid genome expansion, in closely related taxa provides crucial insights into rapid genome evolution, offering a unique window into the mechanisms driving speciation, phenotypic diversification and adaptation to environmental change. Genome expansion can be driven by whole genome duplications (WGD) and/or transposable element (TE) proliferation over short evolutionary timescales. Here we explore these processes in the Corydoradinae, a species rich group of Neotropical catfishes, showing extreme variation in genome size. We utilise a comparative genomics approach using the newly assembled and annotated genomes of *Hoplisoma metae* (a putative polyploid) and *Corydoras fulleri* (a putative diploid). We find substantial differences in genomic composition between the two closely related taxa, with evidence for both WGD and TE expansion in the group. Estimates of the age and origin of the WGD position it as a recent autopolyploidy event, occurring only 4.86 mya, earlier than previously thought. This represents the youngest autoploidy event identified in teleosts to date. Moreover, we find evidence for TE insertion as a driver for intron expansion in the group and highlight TE insertion being highest in the first intron of a gene model, which has regulatory functions. The evidence for both WGD and TE expansion found in this study highlights the importance of the Corydoradinae as a system in which to explore both these drivers of genome expansion.

2.2 Introduction

Genome size varies substantially within animalia, from 19Mbp in the parasitic nematode *Pratylenchus coffeae* to 130Gbp in the marbled lungfish, *Protopterus aethiopicus* (Blommaert, 2020). Genome size is correlated with some phenotypic traits, including amphipod body size, basal metabolic rate in passerine birds and mammals, and embryonic development in salamanders (Roth, Blanke and Wake, 1994; Vinogradov, 1995, 1997; Jockusch, 1997; Jeffery, Yampolsky and Gregory, 2017). However, in general, neither genome size (C-value) or the number of genes (G-value) reflect the morphological complexity of an organism (Elliott and Gregory, 2015). This paradox (the C value paradox or C value enigma) prompted investigations into the architecture of genomes to better understand their make-up (Thomas, 1971; Blommaert, 2020). The subsequent discovery of non-coding and repetitive sequences somewhat resolved this enigma, but also raised questions about the contribution of such elements to changes in genome size, as well as their roles in phenotypic diversification (Elliott and Gregory, 2015; Blommaert, 2020). The main causes of genome expansion include whole gene duplication and transposable element (TE) proliferation, but other processes are also likely to lead to increased

genome sizes and these include tandem gene duplication, insertion/deletion bias, and replication slippage (Gregory 2005). Both WGD and TE expansion mechanisms introduce additional genetic variation and cause changes in gene expression, thus may play an important role in the evolutionary history of a group (Van de Peer, Mizrachi and Marchal, 2017; Bomblies, 2023).

Gene duplications occur most dramatically in the form of whole genome duplication (WGD) events, where the number of chromosomes within an organism is doubled. These events have been suggested to play a role in phenotypic diversification and as well as in tolerance to changing environments (Meyer and Van de Peer, 2005; Mable, Alexandrou and Taylor, 2011; Lorin *et al.*, 2018). Polyploids are traditionally classified as either autopolyploid (derived from a single lineage) or allopolyploid (derived from a hybridization event) (Van de Peer, Mizrachi and Marchal, 2017; Bomblies, 2023). Both auto and allopolyploidy are associated with changes in gene expression (dosage) and the adoption of novel gene functions (neo-functionalisation), increasing the adaptability of a group or tolerance to environmental change (Ohno, 1970; Van de Peer, Mizrachi and Marchal, 2017). For example, WGDs have been linked to the increased salinity tolerance in mangroves (Ma *et al.*, 2022; Xu *et al.*, 2023), drought tolerance in tetraploid *Neobatrachus* frogs (Novikova *et al.*, 2020) and increase in pigmentation cell diversity in teleosts (Braasch, Schartl and Volff, 2007; Braasch, Volff and Schartl, 2008; Lorin *et al.*, 2018). On a macroevolutionary scale, polyploidy is thought to contribute to elevated species diversification due to the novel genotype interacting with ecological factors, although this often occurs after a temporal lag (Van de Peer, Mizrachi and Marchal, 2017). This diversification has been observed in angiosperm (flowering plant) lineages (Schranz, Mohammadin and Edger, 2012; Tank *et al.*, 2015) and within the salmonids (Robertson *et al.*, 2017; Gundappa *et al.*, 2022). In the salmonids, the lag is hypothesised to be due to a delay in the rediploidisation process, where parts of the genome still demonstrate tetrasomic inheritance prior to speciation events, known as the LORe model (Robertson *et al.*, 2017; Gundappa *et al.*, 2022).

An increase in the abundance of TEs can also result in increased genome size in concert with or be independent of WGD events. For example, the Mexican axolotl (*Ambystoma mexicanum*) has a very large genome (32 GB), which is due to an increased abundance of long terminal retroelements within both introns and intergenic regions (Nowoshilow *et al.*, 2018). Conversely, TE expansion within the maize genome occurred after a whole genome duplication event, resulting in 85% of the genome being composed of TEs (Schnable *et al.*, 2009). Traditionally, TEs are considered deleterious, but increasing evidence is demonstrating a role for them in adaptation (Schrader and Schmitz, 2019). For example, variation in melanism in the peppered moth, *Biston betularia,* is the result of TE insertion within the first intron of the gene pigment gene *cortex* (Van't Hof *et al.*, 2016). Additionally, TE

insertions have been shown to alter the expression of *Drosophilia* immunity genes (Ullastres, Merenciano and González, 2021). Transposable elements are also thought to have played a role in both the anolis lizard and cichlid fish adaptive radiations, highlighting the evolutionary significance of proliferating repeat sequences (Brawand *et al.*, 2014; Feiner, 2016).

The Neotropical subfamily *Corydoradinae* is a species rich group of freshwater catfish, with 208 recognized species (Dias *et al.*, 2024). The group has been delimited into nine mtDNA lineages or eight gDNA lineages (Alexandrou *et al,* 2011). These lineages have now been designated as seven separate genera (Dias *et al.*, 2024). Cytogenic and early molecular evidence indicates an increase in genome size within the group, specifically in the *Hoplisoma*, previously known as lineages 6 and 9 (Oliveira *et al.*, 1993; Shimabukuro-Dias *et al*., 2004; Shimabukuro-Dias, Oliveira and Foresti, 2004). Both WGD and TE proliferation were found to occur at the base of *Hoplisoma* indicating these mechanisms are responsible for genome expansion (Marburger *et al.*, 2018). Additionally, these mechanisms may have driven diversification within the group, as *Hoplisoma* are a species rich genus which has undergone rapid phenotypic diversification (Fuller and Evers, 2005; Alexandrou *et al.*, 2011). Thus, they represent a potential model system in which to study the impact of WGD and TE proliferation on speciation and adaptation (Marburger *et al.*, 2018; Bell *et al.*, 2020).

Here we aim to confirm the drivers of genome expansion in the Corydoradinae, using a comparative genome approach, utilising the genome assemblies and initial genome annotations of both *Corydoras fulleri* (a putative diploid) and *H. metae* (a putative polyploid). Using a variety of genome-wide methods, we aim to quantify the contribution of WGDs and TEs to genome expansion in the subfamily. Specifically, using a suite of methods, we test the hypothesis that *H. metae* has undergone one or more WGD events since diverging from *C. fulleri*. If WGD event(s) are confirmed, we will investigate the origin of the WGD event(s) (autopolyploidy vs allopolyploidy) and estimate an evolutionary date(s) of the event(s). Additionally, we aim to investigate additional reasons for the increase in genome size in *H. metae* vs *C. fulleri* by exploring the structure and size of genome features. Finally, we will explore the contribution of transposable element proliferation to genome expansion in the Corydoradinae through elucidation of the composition of both the *C. fulleri* and *H. metae* repeatomes and exploration of the spatial organisation of TE across the genomes.

2.3 Materials and Methods

2.3.1 Genome Assembly

Initial assemblies for both *C. fulleri* and *H. metae* were performed by Dr Ellen Bell in 2019, using the methodology outlined in Bell *et al.* (2022). Briefly, this utilised wtdbg2 (v. 2.5) (Ruan and Li, 2020), to

create an initial assembly from long read PacBio data. Long reads were then mapped to the initial assembly using minimap2 (v. 2.17) (Li, 2018). Short reads were then mapped using BWA-MEM (v. 0.7.17) (Li and Durbin, 2009) and the alignment was merged and sorted using SAMtools (v.1.10)(Danecek *et al.*, 2021). Finally, the assembly was polished using Pilon (v. 1.23) (Walker *et al.*, 2014) and Racon (v. 1.4.15)(Vaser *et al.*, 2017) using the wtdbg2-racon-pilon.pl script (Schellt, 2019; https://github.com/schellt/wtdbg2-racon-pilon).

To obtain a scaffold level assembly for *C. fulleri*, a Dovetail Omni-C library was prepared from flash frozen muscle, heart and liver tissue and sequenced by Dovetail Genomics (Scotts Valley, California, United States). The Dovetail Omni-C reads, along with our initial assembly, were run through the HiRise pipeline by Dovetail Genomics (Putnam *et al.*, 2016).

2.3.2 Genome Annotation

2.3.2.1 Transcriptomic evidence

The transcriptome evidence utilised RNAseq data on NCBI from *H. nattereri* liver (Parente *et al.*, 2017) and *H. julii* gill tissue (Sun *et al.*, 2016), as well as the newly sequenced axillary (venom) glands from *C. narcissus, C. desana, H. arcuatus* and *H. tukano* (Chapter 4). Additionally, we used previously assembled muscle transcriptomes from *H. araguaianesis, H. paleatus, Brochis haroldschultzi* and *C. maculifier* (Bell, Butler and Taylor, unpublished).

Prior to RNA extraction of axillary glands, individuals were euthanised and the gland was dissected and immediately flash frozen using liquid nitrogen. RNA was extracted using an adapted RNAeasy micro or mini kit (Qiagen, Germany) protocol with a prior Trizol step as detailed (http://www.untergasser.de/lab/protocols/rna_prep_comb_trizol_v1_0.html). A volume of 20μl at a minimum concentration of 20 ng/μl was sent to be sequenced. Quality control and library preparation for Illumina sequencing were performed at Novogene Cambridge, using NovaSeq 6000 and each sample was sequenced to 40 million reads.

Existing RNAseq data, generated from previous studies, for liver and gill tissue *H. nattereri* and *H. julli* generated in (Sun *et al.*, 2016; Parente *et al.*, 2017) were downloaded from NCBI as SRA (SRR4069944, SRR5997699) and converted to fastq using SRAtoolkit (v. 2.10.2, trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software /). These, and the axillary gland RNAseq data, were then evaluated using FastQC (v. 0.11.8) (Andrews, 2010) before trimming Illumina adapters and removing sequences with a phred score below 20 (v. 2.10) (Martin, 2011). Filtered reads were first aligned to either the *C. fulleri* (*Corydoras sp*.) or *H. metae* reference genome (*Hoplisoma* and *Brochis sp*.) using Hisat2 (v. 2.2.1) (Kim *et al.*, 2019) before being assembled using default settings

for genome guided assembly in Trinity (v. 2.11.0) (Grabherr *et al.*, 2011). Assembled transcriptomes were grouped by genus and therefore phylogenetic distance to the genome being annotated (*Hoplisoma* and *Brochis* to annotate *H. metae* and *Corydoras* to annotate *C. fulleri*). The transcriptomes were then filtered to remove redundancy using stages 1-8 of the script described in (Cerveau and Jackson, 2016), providing a single more complete transcriptome. This was achieved by first using TransDecoder (v. 5.7.0) (https://github.com/TransDecoder/TransDecoder) to detect open reading frames greater than 100 amino acids, which were then clustered using CD-HIT (v.4.8.1) (Fu *et al.*, 2012) as described in (Cerveau and Jackson, 2016). This reduced the number of transcripts whilst maintaining biological information to reduce the computational time constraint associated with genome annotation (Supplementary Table 2.1).

2.3.2.2 Annotation

Structural annotation was performed with MAKER2 (v.3.01.04) (Chen *et al.*, 2021) using the methodology outlined in Card *et al.* (2019). Genomes were annotated in an iterative process, using transcriptome and proteomic data along with the *ab initio* tools SNAP (v.2006-07-28) (Korf, 2004) and AUGUSTUS (v. 3.3.3)(Stanke *et al.*, 2006; Hoff and Stanke, 2019), which was trained using the results from previous iterations. Protein evidence was obtained from the *Danio rerio* (GRCz11) and *Ictalurus punctatus* (Coco_2.0) genome annotations. The genome was masked using Repeatmasker (v. 4.1.2) (Smit, Hubley and Green, 2013) within MAKER2 using a species specific Corydoradinae library, generated following the method outlined in Bell *et al.* (2022). Functional annotation was obtained via blastp (v. 2.10.1+) (Camacho *et al.*, 2009) of the annotated against all *D. rerio* proteins obtained from UniProt (Chen *et al.*, 2021) with default parameters.

2.3.3 Descriptive statistics

The quality of the genome assembly was assessed using BUSCO (v. 5.3.2) (Simão *et al.*, 2015) with "-m genome" option against the Actinopterygii database. The annotation quality was then assessed using BUSCO with the "-m protein" flag run against the protein output of maker. Annotation statistics were obtained using the AGAT function agat_sp_statistics.pl (v.1.0.0) (Dainat, 2020). Statistical support for the differences in gene length between *H. metae* and *C. fulleri* were obtained using Wilcoxon rank test via the compare_means() function from the ggpubr R package (v.0.6.0)(Kassambara, 2023), in R (v.4.2.3)(R Core Team, 2023). The quality of the assemblies and structural errors were detected using Inspector (v.1.3) (Chen *et al.*, 2021).

To detect syntenic regions within our genomes, we used MCScanX (v.1) (Wang *et al.*, 2012). CDS sequences were extracted from the annotation using the tool agat_sp_extract_sequences.pl, from

AGAT. To obtain the BLAST file input, we ran a self v self blastp of the protein sequences, using default parameters (Camacho *et al.*, 2009). We then ran MCScanX on the directory containing our blastp output to obtain the percentage of the genome that was considered colinear.

2.3.4 Intron, Intergenic and Exon Length

To investigate differences in the length of gene and genomic features, introns were first added to the annotation files using the AGAT tool agat_sp_add_introns.pl. The annotation files could then be filtered to retain either intron or exon regions, which were already annotated. Intergenic regions were identified from the annotation files by first extracting the position of genes. Overlapping genes were collapsed by taking the minimum start position, and the maximum end position to reduce the computational time. Subsequently, the genic positions were converted into a GenomicRanges object using the makeGRangesFromDataFrame function (v.1.50.2) (Lawrence *et al.*, 2013). The regions of the genome that were not covered by a gene were then extracted using the gaps() function from the Genomic Ranges package. The regions with the top and bottom 1% of length were then filtered out.

To estimate whether the observed differences in feature size were statistically significant, we used the lm() function in base R with the length as the effect variable and species as the fixed factor. When analysing intron length, we also included the "peak" the intron was present in (peak 1 < 250, peak 2 >250) as a fixed factor, and the interaction of peak and species. The suitability of our models was checked using the DHARMa package (v. 0.4.6) (Hartig, 2022) and post-hoc response was explored using emmeans(), from the emmeans package (v.1.10.3)(Lenth, 2024).

2.3.5 Evidence for WGD

2.3.5.1 Synonymous substitution rate estimates

The CDS and amino acid models, generated during the genome annotation process, were filtered to retain only models larger than 300 bps and 100 amino acids or smaller than 9000 base pairs and 3000 amino acids, to reduce computational load. Additionally, we obtained the CDS and amino acid sequences of *Cyprinus carpio* (ASM1834038v1, NCBI) and filtered the sequences as above. Including *C. carpio* allowed comparisons to be made between a known polyploid and *H. metae*. To calculate Ks, we used WGDdetector (v.1.1) (Yang *et al.*, 2019) which identifies paralogs and calculates Ks between or among them. A caveat of using Ks to detect WGD events is the presence synonymous site saturation, which can cause spurious peaks (Vanneste, Van de Peer and Maere, 2013). WGDdetector corrects for this using hierarchical clustering and filtering for genes with Ks values less than 5.0 (Yang *et al.*, 2019).

To detect peaks in our data, we used both Gaussian mixture models (GMM) and kernel density estimates (KDE). To reduce the risk of boundary effects when estimating the GMM we first reflected our data and then modelled our data for between one to five clusters, using mclust() (v. 6.0.1)(Scrucca *et al.*, 2023). We selected the model with the lowest Bayesian Information Criterion (BIC) (Supplementary Figure 2.1). We calculated KDE based on the raw Ks values, as KDE is less sensitive to boundary effects. Timing of the *Hoplisoma* WGD events were estimated using the KDE Ks estimates and calibrated using the timing of the teleost specific whole genome (FSGD) duplication.

2.3.5.2 Ploidy estimation using allelic depth

To calculate allelic depth, the Illumina sequences used in genome assembly were aligned to the *C. fulleri* reference genome. Adapters were first identified and removed using Picard (v. 2.20.2-11) (https://broadinstitute.github.io/picard/) MarkIlluminaAdapters and SamToFastq (Van der Auwera *et al.*, 2013). Filtered FASTQ files were then aligned to the *C. fulleri* reference genome using BWA MEM (v. 0.7.17) (Li and Durbin, 2009) and merged with unmapped bam files to retain all relevant sample and quality information using Picard MergeBamAlignment. Individuals sequenced over multiple lanes were merged using SAMtools merge. Optical and sequencing duplicates were removed using Picard MarkDuplicates and overlapping reads are clipped using BamUtils (v. 1.0.15) clipOverlap (Jun *et al.*, 2015). Regions around indels were identified and realigned using the Genome Analysis ToolKit (GATK3, v. 3.8-1-0) (Van der Auwera *et al.*, 2013) RealignerTargetCreater and IndelRealigner. To estimate the allelic depth, we used the software nQuire (v.1), which utilises Gaussian mixture models and maximum likelihood to estimate ploidy level (Weiß *et al.*, 2018). This method relies on observed allele frequencies per site between diploids, triploids and tetraploid levels, where diploids are expected to have a bi-allelic SNP ratio of 0.5, triploids (0.33, 0.66, 0.33), tetraploids (0.25, 0.5, 0.25) across all sequence reads. A binary file was generated from our aligned reads (nQuire create), followed by denoising to remove low quality SNPs and misalignments (nQuire denoise). Finally, to estimate the density of read depth distribution we used the histo command to generate the frequency of reads at each SNP.

2.3.5.3 Smudgeplot and Genomescope

The density and complexity of kmers, substrings of a longer sequence with a fixed length, can be used to make assembly independent inferences about a genome structure (Ranallo-Benavidez, Jaron and Schatz, 2020). To identify kmers within our genomes, we utilised the Illumina raw reads used for genome assembly, with adapters removed using Picard MarkIlluminaAdapters and SamToFastq. We then counted the kmers in each genome with FastK(v. 1.0.0), (https://github.com/thegenemyers/FASTK) using a kmer size of 21. We visualised the kmer distribution

using GenomeScope (v.2.0.1) (Ranallo-Benavidez, Jaron and Schatz, 2020) with default parameters and kmer size of 21. Smudgeplot makes inferences about ploidy using kmers that differ by one nucleotide (het-mers) (Ranallo-Benavidez, Jaron and Schatz, 2020). We identified het-mers using the Smudgeplot (v. 0.3.0) (Ranallo-Benavidez, Jaron and Schatz, 2020) function hetkmers before visualizing using the plot function with default parameters and kmer size of 21.

2.3.5.4 Gene family Identification

Duplications in single gene families have provided some of the earliest evidence for WGD events, particularly the Hox genes, where duplicates appear to be retained for long evolutionary time periods post-duplication (Amores *et al.*, 1998). Homeobox domains were identified using HbxFinder (v.1), using default parameters for vertebrate genomes (--group vertebrate) (Mulhair *et al.*, 2023). This collection of scripts first performs a broad tblastn, utilising the HomeoDB (Zhong and Holland, 2011), followed by a reciprocal blastx and then MMseq2 (v. bad16c765aac60d84a8fde3548adbb06b34980bd) (Steinegger and Söding, 2017) to better detect the highly conserved hox regions. The resulting hox domains were extracted along with 5000 bp either side using SAMtools faidx (Danecek *et al.*, 2021). *Danio rerio* (GRCz11) homeobox proteins were then aligned to candidate homeobox regions using Exonerate (v.2.4.0) (Slater and Birney, 2005) with the protein2genome model to identify the entire hox genes and assign a cluster to a specific scaffold or contig. Clusters were assigned by filtering hits to be greater than 100 amino acids in length with an alignment score of above 200 and then counting the number of genes from each hox cluster on each chromosome.

The Toll-like receptors were chosen as an additional family for investigating evidence of duplications as they are less conserved than the homeobox genes, and subsequently duplications are less likely to be collapsed/amalgamated in the genome assembly process but retain sufficient resolution to identify WGD events if they have occurred. Additionally, there is some evidence to show gene duplications are present/have been retained within the family (Bell *et al.*, 2020). Toll-like receptor proteins were extracted from the genome annotations of *Danio rerio* (GRCz11) and *Ictalurus punctatus* (Coco_2.0) and these were aligned to the Corydoradinae genome assemblies using Exonerate with the protein2genome model. To ensure that additional TLR genes were identified we also aligned teleost TLR amino acid sequences obtained from UniProt (UniProt Consortium, 2021) to the genome assemblies. Finally, to verify our TLR candidates we performed a blastx with the Corydoradinae TLR nucleotides against the downloaded teleost TLR amino acid database.

2.3.6 Transposable Element Proliferation

Transposable Elements were identified using RepeatMasker (v.411) (Smit, Hubley and Green, 2013) using the species-specific transposable element libraries. This output was then parsed using the RM_TRIPS script (https://github.com/clbutler/RM_TRIPS)(Chen *et al.*, 2021), which removes simple repeats, merges repeats with the same annotation within 80bp of each and removes repeats shorter than 80bp. The TE annotation was then used to calculate the abundance and lengths of each repeat family within each genome.

We also investigated TE insertion position within each genome, within the introns, intergenic regions, exons as well as 5000bp upstream and 5000bp downstream of exons, as these are likely to have regulatory functions (Riethoven, 2010). Genome gff files containing structural annotation information were imported into R, using the read.gff function from the ape package (v. 5.7.1) (Paradis and Schliep, 2019) and were filtered to retain the regions of interest (introns, intergenic regions, exons, 5000bp upstream and 5000bp downstream). Both the feature position and the transposable element positions were converted into GenomicRanges objects using makeGRangesFromDataFrame and overlaps were quantified using the countOverlaps() function, both from the GenomicRanges R package (v. 1.50.2)(Wang *et al.*, 2012). To investigate the probability of TE presence or absence in each feature we used a generalized linear model. Using a binary dependent variable, such as presence or absence prevents biases caused by zero-inflated data, estimates of TE abundance being heavily inflated by length of feature (e.g. longer feature more TE insertions) or by the difficulty in estimating insertions within insertions affecting raw TE counts. We modelled the effect of species and feature on TE presence or absence, using the base R glm() function. The residuals were checked using the DHARMa package. The significance of the model was explored using the anova() base R function, with a Chi-squared test. The post hoc response was explored using emmeans(), from the emmeans package (v.1.10.3)(Lenth, 2024). We performed a similar analysis to explore TE insertion within the first intron compared to the other introns within a gene model. We modelled the effect of Corydoradinae species and intron position (first or other) the presence or absence of TE within an intron.

To explore the differences in TE composition of introns between the two species, we calculated the percentage of the intron consisting of TEs. Introns without TE insertions were temporarily removed to improve computational speed. For the remaining introns with TE insertions, the length of each insertion within the intron was calculated using the intersect() function from the GenomicRanges package. The introns without insertions were then added and the length of insertion was set to 0. Finally, the total percentage coverage of the intron by TE insertions was calculated. This reduces bias

where insertions are inserted within each other, which can be difficult to calculate resulting in overrepresenting the impact of the TE on the nucleotide structure.

2.4 Results

2.4.1 Genome annotation

Here we present the first scaffold level assembly of the *C. fulleri* as well as genome annotations of *C. fulleri* and *H. metae*. This represents the first genome annotations from the family Callichthyidae and therefore a valuable contribution to the genomics of the family. The *C. fulleri* assembly is close to complete with a BUSCO score of 95.2% (complete gene models, vertebrate_odb10 database, Supplementary Table 2.2) and a size of 699,783,386 base pairs (bps) with a N50 of 25,128,884 bps (Table 2.1). The N50 score indicates that half of the assembled genome is in scaffolds at least as long as ~ 25 mega bps (mbps). This is half the size of the longest scaffold, at 53,753,736 bps. This is a comparable in quality to the Channel catfish and blue catfish assemblies (Waldbieser *et al.*, 2023). By comparison, the *H. metae* assembly is much larger at 3,813,684,144 bps, making it the largest catfish genome assembled to date (Gong *et al.*, 2018, 2023; Kim *et al.*, 2018; Jiang *et al.*, 2019; Ozerov *et al.*, 2020; He *et al.*, 2021; Kushwaha *et al.*, 2021; Shao *et al.*, 2021; W. Chen *et al.*, 2021; Lemopoulos and Montoya-Burgos, 2022; Xu *et al.*, 2022; Tian *et al.*, 2023; Waldbieser *et al.*, 2023; Ye *et al.*, 2023). The N50 score for *H. metae* is 254,112 bp, which is low compared the contig N50 of published catfish genomes and is far smaller than the longest contig at 3,153,761 bps. Additionally, the BUSCO score of complete gene models 85.7% (vertebrata_odb10 database, Supplementary Table 2.2) and indicates the assembly is fragmented and partially incomplete. Despite being fragmented, the assembly had a similar number of overall structural errors (expansion, collapse, haplotype switch and inversion) to *C. fulleri*. Expansion occurs when regions of the genome are artificially elongated due to misassembly. Collapse is the converse phenomenon, where similar regions are reduced into a single consensus site. There are a similar number of occurrences of haplotype switching between *C. fulleri* and *H. metae*. Haplotype switching occurs at heterozygous structural variant break points, when there are multiple haplotypes. Failure to reconstruct either haplotype creates a mosaic sequences in-between both

Table 2.1 Corydoradinae genome assembly and annotation summary statistics.

	Corydoras fulleri	***Hoplisoma metae***
Assembly Statistics		
Assembly Size (bp)	699783386	3813684144
Longest scaffold/contig	53753736	3153761
N50	25128884	254112
Number of contigs	856	22306
Number of contigs > 10000 bp	603	22306
Number of contigs > 1000000 bp	29	186
Expansion	604	1382
Collapse	406	329
Haplotype Switch	1178	1124
Inversion	3	24
Small scale assembly error /per Mbp	156.16	1249.51
Quality value	29.76	27.96
GC %	42.12	39.69
Annotation Statistics		
Repeat Element Coverage	41.00	74.01
Number of genes	20095	30761
Synteny	5.10	0.15
Number of mrnas with utr both sides	1847	1248
Number of mrnas with at least one utr	8023	9847
Number of exons	204444	178592
Number of five prime utr	5527	8163
Number of three prime utr	4343	2932
Number of exons in cds	201226	175713
Number of introns in cds	181131	144952
Number gene overlapping	236	1401
Number of single exon gene	28	2670
Number of single exons mrna	28	2670
mean mrnas per gene	10	10
mean exons per mrna	102	58
mean gene length	11508	25285
mean cds length	1635	1008
mean exon length	170	171
mean five prime utr length	117	72
mean three prime utr length	352	154
mean intron in cds length	1061	5044

haplotypes, which will appear as either a collapse or expansion depending on the haplotype (Chen *et al.*, 2021). These issues can be particularly pertinent in highly repetitive genomes. The small-scale errors per Mbp (expansion, collapse, haplotype switch and inversions < 50 bp) is far higher in *H. metae* compared to *C. fulleri*, indicating additional differences in genome quality (Table 2.1).

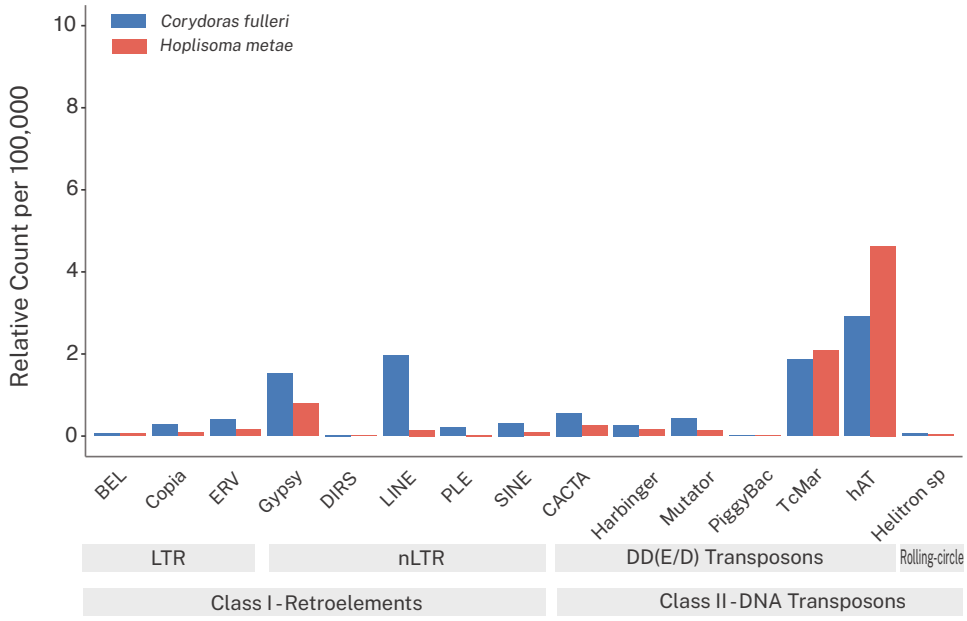


Figure 2.1 The number of each transposable element subclass (per 100,000), relative to genome size of the Corydoradinae species *Corydoras fulleri* (putative diploid) and *Hoplisoma metae* (putative polyploid). The elements are ordered into two classes, Retroelements and DNA transposons, and subsequent sub-class. Blue bars denote *C. fulleri* counts whilst red denotes *H. metae*.

H. metae has the highest coverage of repeat regions found in a catfish genome (Table 2.1, Figure 2.1) at 74%. Conversely, *C. fulleri* has repeat coverage of 41%, which is in keeping with other catfish genomes of a similar size. Relative to the genome size, we identified an expansion in Class II (DNA transposons) in *H. metae* compared to *C. fulleri*. Specifically, an expansion of the hAT elements (Figure 2.1).

Assembly quality, particularly fragmentation, has downstream impacts on the genome annotation process. In *H. metae,* 30,761 gene models were identified, and 20,095 were identified in *C. fulleri*. This is as anticipated for *C. fulleri*, compared to catfish genomes of a similar size (Gong *et al.*, 2023; Waldbieser *et al.*, 2023). The number of gene models identified in *H. metae* was the highest found in any catfish.

2.4.2 Evidence for WGD

2.4.2.1 Ks distribution plots

The estimated number of synonymous substitutions per synonymous site, Ks, is often used to determine the presence of a whole genome duplication. The divergence of two coding sequences serves as a proxy for time, under the assumption of neutral evolution at synonymous sites, thus Ks can be used to estimate the relative time of a duplication event between paralogs (Vanneste, Van de Peer and Maere, 2013; Zwaenepoel and Van de Peer, 2019). Small scale duplications result in an exponential peak at 0, due to a stochastic birth and death process, whilst additional peaks (Ks > 0) can be indicative of WGD events (Tiley, Barker and Burleigh, 2018). Common carp (*C. carpio*) has undergone a whole genome duplication 12-20 mya and shows a clear bimodal distribution of Ks (Figure 2.2A) (Ma *et al.*, 2014; Xu *et al.*, 2019). The initial peak (Ks: KDE= 0.17, GMM=0.18) reflects the carp specific WGD event, whilst the second peak (Ks: KDE=1.78, GMM=2.03) reflects the teleost specific WGD (FSGD) occurring 235 mya (Davesne *et al.*, 2021). The FSGD is also present in *H. metae* (Ks: KDE=1.92, GMM=1.99) and *C. fulleri* (Ks: KDE=1.33, GMM=1.40) (Figure 2.2B-C). Differences in

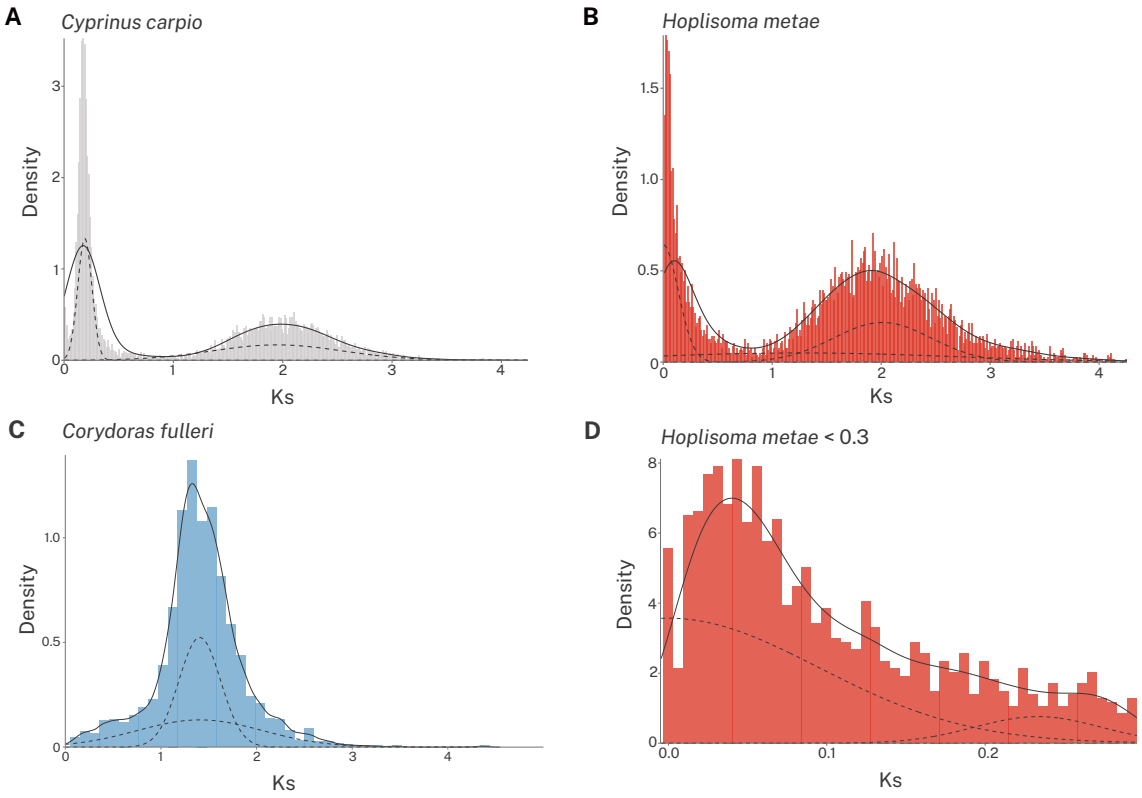


Figure 2.2 The density distribution of the synonymous substitutions per synonymous site (Ks) in the **A** polyploid *Cyprinus carpio,* **B** putative *Hoplisoma metae,* **C** putative diploid *Corydoras fulleri,* **D** Magnified insert of *H. metae* when Ks is less than 0.03. Dashed lines represent Gaussian mixture models and the solid lines represent kernel density estimates.

evolutionary rates can cause peaks at differing Ks values to represent the same WGD event in different species (Yang *et al.*, 2020; Sun *et al.*, 2022). An initial peak (Ks: KDE=0.04, GMM=0.001) in *H. metae* indicates the presence of a lineage specific whole genome duplication (Figure 2.2B). Whilst the GMM model predicts this peak to be very close to zero, it is distinct from the peak at zero characterised by small scale duplications (Figure 2.2D). Using the KDE values for Ks and the current estimates for time of the FSGD place the *Hoplisoma* WGD at 4.86 mya.

2.4.2.2 Kmer distribution

Kmer distributions are an assembly independent method for inferring ploidy level as well as the origin of polyploidy (Ranallo-Benavidez, Jaron and Schatz, 2020). The presence of the two peaks indicate *C. fulleri* is a diploid (Figure 2.3A). The kmer distribution of *C. fulleri* indicates a relatively low heterozygosity, as reflected by the lower initial peak. This is concordant with the reported heterozygosity (AB) of 0.516%. The kmer distribution of *H. metae* resembles that of simulated and empirical tetraploids, due to the presence of four peaks (Figure 2.3B). This is indicated by the peaks at 35, 60, 95, 125x coverage (indicated by the dashed lines). The peaks after this are due to repeated sequences, as indicated by the reduction of the model based on unique sequences to zero (Figure 2.3B). The estimated heterozygosity for the four possible tetraploid heterozygote states in *H. metae* are: AAAB (2.43%), AABB (1.1%), AABb (0.155%), and AaBb (0.121%). The heterozygote state AAAB is greater than AABB, indicating that *H. metae* is likely an autotetraploid (Ranallo-Benavidez, Jaron and Schatz, 2020).

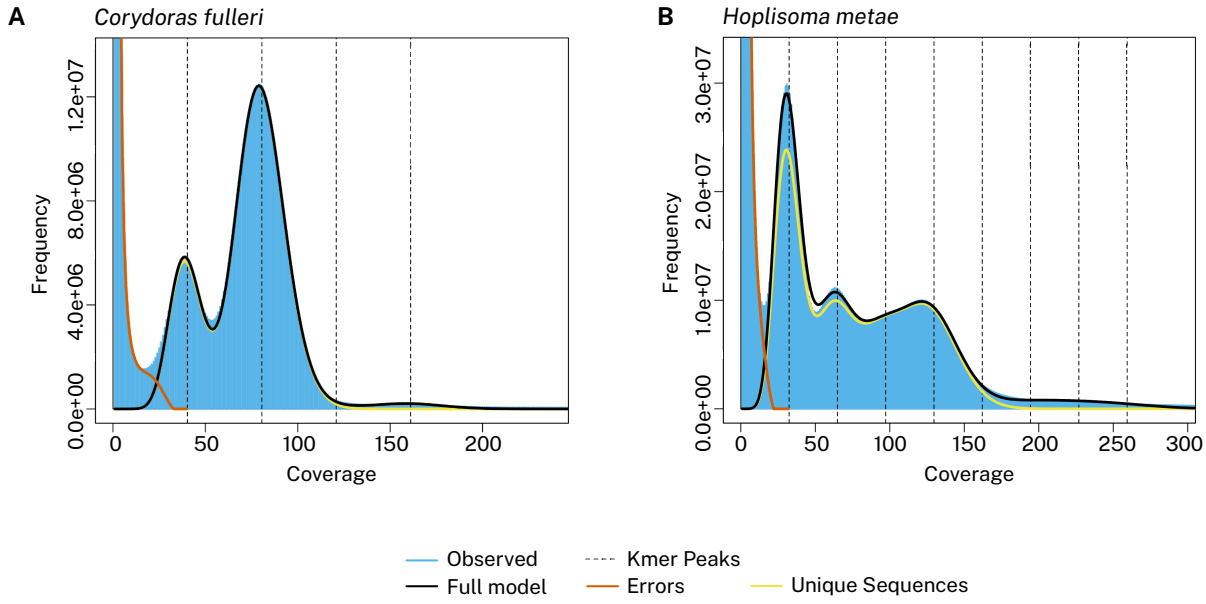


Figure 2.3 The kmer distribution, as estimated by GenomeScope, of **A** the putative diploid *Corydoras fulleri* **B** the putative tetraploid *Hoplisoma metae.*

The het-mer distributions provided by Smudgeplot also indicated *C. fulleri* is a diploid. There is a clear peak at the total density of 2N and a normalized ratio of 1/2, which is indicative of the AB heterozygosity state (Figure 2.4A)(Ranallo-Benavidez, Jaron and Schatz, 2020). The het-mer distribution of *H. metae* is concordant with the kmer distribution, with peaks at a total density of 4n and normalized minor kmer coverage of 1/4 and 1/2. (Figure 2.4B). As with the estimated frequency of heterozygote states estimated from the whole kmer distribution, the AAAB state has a higher density than AABB, which is indicative of autotetraploidy.

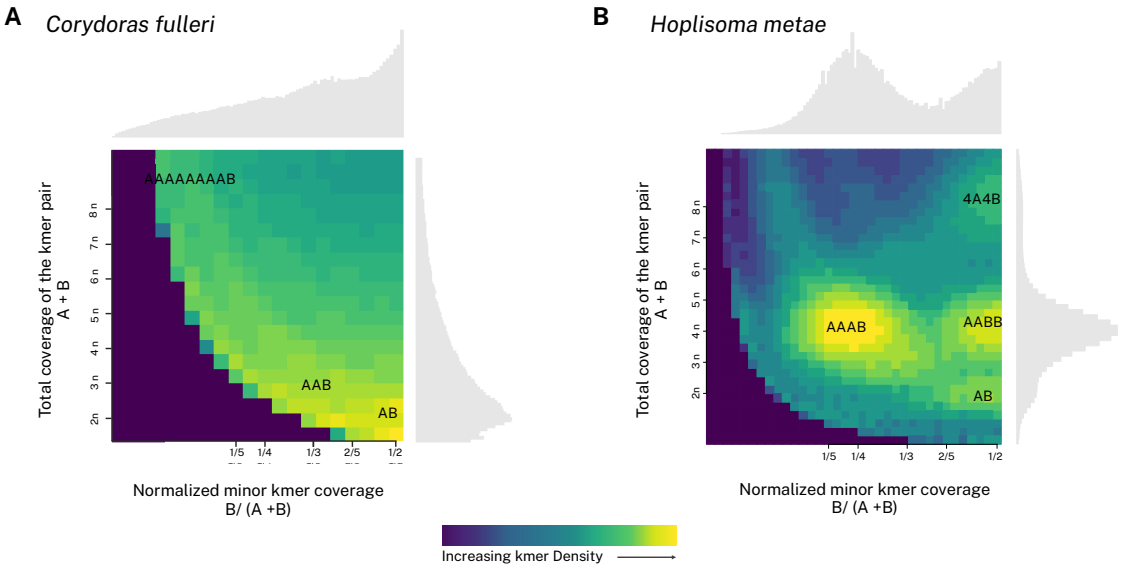


Figure 2.4 The density of kmers which vary by a single nucleotide (het-mers) as estimated by Smudgeplot in the **A** diploid *Corydoras fulleri* **B** putative tetraploid *Hoplisoma metae*.

2.4.2.3 Allele depth

For additional evidence of a WGD and ploidy level we explored the distribution of base frequences via nQuire, which uses a Gaussian mixture model and maximum likelihood model to assess which ploidy distribution best represents empirical data. Base frequences estimated from the *C. fulleri* genome suggested a diploid model (Figure 2.5A). This was confirmed by the likelihood model estimation shows the lowest log-likelihood distance to the diploid model (Figure 2.5B). The *H. metae* base frequency distribution (Figure 2.5A) was more similar to the expected tetraploid distribution, which is confirmed by the distance from the log likelihood model (Figure 2.5C). In tetraploids, the three heterozygous genotypic states are AAAa, AAaa and Aaaa, with relative base frequencies of 0.25, 0.5 and 0.25

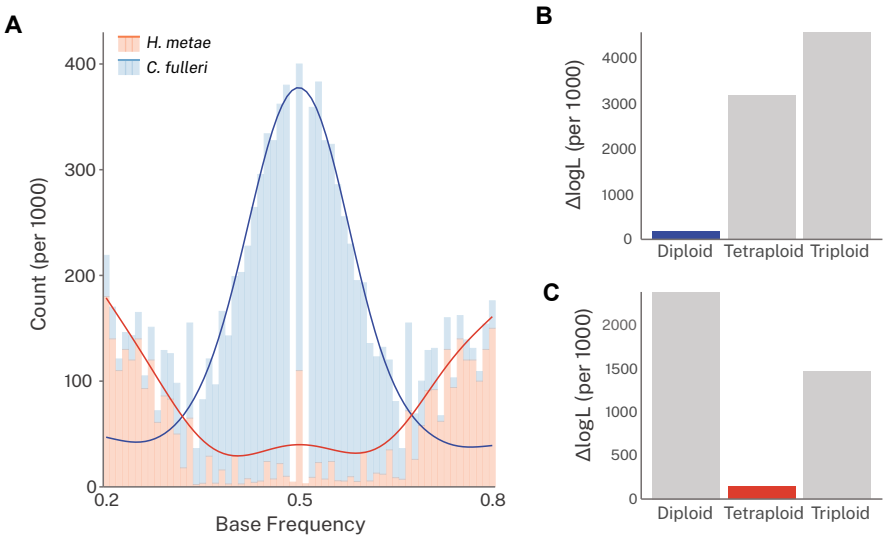


Figure 2.5A The distribution of base frequencies, as estimated by nQuire. The distribution of *Corydoras fulleri*, in blue, is indicative of a diploid genome. The distribution of *Hoplisoma metae* is indicative of an autotetraploid genome. **B** The delta log likelihood, depicting which ploidy model is closest to the observed allele frequency distribution, showing that *C. fulleri* matches the diploid model. **C.** The delta log likelihood suggesting that *H. metae* is tetraploid.

respectively (Scott, Van de Velde and Novikova, 2023). An allopolyploid or autopolyploid that has undergone rediploidization, will have an excess of AAaa sites and this will be reflected in the allelic read depth distribution (Scott, Van de Velde and Novikova, 2023). The distribution for *H. metae* fits the expected distribution for an autotetraploid.

2.4.2.4 Gene family duplications

The Hox clusters identified in *C. fulleri* and *H. metae* were largely concordant with findings in other catfishes (*P. hypophthalmus* and *I. punctatus)* (Figure 2.6). As in *D. rerio*, *I. punctatus* and *P. hypophthalmus,* the two Corydoradinae species have lost the HoxD duplicate (HoxDb) originating from the teleost specific whole genome duplication event (Kurosawa *et al.*, 2006). Additionally, the genes HoxA2a, HoxA7a, HoxA10a, HoxC8b and HoxC10b were absent in the Corydoradinae, in concordance with other teleost fishes (Kuraku and Meyer, 2009). In *D. rerio* these became pseudogenes but appear to have been lost in catfish species (Ma *et al.*, 2014). In contrast to *I. punctatus* and *P. hypophthalmus*, the Corydoradinae species have lost the Hox genes HoxC1a and HoxB2b. Some additional lineage specific loss was evident in *H. metae* compared to *C. fulleri*, including the HoxB8b, HoxA9a, HoxB9a. The gene HoxA11a is present in *H. metae* but absent in other catfish species. This may represent a

misannotation. Whilst HoxA13a was not present on ctg9817 with rest of the HoxAa cluster, it was identified on ctg222. This could represent either a gene duplication or fragmentation of a cluster. To determine either scenario would require a more contiguous assembly. An additional HoxB13a gene was identified in *H. metae*. This was not identified with any other genes representing a cluster but shows high affinity to HoxB13a, with an Exonerate alignment score of 1080. This is higher than the mean score for *H. metae* Homeobox genes, within clusters. As above, this could represent a cluster that has become fragmented or the remains of a duplicated cluster. Due to the conserved nature of the Homeobox domain, it is possible that the Hox genes may be prone to genome collapse assembly errors. We explored this by identifying structural errors within this region using the Inspector output. We did not identify any small errors across the Hox regions but did identify larger structural errors including haplotype switches and collapses on contigs/scaffolds containing Hox clusters. However, these were identified in greater quantity in *C. fulleri*, suggesting that perhaps this is not the cause of the loss of additional hox clusters in *H. metae*. However, Inspector is intended for diploid genomes and there is currently no polyploid alternative.

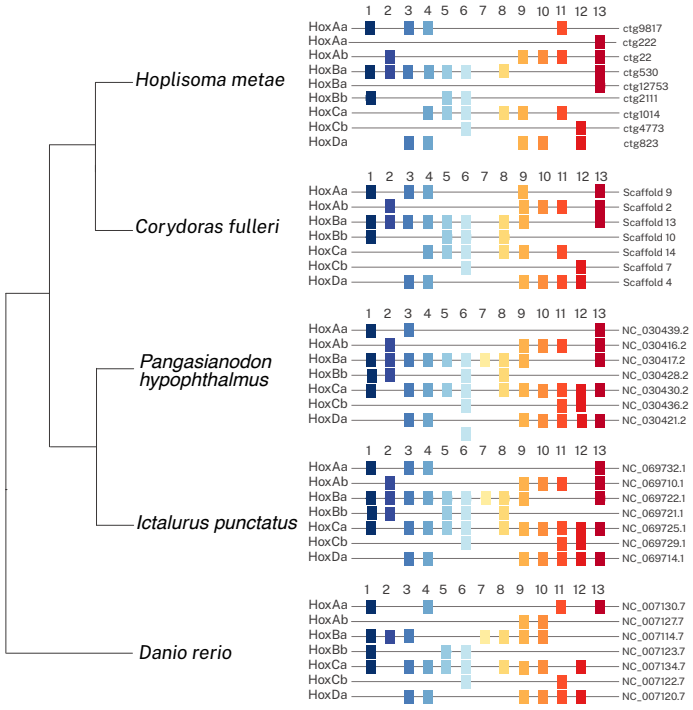


Figure 2.6 Schematic showing the presence and location Hox genes across the *Hoplisoma metae*, *Corydoras fulleri*, *Pangasianodon hypophthalmus*, *Ictalurus punctatus* and *Danio rerio* genomes.

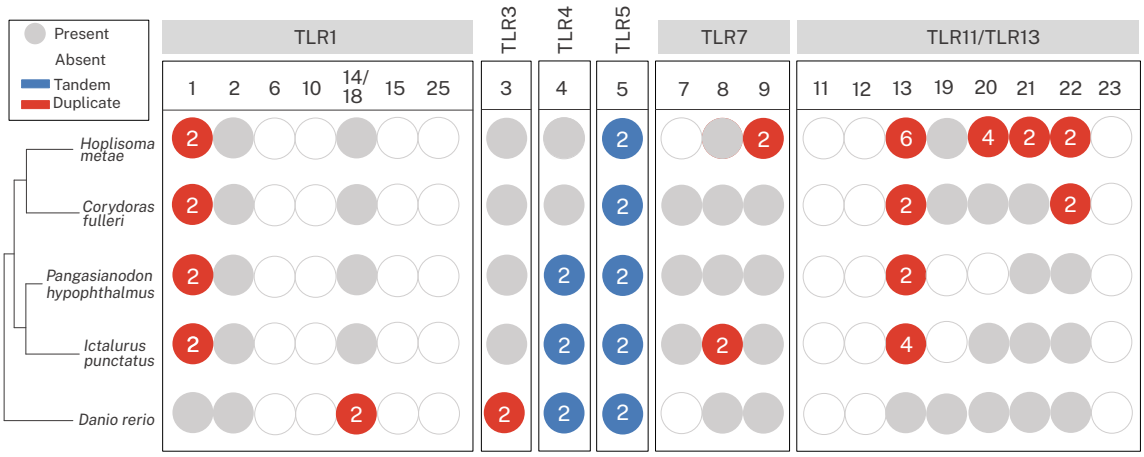


Figure 2.7 Schematic showing the presence and absence of the Toll-like receptor genes in *Hoplisoma metae*, *Corydoras fulleri*, *Pangasianodon hypophthalmus*, *Ictalurus punctatus* and *Danio rerio* genomes. The TLR genes are organised by family in grey and the specific gene number is given below. The numbers within the circles represent the number of gene copies, whilst the colour denotes the type of duplication, blue – tandem duplication and red – duplication on different ctgs/scaffolds/chromosomes.

To investigate whether there was evidence for WGD in a less conserved family, we explored potential duplications in the Toll-like receptors. As with the Homeobox genes, we saw similarities in the presence and absence of genes to those in *P. hypopthalmus* and *I. punctatus*. For example, whilst TLR8 is lost in *D. rerio*, it is retained in the catfish. Additionally, in the catfish, TLR5 is tandemly duplicated (Figure 2.7). Conversely, TLR19 is present in the Corydoradinae and *D. rerio*, but absent in the two additional catfish genomes. In *H. metae* we could not identify TLR7. Additionally, we identified two TLR1 genes and only one TLR2 gene. This is in contrast to Bell *et al.* (2020), who found signatures of gene duplication in both TLR1 and TLR2 in *H. araguaiaensis*, a sister species to *H. metae*. Additionally, we did identify TLR1 in *C. fulleri*, which was also duplicated, as in other catfish species. Whilst previous work on *I. punctatus* and *P. hypophthalamus* has not explored multiple TLR1 copies, an additional copy is present in the genome. There is some evidence for gene duplication in *H. metae*, however. The genes TLR9, TLR13, TLR20 and TLR21 have been duplicated in *H. metae* only (Figure 2.7). In TLR22, there has been a duplication in both *C. fulleri* and *H. metae* (Figure 2.7). Whilst we would expect another two copies of TLR22 in *H. metae,* this does indicate a Corydoradinae specific gene duplication of TLR22.

2.4.3 Genome feature length

As well as whole genome duplication, we explored other potential reasons for genome size shifts including the comparative lengths of genes as well as introns, exons and intergenic regions. The average gene size in *H. metae* was 25285 bp, which was two-fold greater than the gene size in *C. fulleri* ($p < 0.001$). Whilst exon size was very similar between the species at 170 bp and 178bp for *C. fulleri* and *H. metae*, respectively (Table 2.1, Figure 2.8B). The difference was statistically significant ($p < 0.001$) with *H. metae* significantly larger than *C. fulleri*. However, the difference between the estimated marginal means was small, 4.85 in *C. fulleri* and 4.86 in *H. metae* (Supplementary Table 2.3). When the log distribution of intron size was visualised, a bimodal peak, which is characteristic of teleost intron sizes was apparent in all teleost species (Figure 2.8A)(Jakt, Dubin and Johansen, 2022). This was less pronounced in *Protopterus annectens* (African Lungfish), which diverged from teleosts 460 mya (Amores *et al.*, 1998; Jaillon *et al.*, 2004; Kurosawa *et al.*, 2006; Hoegg *et al.*, 2007; Zou *et al.*, 2007; Kim *et al.*, 2018; Ozernyuk and Schepetov, 2022). The first intron peak was significantly longer in *C. fulleri* than *H. metae* ($p < 0.001$), although the difference between the estimated marginal means was very small (4.762 in *C. fulleri*, and 4.709 *in H. metae,* Supplementary Table 2.3). Whilst the first peak is generally conserved in size, *H. metae* had significantly longer introns in peak 2 than *C. fulleri* ($p < 0.001$, Supplementary Table 2.3). Additionally, we identified a significant increase in the size of the intergenic regions in *H. metae* in comparison to *C. fulleri* ($p < 0.0001$)(Figure 2.8C).

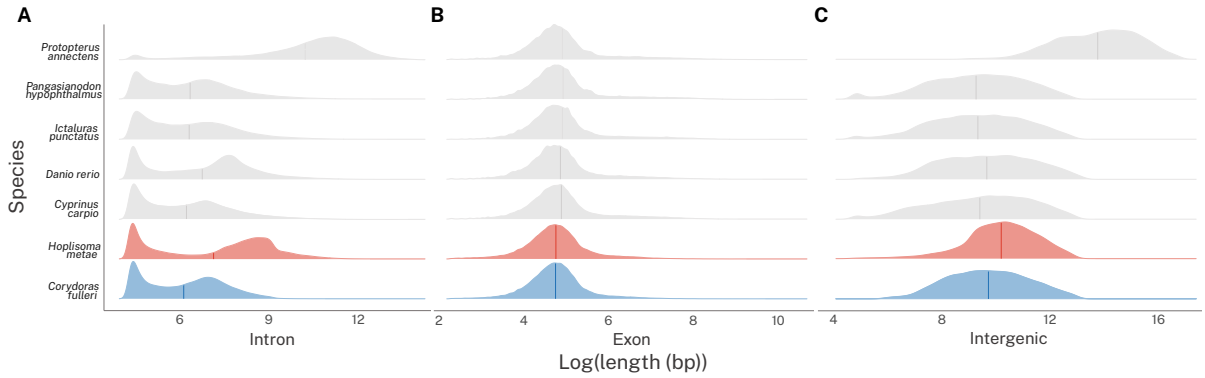


Figure 2.8 Distribution of genomic feature length (log transformed) in base pairs of **A** introns, **B** exons, **C** intergenic regions in *Protopterus anectens*, *Pangasianodon hypothalamus*, *Ictalurus punctatus*, *Danio rerio*, *Cyrpinus carpio*, *Hoplisoma metae* and *Corydoras fulleri*.

Table 2.2 Results of the generalized linear mixed model exploring the effect of genomic feature and Corydoradinae species on transposable element presence.

	df	Deviance	Resid. df	Resid. Dev	p.value
NULL			864656	1102322.00	
Species	1	9772.78	864655	1092548.81	<.0001 ***
Feature	4	443181.00	864651	649367.82	<.0001 ***
Species * Feature	4	9181.95	864647	640185.87	<.0001 ***

To explore the potential underlying causes of the expansion of specific genomic regions in *H. metae* compared to *C. fulleri*, we investigated transposable element insertion within different features of the genome. We fitted a Generalized Linear Mixed Model (GLMM) to investigate the effect of species and feature on the presence or absence of a transposable element. The results (summarized in Table 2.2) showed the effect of species (X^2 = 9773, df=1, p < 0.0001), feature (X^2= 443180, df=4, p < 0.0001)

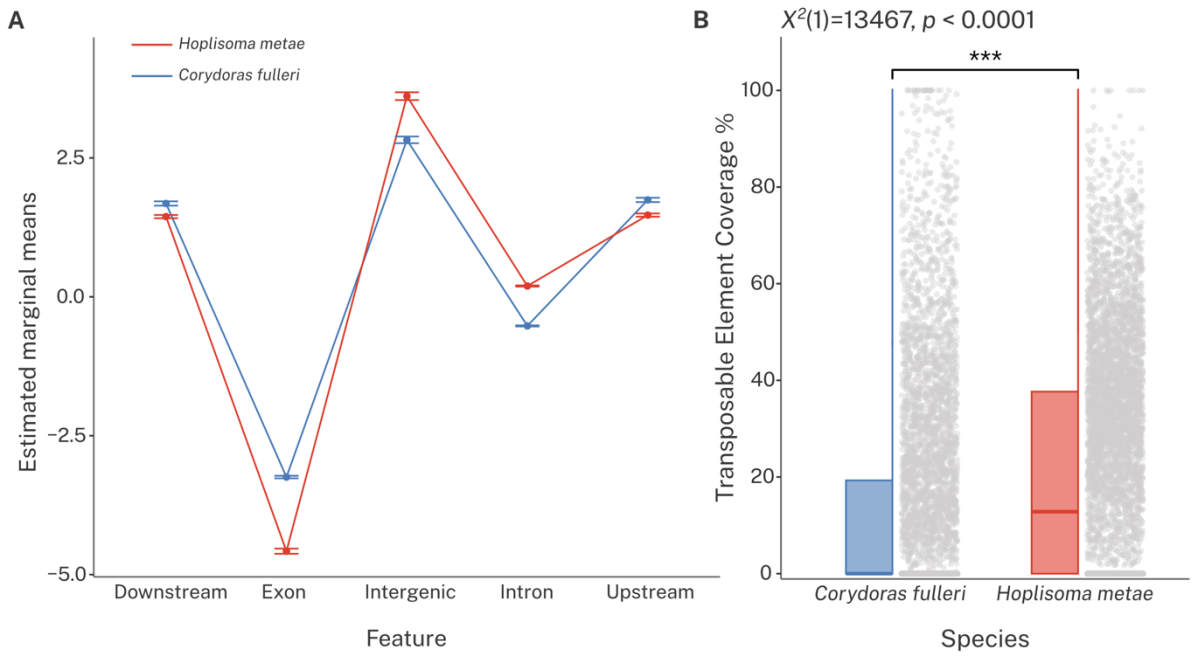


Figure 2.9 A The interaction of Corydoradinae species and genomic feature on the transposable element presence or absence, represented by the estimated marginal means (EMM). Error bars represent the 95% confidence intervals. **B** The percentage of introns covered by transposable elements in *C. fulleri* and *H. metae*. Significant (p < 0.0001) differences between species were assessed using a Kruskal-Wallis test. Each point shows the distribution of introns.

Table 2.3 Post-hoc test, exploring the effect of genomic feature and Corydoradinae species (*Corydoras fulleri* and *Hoplisoma metae*) on the presence or absence of transposable elements.

Contrast	Feature / Species	Estimate	SE	df	z.ratio	p.value
Between Species						
C. fulleri - H. metae	Downstream	0.24	0.03	Inf	9.51	<.0001 ***
	Exon	1.33	0.03	Inf	50.72	<.0001 ***
	Intergenic	-0.79	0.05	Inf	-16.65	<.0001 ***
	Intron	-0.72	0.01	Inf	-100.98	<.0001 ***
	Upstream	0.27	0.03	Inf	11.05	<.0001 ***
Within Species						
downstream - exon	*C. fulleri*	4.92	0.023	Inf	213.40	<.0001 ***
downstream - intergenic		-1.15	0.037	Inf	-31.20	<.0001 ***
downstream - intron		2.20	0.021	Inf	107.48	<.0001 ***
downstream - upstream		-0.06	0.028	Inf	-2.28	0.15
exon - intergenic		-6.07	0.033	Inf	-184.28	<.0001 ***
exon - intron		-2.72	0.013	Inf	-216.05	<.0001 ***
exon - upstream		-4.99	0.023	Inf	-216.98	<.0001 ***
intergenic - intron		3.35	0.031	Inf	107.36	<.0001 ***
intergenic - upstream		1.08	0.037	Inf	29.49	<.0001 ***
intron - upstream		-2.27	0.020	Inf	-111.12	<.0001 ***
downstream - exon	*H. metae*	6.02	0.028	Inf	216.35	<.0001 ***
downstream - intergenic		-2.17	0.039	Inf	-55.95	<.0001 ***
downstream - intron		1.25	0.016	Inf	79.68	<.0001 ***
downstream - upstream		-0.03	0.021	Inf	-1.31	0.69
exon - intergenic		-8.19	0.043	Inf	-191.00	<.0001 ***
exon - intron		-4.77	0.024	Inf	-197.56	<.0001 ***
exon - upstream		-6.05	0.028	Inf	-217.48	<.0001 ***
intergenic - intron		3.42	0.036	Inf	94.42	<.0001 ***
intergenic - upstream		2.14	0.039	Inf	55.27	<.0001 ***
intron - upstream		-1.28	0.016	Inf	-81.61	<.0001 ***

and their interaction (X^2= 9182, df=4, p <0.0001) were significant. Further post-hoc tests showed significant differences between the two species in number of TE insertions at each feature (Table 2.3). The probability of a TE being in a feature was higher for *C. fulleri* than *H. metae* in the downstream, upstream and exon regions (Figure 2.9A). Conversely, the probability of a TE being in the introns and intergenic regions of *H. metae* was higher than for *C. fulleri* (Figure 2.9A). Additionally, many of the features showed a significant deviation from each other within the species (Table 2.3). In both species, exons had the significantly lower probability of TE insertion compared to other features (Table 2.3, Figure 2.9A). Additionally, TE insertions into intergenic, downstream and upstream regions were significantly higher than intron regions in both species. Intergenic regions had the highest probability of TE insertions compared to all features. Additionally, downstream and upstream regions, did not significantly differ in insertion probability between both features. To further explore TE insertion in introns, we performed a Kruskal-Wallis rank sum test. This revealed a difference in significant coverage (as a percentage of length) of the introns ($X^2(1)$= 13467, p < 0.0001, Figure 2.9B). This highlights that not only do *H. metae* introns have a higher probability of TE insertion, but more of the intron is composed of TEs. To explore whether the effect of species and intron position on the probability of TEs being present, we performed an additional GLM analysis. We found that both species (X^2= 10343.9, p < 0.0001) and rank (X^2= 926.6, p < 0.0001), and their interaction (X^2= 122.6, p < 0.0001) had a significant effect on TE presence. Post-hoc analysis revealed that TEs were significantly more likely to be present in the first intron of both species (Supplementary Table 2.4, p < 0.0001), but

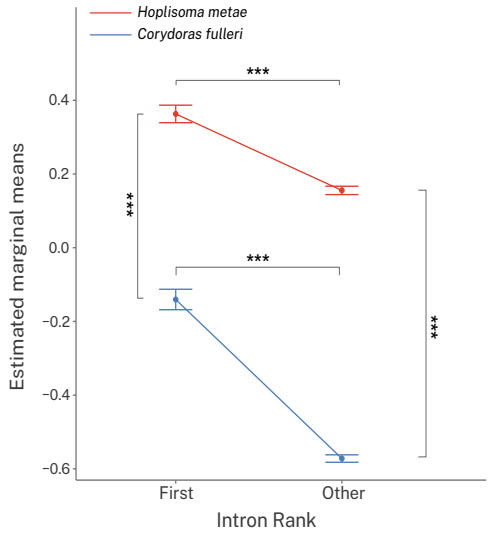


Figure 2.10 The interaction of Corydoradinae species and intron position on the transposable element presence or absence, represented by the estimated marginal means (EMM). Error bars represent the 95% confidence intervals. Asterisks denote significance p < 0.0001.

significantly higher in both positions within *H. metae* (Figure 2.10, Supplementary Table 2.4, $p < 0.0001$).

2.5 Discussion

Here, we used a comparative genomic approach to investigate genomic characteristics in putative diploid and polyploid Corydoradinae catfish species. We found strong evidence for a single WGD in *H. metae* using allelic depth distribution, synonymous substitutions per synonymous site (Ks) and kmer depth distribution (implemented through GenomeScope and Smudgeplot). Evidence from Homeobox and TLR gene families was more equivocal with little evidence of a WGD found using Hox clusters and some evidence in TLRs. We identified considerable differences in genome architecture between the two Corydoradinae species. Intron lengths were found to be significantly longer in *H. metae* than in *C. fulleri,* as were intergenic regions. We found that in both regions, this increase in length accompanied by an increase in transposable element insertions, indicating a role for transposable element proliferation as a driver of genome size in the group.

The Ks distribution provided clear evidence of a very recent WGD in *H. metae* with the first peak at Ks=0.04 (KDE). When calibrated to the FSGD event, the event is estimated to have occurred 4.86 mya. Previously the putative WGD was estimated to have occurred 20-30 mya, so the current estimate suggests an earlier event than previously anticipated (Marburger *et al. 2018*). This is younger than the Cyprinidae specific duplication, which occurred between 12.4 mya and 23 mya (Ma *et al.*, 2014). Additionally, both the allele depth distribution and kmer density distribution techniques not only showed evidence for *H. metae* being a tetraploid, but suggested it was an autotetraploid. This was demonstrated in both cases by the ratio of heterozygotes frequencies. Autopolyploids arise within a single lineage, and subsequently have two sets of homologous chromosomes, each able to pair with any of the other homologs during meiosis (tetrasomal inheritance)(Bomblies and Madlung, 2014; Bomblies, 2023). The timing of this WGD indicates that this may be one of the most recent autopolyploid WGD events known in vertebrates. Both the sturgeon/paddlefish and Salmonoid specific WGD events originate via autopolyploidy, but both are much older (Redmond *et al.*, 2023). As a result, the *Hoplisoma* specific duplication could represent the opportunity to study a much earlier stage of autopolyploid evolution in a vertebrate system. Additionally, some evidence suggests the origin of FSGD event was autoploidy (Braasch and Postlethwait, 2012; Qi *et al.*, 2024). This event is thought to be key in the subsequent diversification of teleosts, and as such an empirical example of an early teleost autoploidy event , such as that in *Hoplisoma*, could shed light on the forces shaping the genome shortly after the FSGD (Braasch and Postlethwait, 2012; Qi *et al.*, 2024). Although we have used only one *Hoplisoma* species, it is likely the WGD occurred at the base of the group, as indicated

by Marburger *et al.* (2018). The very recent nature of the WGD event is likely to have an impact on our ability to detect the WGD using some methods e.g. Hox gene family duplicate retention. Particularly in the case of a recent autopolyploid event, the gene copies have had minimal time to accumulate differences. This high sequence similarity among duplicated regions also increases the probability of structural errors during the genome assembly process.

The Hox genes play a vital role in morphogenesis, particularly in the anterior-posterior body axis and are often used to identify WGD events (Ozernyuk and Schepetov, 2022). Here, we have identified seven clusters of hox genes in both *C. fulleri* and *H. metae*. In H*. metae,* we failed to find any additional clusters, and only two additional Hox genes. This does not directly support a hypothesis of WGD in the group, in contrast to the whole genome evidence. In the evolutionary history of teleosts, it is not unheard of for whole clusters to be lost. For example, HoxCb has been lost in the pufferfish, stickleback, medaka and tilapia, whilst HoxDb has been lost in the zebrafish, blunt snout bream and catfish (Amores *et al.*, 1998; Jaillon *et al.*, 2004; Kurosawa *et al.*, 2006; Hoegg *et al.*, 2007; Zou *et al.*, 2007; Kim *et al.*, 2018; Ozernyuk and Schepetov, 2022). Based on the estimated recent age of the putative *Hoplisoma* WGD at 4.86 mya, we believe it to be unlikely that gene loss of seven clusters has occurred. It is more likely that either assembly collapse or concatenation of haplotypes has resulted in the difference between expected (based on a WGD) and observed Hox clusters diversity.

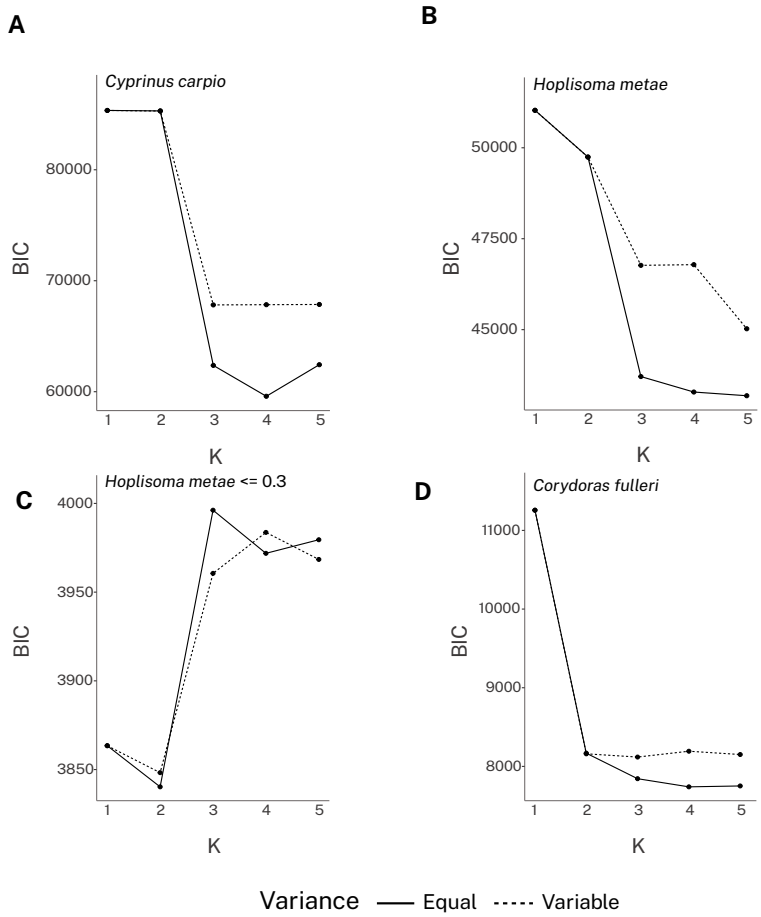
The Hox genes are highly conserved, increasing the likelihood of them being misassembled. To account for this, we investigated a second gene family to determine if evidence could be found elsewhere. The toll-like receptors (TLR) are a group of type I transmembrane proteins which are involved in eliciting an immune response (Nie *et al.*, 2018). In teleosts, 21 TLRs have been identified, some of which are orthologs of mammal TLRs and others are teleost specific (Nie *et al.*, 2018). We found some evidence for duplication in the *H. metae* TLRs. The genes TRL13, TLR20, TLR21 were duplicated in *H. metae* relative to *C. fulleri*, but we did not see absolute duplication across the TLRs. This could the result of gene loss following the duplication event, or due to collapse in assembly, as previously discussed. Previous work did find evidence for gene duplication using allele frequencies in TLR1 and TLR2 in *H. araguaiaensis*, a sister species to *H. metae*, but not in *C. maculifer*, a sister to *C. fulleri* (Bell *et al.*, 2020). This indicates perhaps assembly collapse is masking patterns of duplication in the gene families. However, these hypotheses could only be uncoupled with a better understanding of gene loss and retention after this putative WGD and a more contiguous genome assembly.

Significant increases in genome size can occur without whole genome duplications. In the Antarctic notothenioid fishes, TE have almost entirely accounted for changes in genome size (Bista *et al.*, 2023). Additionally, the same TE genus, long terminal repeats (LTR), are found to have undergone expansion

within many of the giant amphibian genomes, suggesting this specific genus plays a key role in genome expansion (Nowoshilow *et al.*, 2018; Dittrich *et al.*, 2024). Here, we did find an increase in overall abundance of transposable elements as well as an increase in those in the DNA transposon genera hAT in *H. metae*, relative to *C. fulleri*. Additionally, we showed there was a higher probability of TE insertion into the introns and intergenic regions in *H. metae* relative to *C. fulleri* but not in other genomic regions. This indicates that TE insertion within the introns and intergenic regions could be a driver for genome size increase in *H. metae*. Additionally, we showed the probability of TE insertions is higher in the first introns of *H. metae* compared to other introns and those of *C. fulleri*. The first intron plays a role in regulating gene expression and thus, insertions into this region could result in phenotypic change, increasing the adaptive potential *H. metae*. This was observed in the peppered moth *Biston betularia* and in the Midas cichlid *Amphilophus spp.,* where TE insertions into the first intron of pigment genes resulted in colour pattern polymorphism (Van't Hof *et al.*, 2016; Kratochwil *et al.*, 2022). *Hoplisoma* is a species rich genus displaying considerable phenotypic diversity, particularly with regards to colour pattern (Alexandrou *et al.*, 2011).

In summary, the work presented here adds to the current evidence for a *Hoplisoma* specific WGD event, which we have shown to have originated from a single lineage, known as autopolyploidy (Marburger *et al.*, 2018). We have highlighted the recent origin for this event, making *Hoplisoma* the youngest autopolyploid teleost lineage currently identified. Whilst this offers ample opportunity to explore the early evolution of autopolyploids, it also represents some logistical challenges with regards to genome assembly. Additionally, we have shown considerable differences in genomic architecture, independent of the WGD event, between *H. metae* and *C. fulleri*. We also demonstrated intron expansion to be a major contributor to genome size increase within the Corydoradinae and provided evidence to suggest that TE proliferation is driving this shift in intron size in the subfamily.

2.6 Supplementary Materials



Supplementary Figure 2.1 The Bayesian Information Criterion (BIC) for Generalised mixture models (GMM) used to estimate peak location in synonymous substitutions per synonymous site (Ks) in **A** *Cyprinus carpio* **B** *Hoplisoma metae* **C** *Hoplisoma metae* <=0.3 **D** *Corydoras fulleri.*

Supplementary Table 2.1 The assembly statistics of transcriptomes used as evidence to annotate *Corydoras fulleri* and *Hoplisoma metae* genomes.

Species	Tissue	Assembly Method	Transcript Number	Busco Score				
				Complete	Single	Duplicated	Fragmented	Missing
Hoplisoma thinned	Mixed	-	154224	77.3	45.6	31.7	4.8	17.9
Corydoras thinned	Mixed	-	178883	78.5	18.7	59.8	5	16.5
Hoplisoma araguaiaensis	Muscle	*de novo*	65992	27	22.7	4.3	9.8	63.2
Hoplisoma arcuatus	Venom Gland	Genome Guided	157773	32.5	25.9	6.6	15.2	52.3
Hoplisoma arcuatus	Venom Gland	Genome Guided	142078	39	33.3	5.7	15	46
Brochis haroldschultzei	Muscle	*de novo*	115020	70.5	53.2	17.3	5.6	23.9
Corydoras maculifer	Muscle	*de novo*	66579	20.9	14.9	6	13	66
Corydoras narcissus	Venom Gland	Genome Guided	227443	72.3	46.9	26.3	6.6	21.1
Corydoras narcissus	Venom Gland	Genome Guided	201481	72.9	47.5	25.4	6	21.1
Hoplisoma paleatus	Muscle	*de novo*	118445	65.9	51.5	13.3	7.8	26.3
Corydoras desana	Venom Gland	Genome Guided	110198	53	39.9	13.1	20.4	36.6
Corydoras desana	Venom Gland	Genome Guided	116793	66.5	46.5	20	7	26.5
Corydoras desana	Venom Gland	Genome Guided	125717	72.3	49.6	22.7	5.7	22
Hoplisoma nattereri	Liver	Genome Guided	15172	7.9	7.6	0.3	8	84.1
Hoplisoma julii	Gill	Genome Guided	92302	26.5	23.1	3.4	15.4	58.1
Hoplisoma tukano	Venom Gland	Genome Guided	92408	19.2	17.1	2.1	4.86	65
Hoplisoma tukano	Venom Gland	Genome Guided	68774	23.3	20.8	2.5	17.6	59.1

Supplementary Table 2.2 The BUSCO scores for both the *Corydoras fulleri* and *Hoplisoma metae* genome assembly and annotations, indicating quality of annotation or assembly.

Species	Input	Database	Busco				
			Complete	Single	Duplicated	Fragmented	Missing
H. metae	Assembly	vertebrata_odb10	85.7	75.3	10.4	7.6	6.7
H. metae	Assembly	actinopterygii_odb10	83.5	74	9.5	4.6	11.9
H. metae	Annotation	vertebrata_odb10	35.9	33.2	2.7	15.6	48.5
H. metae	Annotation	vertebrata_odb10	38.4	34.9	3.5	30	31
C. fulleri	Assembly	actinopterygii_odb10	94	94.2	1.3	1.4	4.6
C. fulleri	Assembly	vertebrata_odb10	95.2	94.2	1	1.1	3.7
C. fulleri	Annotation	actinopterygii_odb10	76.8	75	1.8	5.5	17.7
C. fulleri	Annotation	vertebrata_odb10	79.2	77.9	1.3	8.6	12.2

Supplementary Table 2.3 Post-hoc testing, exploring differences in length of genomic features between *Corydoras fulleri* and *Hoplisoma metae*

Feature	Estimate	SE	df	t.ratio	p.value
Intron- Peak1	0.054	0.01	317369	10.34	<.0001
Intron- Peak2	-1.28	0.00	317369	-348.09	<.0001
Intergenic	-0.45	0.01	50739	-31.63	<.0001
Exon	-0.01	0.00	374851	-3.39	<.0001

Supplementary Table 2.4 Post-hoc testing, exploring the effect of intron position within gene models and Corydoradinae species (*Corydoras fulleri* and *Hoplisoma metae*) on transposable element presence or absence.

Contrast	Feature / Species	Estimate	SE	df	z.ratio	p.value
Between Species						
C. fulleri - H. metae	First	-0.50	0.02	Inf	-26.99	<.0001***
C. fulleri - H. metae	Other	-0.73	0.01	Inf	-93.93	<.0001***
Within Species						
First - Other	*C. fulleri*	0.43	0.02	Inf	28.63	<.0001***
First - Other	*H. metae*	0.21	0.01	Inf	15.43	<.0001***

2.7 Attribution Statement

Project conception by Martin Taylor, Emily Phelps and Ellen Bell. The initial genome assembly of *Corydoras fulleri* as well as the genome assembly for *Hoplisoma metae* were assembled by Ellen Bell in 2019. Additionally, the species-specific repeat library for *Corydoras fulleri* was curated by Ellen Bell and Chris Butler for the publication Bell, E.A., Butler, C.L. *et al.* (2023), Mol. Ecol. Res. 22(2):823-833. The *Hoplisoma metae* library was also curated by Ellen Bell and Chris Butler. Genome annotations and all other analysis was performed by Emily Phelps. Supervision by Martin Taylor.

2.8 References

Alexandrou, M.A., Oliveira, C., Maillard, M., *et al.* (2011) Competition and phylogeny determine community structure in Müllerian co-mimics, *Nature*, 469(7328), pp. 84–88. https://doi.org/10.1038/nature09660.

Amores, A., Force, A., Yan, Y.L., *et al.* (1998) Zebrafish hox clusters and vertebrate genome evolution, *Science*, 282(5394), pp. 1711–1714. https://doi.org/10.1126/science.282.5394.1711.

Andrews, S. (2010) *FastQC: A Quality Control Tool for High Throughput Sequence*. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Bell, E.A., Cable, J., Oliveira, C., *et al.* (2020) Help or hindrance? The evolutionary impact of whole-genome duplication on immunogenetic diversity and parasite load, *Ecology And Evolution*, 10(24), pp. 13949–13956. https://doi.org/10.1002/ece3.6987.

Bell, E.A., Butler, C.L., Oliveira, C., *et al.* (2022) Transposable element annotation in non-model species: The benefits of species-specific repeat libraries using semi-automated EDTA and DeepTE de novo pipelines, *Molecular Ecology Resources*, 22(2), pp. 823–833. https://doi.org/10.1111/1755-0998.13489.

Bi, X., Wang, K., Yang, L., *et al.* (2021) Tracing the genetic footprints of vertebrate landing in non-teleost ray-finned fishes, *Cell*, 184(5), pp. 1377–1391. https://doi.org/10.1016/j.cell.2021.01.046.

Bista, I., Wood, J.M.D., Desvignes, T., *et al.* (2023) Genomics of cold adaptations in the Antarctic notothenioid fish radiation, *Nature Communications*, 14(1). https://doi.org/10.1038/s41467-023-38567-6.

Blommaert, J. (2020) Genome size evolution: towards new model systems for old questions, *Proceedings. Biological Sciences*, 287(1933). https://doi.org/10.1098/rspb.2020.1441.

Bomblies, K. (2023) Learning to tango with four (or more): the molecular basis of adaptation to polyploid meiosis, *Plant Reproduction*, 36(1), pp. 107–124. https://doi.org/10.1007/s00497-022-00448-1.

Bomblies, K. and Madlung, A. (2014) Polyploidy in the Arabidopsis genus, *Chromosome Research*, 22(2), pp. 117–134. https://doi.org/10.1007/s10577-014-9416-x.

Braasch, I. and Postlethwait, J.H. (2012) Polyploidy in Fish and the Teleost Genome Duplication, in P.S. Soltis and D.E. Soltis (eds) *Polyploidy And Genome Evolution*. Berlin, Heidelberg: Springer, pp. 341–383. https://doi.org/10.1007/978-3-642-31442-1_17.

Braasch, I., Schartl, M. and Volff, J.-N. (2007) Evolution of pigment synthesis pathways by gene and genome duplication in fish, *BMC Evolutionary Biology*, 7(1), pp. 1–18. https://doi.org/10.1186/1471-2148-7-74.

Braasch, I., Volff, J.N. and Schartl, M. (2008) The evolution of teleost pigmentation and the fish-specific genome duplication, *Journal Of Fish Biology*, 73(8), pp. 1891–1918. https://doi.org/10.1111/j.1095-8649.2008.02011.x.

Brawand, D., Wagner, C.E., Li, Y.I., *et al.* (2014) The genomic substrate for adaptive radiation in African cichlid fish, *Nature*, 513(7518), pp. 375–381. https://doi.org/10.1038/nature13726.

Camacho, C., Coulouris, G., Avagyan, V., *et al.* (2009) BLAST+: architecture and applications, *BMC Bioinformatics*, 10. https://doi.org/10.1186/1471-2105-10-421.

Cantarel, B.L., Korf, I., Robb, S.M.C., *et al.* (2008) MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes, *Genome Research*, 18(1), pp. 188–196. https://doi.org/10.1101/gr.6743907.

Card, D.C., Adams, R.H., Schield, D.R., *et al.* (2019) Genomic Basis of Convergent Island Phenotypes in Boa Constrictors, *Genome Biology And Evolution*, 11(11), pp. 3123–3143. https://doi.org/10.1093/gbe/evz226.

Cerveau, N. and Jackson, D.J. (2016) Combining independent de novo assemblies optimizes the coding transcriptome for nonconventional model eukaryotic organisms, *BMC Bioinformatics*, 17(1). https://doi.org/10.1186/s12859-016-1406-x.

Chen, W., Zou, M., Li, Y., *et al.* (2021) Sequencing an F1 hybrid of Silurus asotus and S. meridionalis enabled the assembly of high-quality parental genomes, *Scientific Reports*, 11(1). https://doi.org/10.1038/s41598-021-93257-x.

Chen, Y., Zhang, Y., Wang, A.Y., *et al.* (2021) Accurate long-read de novo assembly evaluation with Inspector, *Genome Biology*, 22(1). https://doi.org/10.1186/s13059-021-02527-4.

Dainat, J. (2020) *AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format*. Available at: https://www.doi.org/10.5281/zenodo.355271.

Danecek, P., Bonfield, J.K., Liddle, J., *et al.* (2021) Twelve years of SAMtools and BCFtools, *GigaScience*, 10(2). https://doi.org/10.1093/gigascience/giab008.

Davesne, D., Friedman, M., Schmitt, A.D., *et al.* (2021) Fossilized cell structures identify an ancient origin for the teleost whole-genome duplication, *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 118(30). https://doi.org/10.1073/pnas.2101780118.

Dias, A.C., Tencatt, L.F.C., Roxo, F.F., *et al.* (2024) Phylogenomic analyses in the complex Neotropical subfamily Corydoradinae (Siluriformes: Callichthyidae) with a new classification based on morphological and molecular data, *Zoological Journal Of The Linnean Society*. https://doi.org/10.1093/zoolinnean/zlae053

Dittrich, C., Hoelzl, F., Smith, S., *et al.* (2024) Genome assembly of the dyeing poison frog provides insights into the dynamics of transposable element and genome-size evolution, *Genome Biology And Evolution*, 16(6). https://doi.org/10.1093/gbe/evae109.

Elliott, T.A. and Gregory, T.R. (2015) What's in a genome? The C-value enigma and the evolution of eukaryotic genome content, *Philosophical Transactions Of The Royal Society Of London. Series B*, 370(1678). https://doi.org/10.1098/rstb.2014.0331.

Feiner, N. (2016) Accumulation of transposable elements in Hox gene clusters during adaptive radiation of Anolis lizards, *Proceedings. Biological Sciences*, 283(1840). https://doi.org/10.1098/rspb.2016.1555.

Fu, L., Niu, B., Zhu, Z., *et al.* (2012) CD-HIT: accelerated for clustering the next-generation sequencing data, *Bioinformatics*, 28(23), pp. 3150–3152. https://doi.org/10.1093/bioinformatics/bts565.

Fuller, I.A.M. and Evers, H.E. (2005) *Identifying Corydoradinae Catfishe*. Kidderminster, England: Ian Fuller Enterprises.

Gong, G., Dan, C., Xiao, S., *et al.* (2018) Chromosomal-level assembly of yellow catfish genome using third-generation DNA sequencing and Hi-C analysis, *GigaScience*, 7(11). https://doi.org/10.1093/gigascience/giy120.

Gong, G., Ke, W., Liao, Q., *et al.* (2023) A chromosome-level genome assembly of the darkbarbel catfish Pelteobagrus vachelli, *Scientific Data*, 10(1). https://doi.org/10.1038/s41597-023-02509-0.

Grabherr, M.G., Haas, B.J., Yassour, M., *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome, *Nature Biotechnology*, 29(7), pp. 644–652. https://doi.org/10.1038/nbt.1883.

Gundappa, M.K., To, T.-H., Grønvold, L., *et al.* (2022) Genome-wide reconstruction of rediploidization following autopolyploidization across one hundred million years of Salmonid evolution, *Molecular Biology And Evolution*, 39(1). https://doi.org/10.1093/molbev/msab310.

Hartig, F. (2022) DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. Available at: https://CRAN.R-project.org/package=DHARMa.

He, W.-P., Zhou, J., Li, Z., *et al.* (2021) Chromosome-level genome assembly of the Chinese longsnout catfish Leiocassis longirostris, *Zoological Research*, 42(4), pp. 417–422. https://doi.org/10.24272/j.issn.2095-8137.2020.327.

Hoegg, S., Boore, J.L., Kuehl, J.V., *et al.* (2007) Comparative phylogenomic analyses of teleost fish Hox gene clusters: lessons from the cichlid fish Astatotilapia burtoni, *BMC Genomics*, 8(1). https://doi.org/10.1186/1471-2164-8-317.

Hoff, K.J. and Stanke, M. (2019) Predicting Genes in Single Genomes with AUGUSTUS, *Current Protocols In Bioinformatics*, 65(1). https://doi.org/10.1002/cpbi.57.

Holt, C. and Yandell, M. (2011) MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects, *BMC Bioinformatics*, 12(1), pp. 1–14. https://doi.org/10.1186/1471-2105-12-491.

Jaillon, O., Aury, J.-M., Brunet, F., *et al.* (2004) Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype, *Nature*, pp. 946–957. https://doi.org/10.1038/nature03025.

Jakt, L.M., Dubin, A. and Johansen, S.D. (2022) Intron size minimisation in teleosts, *BMC Genomics*, 23(1). https://doi.org/10.1186/s12864-022-08760-w.

Jeffery, N.W., Yampolsky, L. and Gregory, T.R. (2017) Nuclear DNA content correlates with depth, body size, and diversification rate in amphipod crustaceans from ancient Lake Baikal, Russia, *Genome*, 60(4), pp. 303–309. https://doi.org/10.1139/gen-2016-0128.

Jiang, W., Lv, Y., Cheng, L., *et al.* (2019) Whole-Genome Sequencing of the Giant Devil Catfish, Bagarius yarrelli, *Genome Biology And Evolution*, 11(8), pp. 2071–2077. https://doi.org/10.1093/gbe/evz143.

Jockusch, E.L. (1997) An evolutionary correlate of genome size change in plethodontid salamanders, *Proceedings. Biological Sciences*, 264(1381), pp. 597–604. https://doi.org/10.1098/rspb.1997.0085.

Jun, G., Wing, M.K., Abecasis, G.R., *et al.* (2015) An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data, *Genome Research*, 25(6), pp. 918–925. https://doi.org/10.1101/gr.176552.114.

Kassambara, A. (2023) ggpubr: 'ggplot2' Based Publication Ready Plots. Available at: https://CRAN.R-project.org/package=ggpubr.

Kim, D., Paggi, J.M., Park, C., *et al.* (2019) Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype, *Nature Biotechnology*, 37(8), pp. 907–915. https://doi.org/10.1038/s41587-019-0201-4.

Kim, O.T.P., Nguyen, P.T., Shoguchi, E., *et al.* (2018) A draft genome of the striped catfish, Pangasianodon hypophthalmus, for comparative analysis of genes relevant to development and a resource for aquaculture improvement, *BMC Genomics*, 19(1), pp. 1–16. https://doi.org/10.1186/s12864-018-5079-x.

Korf, I. (2004) Gene finding in novel genomes, *BMC Bioinformatics*, 5. https://doi.org/10.1186/1471-2105-5-59.

Kratochwil, C.F., Kautt, A.F., Nater, A., *et al.* (2022) An intronic transposon insertion associates with a trans-species color polymorphism in Midas cichlid fishes, *Nature Communications*, 13(1). https://doi.org/10.1038/s41467-021-27685-8.

Kuraku, S. and Meyer, A. (2009) The evolution and maintenance of Hox gene clusters in vertebrates and the teleost-specific genome duplication, *The International Journal Of Developmental Biology*, 53(5–6), pp. 765–773. https://doi.org/10.1387/ijdb.072533km.

Kurosawa, G., Takamatsu, N., Takahashi, M., *et al.* (2006) Organization and structure of hox gene loci in medaka genome and comparison with those of pufferfish and zebrafish genomes, *Gene*, 370, pp. 75–82. https://doi.org/10.1016/j.gene.2005.11.015.

Kushwaha, B., Pandey, M., Das, P., *et al.* (2021) The genome of walking catfish Clarias magur (Hamilton, 1822) unveils the genetic basis that may have facilitated the development of environmental and terrestrial adaptation systems in air-breathing catfishes, *DNA Research*, 28(1). https://doi.org/10.1093/dnares/dsaa031.

Lawrence, M., Huber, W., Pagès, H., *et al.* (2013) Software for computing and annotating genomic ranges, *PLoS Computational Biology*, 9(8). https://doi.org/10.1371/journal.pcbi.1003118.

Lemopoulos, A. and Montoya-Burgos, J.I. (2022) Whole genome assembly of the armored loricariid catfish Ancistrus triradiatus highlights herbivory signatures, *Molecular Genetics And Genomics: MGG*, 297(6), pp. 1627–1642. https://doi.org/10.1007/s00438-022-01947-6.

Lenth, R.V. (2024) emmeans: Estimated Marginal Means, aka Least-Squares Means. Available at: https://CRAN.R-project.org/package=emmeans.

Li, H. (2018) Minimap2: pairwise alignment for nucleotide sequences, *Bioinformatics (Oxford, England)*, 34(18), pp. 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform, *Bioinformatics*, 25(14), pp. 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Lorin, T., Brunet, F.G., Laudet, V., *et al.* (2018) 1 Teleost fish-specific preferential retention of pigmentation gene-containing families after 2 whole genome duplications in vertebrates, *G3* , 8(5), pp. 1795–1806. https://doi.org/10.1534/g3.118.200201.

Ma, D., Ding, Q., Guo, Z., *et al.* (2022) The genome of a mangrove plant, Avicennia marina, provides insights into adaptation to coastal intertidal habitats, *Planta*, 256(1). https://doi.org/10.1007/s00425-022-03916-0.

Ma, W., Zhu, Z.-H., Bi, X.-Y., *et al.* (2014) Allopolyploidization is not so simple: evidence from the origin of the tribe Cyprinini (Teleostei: Cypriniformes), *Current Molecular Medicine*, 14(10), pp. 1331–1338. https://doi.org/10.2174/1566524014666141203101543.

Marburger, S., Alexandrou, M.A., Taggart, J.B., *et al.* (2018) Whole genome duplication and transposable element proliferation drive genome expansion in Corydoradinae catfishes, *Proceedings Of The Royal Society B*, 285(1872). https://doi.org/10.1098/rspb.2017.2732.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads, *EMBnet.Journal*, 17(1), pp. 10–12. https://doi.org/10.14806/ej.17.1.200.

Mulhair, P.O., Crowley, L., Boyes, D.H., *et al.* (2023) Diversity, duplication, and genomic organization of homeobox genes in Lepidoptera, *Genome Research*, 33(1), pp. 32–44. https://doi.org/10.1101/gr.277118.122.

Nie, L., Cai, S.-Y., Shao, J.-Z., *et al.* (2018) Toll-like receptors, associated biological roles, and signaling networks in non-mammals, *Frontiers In Immunology*, 9. https://doi.org/10.3389/fimmu.2018.01523.

Novikova, P.Y., Brennan, I.G., Booker, W., *et al.* (2020) Polyploidy breaks speciation barriers in Australian burrowing frogs Neobatrachus, *PLoS Genetics*, 16(5). https://doi.org/10.1371/journal.pgen.1008769.

Nowoshilow, S., Schloissnig, S., Fei, J.-F., *et al.* (2018) The axolotl genome and the evolution of key tissue formation regulators, *Nature*, 554(7690), pp. 50–55. https://doi.org/10.1038/nature25458.

Ohno, S. (1970) *Evolution by gene duplication*. New York: Springer.

Ozernyuk, N. and Schepetov, D. (2022) Hox gene cluster organization and genome duplications in fishes and mammals: transcript variant distribution along the anterior-posterior axis, *International Journal Of Molecular Sciences*, 23(17). https://doi.org/10.3390/ijms23179990.

Ozerov, M.Y., Flajšhans, M., Noreikiene, K., *et al.* (2020) Draft genome assembly of the freshwater apex predator wels catfish (Silurus glanis) using linked-read sequencing, *G3 Genes|Genomes|Genetics*, 10(11), pp. 3897–3906. https://doi.org/10.1534/g3.120.401711.

Paradis, E. and Schliep, K. (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R, *Bioinformatics*, pp. 526–528. https://doi.org/10.1093/bioinformatics/bty633.

Parente, T.E., Moreira, D.A., Magalhães, M.G.P., *et al.* (2017) The liver transcriptome of suckermouth armoured catfish (Pterygoplichthys anisitsi, Loricariidae): Identification of expansions in defensome gene families, *Marine Pollution Bulletin*, 115(1–2), pp. 352–361. https://doi.org/10.1016/j.marpolbul.2016.12.012.

Parey, E., Louis, A., Montfort, J., *et al.* (2022) An atlas of fish genome evolution reveals delayed rediploidization following the teleost whole-genome duplication, *Genome Research*, 32(9), pp. 1685–1697. https://doi.org/10.1101/gr.276953.122.

Putnam, N.H., O'Connell, B.L., Stites, J.C., *et al.* (2016) Chromosome-scale shotgun assembly using an in vitro method for long-range linkage, *Genome Research*, 26(3), pp. 342–350. https://doi.org/10.1101/gr.193474.115.

Qi, M., Clark, J., Moody, E.R.R., *et al.* (2024) Molecular dating of the teleost whole genome duplication (3R) is compatible with the expectations of delayed rediploidization, *Genome Biology And Evolution*, 16(7). https://doi.org/10.1093/gbe/evae128.

R Core Team (2023) R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/.

Ranallo-Benavidez, T.R., Jaron, K.S. and Schatz, M.C. (2020) GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes, *Nature Communications*, 11(1). https://doi.org/10.1038/s41467-020-14998-3.

Redmond, A.K., Casey, D., Gundappa, M.K., *et al.* (2023) Independent rediploidization masks shared whole genome duplication in the sturgeon-paddlefish ancestor, *Nature Communications*, 14(1). https://doi.org/10.1038/s41467-023-38714-z.

Riethoven, J.-J.M. (2010) Regulatory regions in DNA: promoters, enhancers, silencers, and insulators, *Methods In Molecular Biology*, 674, pp. 33–42. https://doi.org/10.1007/978-1-60761-854-6_3.

Robertson, F.M., Gundappa, M.K., Grammes, F., *et al.* (2017) Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification, *Genome Biology*, 18(1). https://doi.org/10.1186/s13059-017-1241-z.

Roth, G., Blanke, J. and Wake, D.B. (1994) Cell size predicts morphological complexity in the brains of frogs and salamanders, *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 91(11), pp. 4796–4800. https://doi.org/10.1073/pnas.91.11.4796.

Ruan, J. and Li, H. (2020) Fast and accurate long-read assembly with wtdbg2, *Nature Methods*, 17(2), pp. 155–158. https://doi.org/10.1038/s41592-019-0669-3.

Schnable, P.S., Ware, D., Fulton, R.S., *et al.* (2009) The B73 maize genome: complexity, diversity, and dynamics, *Science*, 326(5956), pp. 1112–1115. https://doi.org/10.1126/science.1178534.

Schrader, L. and Schmitz, J. (2019) The impact of transposable elements in adaptive evolution, *Molecular Ecology*, 28(6), pp. 1537–1549. https://doi.org/10.1111/mec.14794.

Schranz, M.E., Mohammadin, S. and Edger, P.P. (2012) Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model, *Current Opinion In Plant Biology*, 15(2), pp. 147–153. https://doi.org/10.1016/j.pbi.2012.03.011.

Scott, A.D., Van de Velde, J.D. and Novikova, P.Y. (2023) Inference of polyploid origin and inheritance mode from population genomic data, *Methods In Molecular Biology (Clifton, N.J.)*, 2545, pp. 279–295. https://doi.org/10.1007/978-1-0716-2561-3_15.

Scrucca, L., Fraley, C., Murphy, T.B., *et al.* (2023) *Model-Based Clustering, Classification, and Density Estimation Using mclust in R*. https://doi.org/10.1201/9781003277965.

Shao, F., Pan, H., Li, P., *et al.* (2021) Chromosome-Level Genome Assembly of the Asian Red-Tail Catfish (Hemibagrus wyckioides), *Frontiers In Genetics*, 12. https://doi.org/10.3389/fgene.2021.747684.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., *et al.* (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, *Bioinformatics* , 31(19), pp. 3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

Slater, G.S.C. and Birney, E. (2005) Automated generation of heuristics for biological sequence comparison, *BMC Bioinformatics*, 6. https://doi.org/10.1186/1471-2105-6-31.

Smit, A.F.A., Hubley, R. and Green, R. (2013) *RepeatMasker Open-4.0*. Available at: http://www.repeatmasker.org.

Stanke, M., Keller, O., Gunduz, I., *et al.* (2006) AUGUSTUS: ab initio prediction of alternative transcripts, *Nucleic Acids Research*, 34, pp. 435–439. https://doi.org/10.1093/nar/gkl200.

Steinegger, M. and Söding, J. (2017) MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets, *Nature Biotechnology*, 35(11), pp. 1026–1028. https://doi.org/10.1038/nbt.3988.

Sun, Y., Huang, Y., Li, X., *et al.* (2016) Fish-T1K (Transcriptomes of 1,000 Fishes) Project: large-scale transcriptome data for fish evolution studies, *GigaScience*, 5. https://doi.org/10.1186/s13742-016-0124-7.

Tank, D.C., Eastman, J.M., Pennell, M.W., *et al.* (2015) Nested radiations and the pulse of angiosperm diversification: increased diversification rates often follow whole genome duplications, *The New Phytologist*, 207(2), pp. 454–467. https://doi.org/10.1111/nph.13491.

Thomas, C.A., Jr (1971) The genetic organization of chromosomes, *Annual Review Of Genetics*, 5(1), pp. 237–256. https://doi.org/10.1146/annurev.ge.05.120171.001321.

Tian, C.-X., Lin, X.-H., Zhou, D.-Y., *et al.* (2023) A chromosome-level genome assembly of Hong Kong catfish (Clarias fuscus) uncovers a sex-determining region, *BMC Genomics*, 24(1). https://doi.org/10.1186/s12864-023-09394-2.

Tiley, G.P., Barker, M.S. and Burleigh, J.G. (2018) Assessing the performance of Ks plots for detecting ancient whole genome duplications, *Genome Biology And Evolution*, 10(11), pp. 2882–2898. https://doi.org/10.1093/gbe/evy200.

Ullastres, A., Merenciano, M. and González, J. (2021) Regulatory regions in natural transposable element insertions drive interindividual differences in response to immune challenges in Drosophila, *Genome Biology*, 22(1). https://doi.org/10.1186/s13059-021-02471-3.

UniProt Consortium (2021) UniProt: the universal protein knowledgebase in 2021, *Nucleic Acids Research*, 49(1). https://doi.org/10.1093/nar/gkaa1100.

Van de Peer, Y., Mizrachi, E. and Marchal, K. (2017) The evolutionary significance of polyploidy, *Nature Reviews. Genetics*, 18(7), pp. 411–424. https://doi.org/10.1038/nrg.2017.26.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., *et al.* (2013) From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline, *Current Protocols In Bioinformatics*, 43(1). https://doi.org/10.1002/0471250953.bi1110s43.

Vanneste, K., Van de Peer, Y. and Maere, S. (2013) Inference of genome duplications from age distributions revisited, *Molecular Biology And Evolution*, 30(1), pp. 177–190. https://doi.org/10.1093/molbev/mss214.

Van't Hof, A.E., Campagne, P., Rigden, D.J., *et al.* (2016) The industrial melanism mutation in British peppered moths is a transposable element, *Nature*, 534(7605), pp. 102–105. https://doi.org/10.1038/nature17951.

Vaser, R., Sović, I., Nagarajan, N., *et al.* (2017) Fast and accurate de novo genome assembly from long uncorrected reads, *Genome Research*, 27(5), pp. 737–746. https://doi.org/10.1101/gr.214270.116.

Vinogradov, A.E. (1995) Nucleotypic effect in homeotherms: Body-mass-corrected basal metabolic rate of mammals is related to genome size, *Evolution*, 49(6), pp. 1249–1259. https://doi.org/10.1111/j.1558-5646.1995.tb04451.x.

Vinogradov, A.E. (1997) Nucleotypic effect in homeotherms: Body-mass independent resting metabolic rate of passerine birds is related to genome size, *Evolution*, 51(1), pp. 220–225. https://doi.org/10.1111/j.1558-5646.1997.tb02403.x.

Waldbieser, G.C., Liu, S., Yuan, Z., *et al.* (2023) Reference genomes of channel catfish and blue catfish reveal multiple pericentric chromosome inversions, *BMC Biology*, 21(1). https://doi.org/10.1186/s12915-023-01556-8.

Walker, B.J., Abeel, T., Shea, T., *et al.* (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement, *PloS One*, 9(11). https://doi.org/10.1371/journal.pone.0112963.

Wang, Y., Tang, H., Debarry, J.D., *et al.* (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity, *Nucleic Acids Research*, 40(7). https://doi.org/10.1093/nar/gkr1293.

Weiß, C.L., Pais, M., Cano, L.M., *et al.* (2018) nQuire: a statistical framework for ploidy estimation using next generation sequencing, *BMC Bioinformatics*, 19(1). https://doi.org/10.1186/s12859-018-2128-z.

Xu, P., Xu, J., Liu, G., *et al.* (2019) The allotetraploid origin and asymmetrical genome evolution of the common carp Cyprinus carpio, *Nature Communications*, 10(1). https://doi.org/10.1038/s41467-019-12644-1.

Xu, S., Guo, Z., Feng, X., *et al.* (2023) Where whole-genome duplication is most beneficial: Adaptation of mangroves to a wide salinity range between land and sea, *Molecular Ecology*, 32(2), pp. 460–475. https://doi.org/10.1111/mec.16320.

Xu, Y., Shao, F., Chen, W., *et al.* (2022) A chromosome-level genome of the helmet catfish (Cranoglanis bouderius), *Frontiers In Genetics*, 13. https://doi.org/10.3389/fgene.2022.962406.

Yang, Y., Li, Y., Chen, Q., *et al.* (2019) WGDdetector: a pipeline for detecting whole genome duplication events using the genome or transcriptome annotations, *BMC Bioinformatics*, 20(1). https://doi.org/10.1186/s12859-019-2670-3.

Ye, H., Fan, J., Hou, Y., *et al.* (2023) Chromosome-level genome assembly of the largefin longbarbel catfish (Hemibagrus macropterus), *Frontiers In Genetics*, 14. https://doi.org/10.3389/fgene.2023.1297119.

Zhong, Y.-F. and Holland, P.W.H. (2011) HomeoDB2: functional expansion of a comparative homeobox gene database for evolutionary developmental biology, *Evolution & Development*, 13(6), pp. 567–568. https://doi.org/10.1111/j.1525-142X.2011.00513.x.

Zou, S.-M., Jiang, X.-Y., He, Z.-Z., *et al.* (2007) Hox gene clusters in blunt snout bream, Megalobrama amblycephala and comparison with those of zebrafish, fugu and medaka genomes, *Gene*, 400(1–2), pp. 60–70. https://doi.org/10.1016/j.gene.2007.05.021.

Zwaenepoel, A. and Van de Peer, Y. (2019) wgd-simple command line tools for the analysis of ancient whole-genome duplications, *Bioinformatics* , 35(12), pp. 2153–2155. https://doi.org/10.1093/bioinformatics/bty915.

3 The genetics of polymorphic mimicry in the Neotropical catfish, *Corydoras fulleri.*

3.1 Abstract

Müllerian mimicry is widespread across the tree of life, but relatively little is known about the genetics underpinning mimetic warning colouration, outside of a few well studied species. Here we introduce *Corydoras fulleri*, a putative polymorphic Müllerian mimic with three morphs. All three morphs are found in sympatry and form a probable Müllerian mimicry ring with four additional (non-polymorphic) sympatric Corydoradinae *spp,* which resemble two of the polymorphic patterns found in *C. fulleri.* Here using mtDNA sequencing we confirm *C. fulleri* is a single polymorphic species rather than a contact zone between hybridizing sister taxa. We then utilise the natural variation in this system to identify adaptive loci associated with the presence or absence of a melanic blotch on the flank, an essential characteristic allowing *C. fulleri* to share a resemblance with its co-mimics. By utilising low coverage whole genome resequencing and sliding window population genetic methods, we have identified candidate regions of the genome associated with melanin dispersion and density, indicating patterning in this system is utilising known pigmentation pathways. Additionally, the presence of these genes and associated loci across the genome indicates this trait is underpinned by a more complex architecture than those previously identified in mimetic systems, highlighting the value of understanding the genetics of mimicry across a diverse spectrum of mimetic taxon.

3.2 Introduction

The striking visual parallels between some sympatric species inspired early naturalists and subsequently became a key tenet of Wallace's hypotheses on animal colour patterns (Wallace, 1877; Caro, 2017). Subsequently, colour pattern mimicry has become a powerful example of natural selection (Darwin, 1859; Mallet and Joron, 1999). Colour pattern mimicry comes in different forms being broadly defined as ‘protective’, ‘reproductive’ and ‘aggressive’ (Pasteur, 1982). Protective mimicry occurs when two or more species adopt similar or identical coloration to avoid or reduce predation (Ruxton *et al.*, 2018). This includes both Batesian, where a palatable species mimics a non-palatable species, and Müllerian mimicry, where multiple non-palatable species share a resemblance, all benefiting from the reduced costs of predator education (Bates, 1862; Müller, 1878; Ruxton *et al.*, 2018). Protective mimicry may lead to phenotypic resemblance between or among species and colour pattern polymorphism within a single species (Ruxton *et al.*, 2018; Twomey *et al.*, 2020). Occurrences of this phenomenon are found in a variety of taxa including bees (Chatelain *et al.*, 2023), velvet ants (Wilson *et al.*, 2012, 2018), millipedes (Marek and Bond, 2009), *Heliconius* (Merrill *et al.*, 2015) and

Papilio butterflies (Nishikawa *et al.*, 2015), poison dart frogs (Twomey, Vestergaard and Summers, 2014), and fish (Alexandrou *et al.*, 2011; Wright, 2011). Consequently, the phenomenon of colour pattern mimicry provides repeated, independent case studies through which we can explore the forces driving both phenotypic diversification and similarity (Twomey *et al.*, 2020).

Under Müller's original theory, novel phenotypes were preferentially predated, resulting in the convergence of all aposematic taxa to a single geographically widespread phenotype (Mallet and Joron, 1999; Merrill *et al.*, 2015; Ezray *et al.*, 2019). Conversely, in nature, considerable variation in warning colouration exists both within single localities and across wider geographical areas (Marek and Bond, 2009; Joron *et al.*, 2011; Wilson *et al.*, 2012; Twomey, Vestergaard and Summers, 2014; Merrill *et al.*, 2015; Jay *et al.*, 2018; Motyka, Kampova and Bocak, 2018; Ezray *et al.*, 2019). Such polymorphism is paradoxical in Müllerian mimicry systems, as all taxa should gain the greatest protection by converging on a single phenotype, due to frequency-dependent selection purging phenotypes that deviate from the optimum (Mallet and Joron, 1999). The existence of regional polymorphism can be explained largely by the temporal and spatial presence of different model taxa (Mallet, 1999). Conversely, the evolution of locally stable polymorphisms is still not fully understood but evidence in *He. numata* supports a role for predator range, assortative mating and a complex genetic architecture (Chouteau *et al.*, 2017; Holmes, Grundler and Davis Rabosky, 2017; Jay *et al.*, 2018). In further contrast to Müller's original theory, there is limited support for colour pattern convergence, with empirical and theoretical evidence largely supporting the advergence of one aposematic taxon (the mimic) to another (the model)(Mallet, 1999; Symula, Schulte and Summers, 2001; Marek and Bond, 2009; Wilson *et al.*, 2012; Muell *et al.*, 2022). Thus, by better understanding the genetics underpinning polymorphism we can make inferences about the origin and maintenance of such systems.

The Corydoradinae are a diverse subfamily of freshwater catfish, from the family Callichtyidae, with over 208 species distributed across South America (Dias *et al.*, 2024). As with all Callichthyidae, the subfamily Corydoradinae are venomous, possessing venom glands at the base of their dorsal and pectoral fins which are accompanied by sharp lockable spines (Greven, Flasbeck and Passia, 2006). The group shows a huge diversity of pigmentation patterns and have elements of both conspicuous and cryptic patterning (Alexandrou *et al.*, 2011). Many species of Corydoradinae participate in Müllerian mimicry rings, with up to five sympatric mimetic species coexisting in mixed shoals (Alexandrou *et al.*, 2011). Within the group, 27 Müllerian mimicry rings have been identified, with 60 species being classed as co-mimics (Alexandrou *et al.*, 2011; Tencatt and Ohara, 2016; Tencatt *et al.*, 2021, 2024).

Corydoras fulleri is a polymorphic species found in the Rio Madeira basin in southern Peru. *Corydoras fulleri* is found in sympatry with several additional Corydoradinae species, with which it is thought to be a co-mimic. These includes *Hoplisoma cf. acrensis* and *Osteogaster aeneus,* as well as two additional undescribed species (CW62, CW040)(Figure 3.1A)(Tencatt *et al.*, 2021). Three colour morphs of *C. fulleri* have been identified (Figure 3.1A); 1) a horizontal stripe stretching from below the dorsal fin along the caudal peduncle and to the caudal fin 2) a horizontal stripe as previous described, along with a blotch of dark pigmentation on the flank of the individual 3) a blotch on the flank of the individual but lacking the horizontal stripe. However, it appears that there is substantial variation among these morphs, suggesting more of a continuum of phenotypes rather than discrete forms (Figure 3.1B). Individuals with differing colouration have been shown to breed freely in the aquarium, suggesting these phenotypes represent different morphs rather than distinct species.

The variation observed in *C. fulleri* potentially offers an important opportunity to expand the current knowledge regarding the evolution and maintenance of polymorphism within a mimetic species, the understanding of which is largely restricted to Lepidoptera (Merrill *et al.*, 2015; Nishikawa *et al.*,

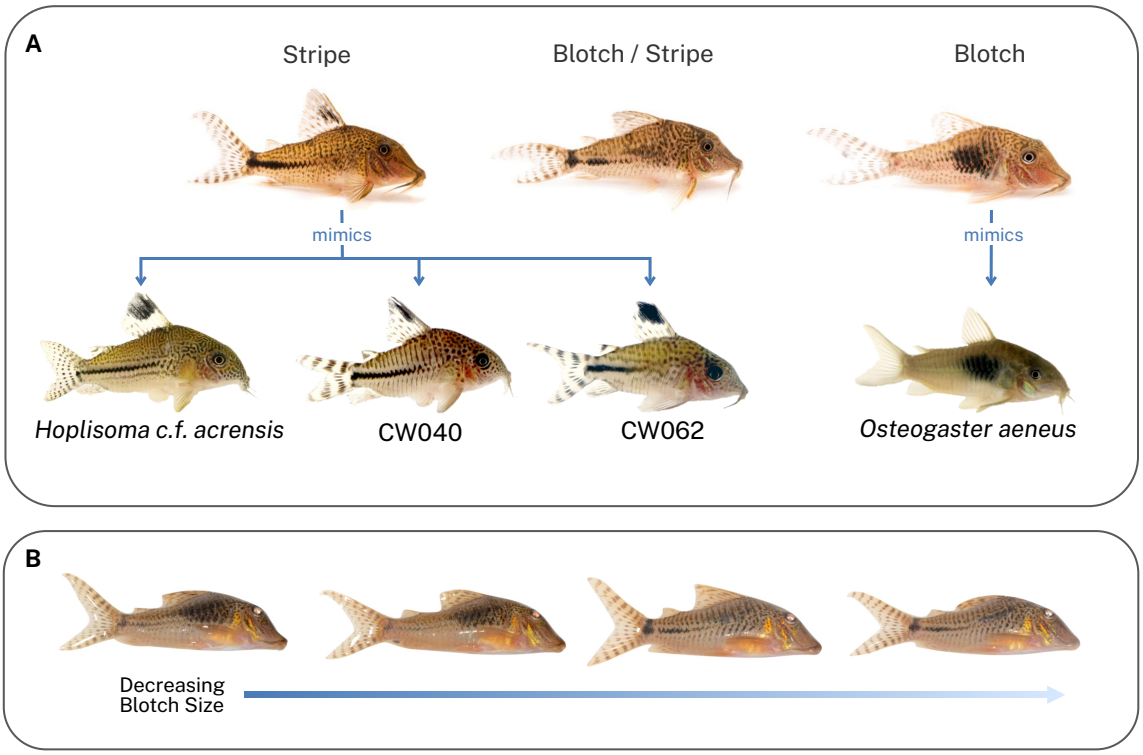


Figure 3.1A The morphs of *Corydoras fulleri* and their putative mimetic interactions with sympatric Corydoradinae species. **B** The variation observed in melanic blotch density and size. Images courtesy of Martin Taylor and Ian Fuller.

2015). Moreover, as one of few potentially polymorphic mimetic species in the Corydoradinae, it offers an opportunity to identify genes associated with pigmentation in the group.

To investigate the molecular basis for mimicry in this system, we have explored the following objectives:

1) Use mtDNA sequencing to establish whether *C. fulleri* is a single polymorphic taxon or represents the hybridization of two distinct species hybridising at a contact zone or zones.

2) Use low coverage whole genome resequencing to investigate the genetic basis of the blotch pigmentation on the flank using a comparison between Blotch/Stripe and Stripe individuals. We aim to identify putative adaptive loci and linked genes to better elucidate the genetic architecture responsible for this putative polymorphic mimic.

3.3 Methods

3.3.1 Sample Collection and DNA extraction

Live individuals were obtained by hand-net or cast-net fishing from two locations in the Madre de Dios region in Peru, in May of 2022 (Cheese Creek and Alegría, n=21, Figure 3.3A). Individuals were euthanised using 0.4ml of clove oil per litre of river water. An additional five ethanol preserved specimens of unknown origin were used for population genetic analysis whilst a further three ethanol preserved specimens, and eight aquarium individuals were used for mitochondrial analysis. DNA was extracted from fin clips using the Qiagen DNeasy Tissue Kit (Qiagen, Germany), following the manufacturer's protocols. Quality and quantity were determined using the Qubit™ 4 fluorometer (Thermofisher Scientific, USA), following the protocol associated with the Qubit™ dsDNA Broad Range Quantification Assay kit (Thermofisher Scientific, USA). Gel electrophoresis was performed to assess DNA fragmentation, and a minimum volume of 20 µL at a concentration of 10 ng/µL was prepared for sequencing. Library preparation and sequencing of individuals was completed at Novogene, Cambridge, using NovaSeq 6000 at a desired depth of 5x per individual.

3.3.1 Sequence filtering, alignment

Raw reads were examined using FastQC (v. 0.11.9) (Andrews, 2010) and trimmed and filtered following the Genome Analysis Toolkit pre-processing pipeline through a custom script (available at https://github.com/EmilyPhelps/CorydoradinaeMimicry2024). Picard (v. 2.20.2-11) (https://broadinstitute.github.io/picard/) tools MarkIlluminaAdapters and SamToFastq were used to filter low quality sequences and remove adapters (Van der Auwera *et al.*, 2013). Filtered FASTQ files were aligned to the *C. fulleri* reference genome using BWA MEM (v. 0.7.17) (Li and Durbin, 2009) and merged with unmapped bam files to retain all relevant sample and quality information using Picard (v.

2.20.2-11) MergeBamAlignment. Files resulting from samples sequenced over multiple lanes were merged using SAMtools merge (v. 1.9)(Danecek *et al.*, 2021). Optical and sequencing duplicates were removed using Picard MarkDuplicates and overlapping reads were clipped by BamUtils (v. 1.0.15) clipOverlap (Jun *et al.*, 2015). Regions around indels were identified and realigned using Genome Analysis ToolKit (v. 3.8-1-0) RealignerTargetCreater and IndelRealigner (Van der Auwera *et al.*, 2013). Finally, read depth was calculated per sample using SAMtools depth.

3.3.2 Mitochondrial Haplotypes

We used Norgal (v.1.0.0)(Al-Nakeeb, Petersen and Sicheritz-Pontén, 2017) to assemble a mitochondrial reference genome. This process utilises the raw reads used to create the nuclear genome assembly and final draft assembly to create a mitochondrial assembly. To annotate the assembly, a NCBI blastn (v.2.9.0+)(Camacho *et al.*, 2009) of the *Hoplisoma trilineatum* mitochondrial genes (NC_049098.1, NCBI) was performed against the *de novo* mitochondrial genome. To obtain individual mitochondrial sequences, the *C. fulleri* whole genome resequencing raw reads where then aligned to the mitochondrial reference using BWA MEM, and variants were called using FreeBayes (v.1.3.1)(Garrison and Marth, 2012) using the --standard-filters (to filter variants with default settings) and--throw-away-indels-obs (to remove indels) options. The resulting vcf file was indexed using GATK (v. 4.2.6.1) IndexFeature and split into sample-specific vcf files using BCFtools (v. 1.10.2) and GATK SelectVariants. The resulting vcf files were then phased using WhatsHap phase (v. 1.7)(Martin, Ebert and Marschall, 2023). Consensus sequences were generated by BCFtools consensus using the-H option, providing a full mitochondrial genome for each individual. Annotated mitochondrial gene sequences were extracted using SAMtools faidx. Sequences were grouped by gene and aligned using the default parameters in MAFFT (v.7.453)(Katoh and Standley, 2013). Gene alignments were loaded into R (v. 4.2.2) using the function read.dna() from the package ape (v. 5.7-1)(Paradis and Schliep, 2019) and haplotypes were estimated using haplotype() from the package pegas (v. 1.2)(Paradis, 2010). Resulting haplotype networks were generated using the pegas function haplonet(). The aesthetics of the resulting plots were edited in Affinity Designer (v. 2.5.3).

3.3.3 Genotype likelihood estimates

ANGSD (v. 0.939-8)(Korneliussen, Albrechtsen and Nielsen, 2014) was used to estimate genotype likelihoods using the GATK model (-GL 2). ANGSD estimates genotype likelihoods based on the mapping and sequence quality scores, thus considering the level of uncertainty associated with low coverage resequencing data (Korneliussen, Albrechtsen and Nielsen, 2014; Lou *et al.*, 2021). We used the *C. fulleri* assembly as the reference and ancestral sequence (-ref/-anc). Initially, SNPs were called across all phenotypes and populations, providing a list of variable sites that could be used in downstream analyses. Sites with a probability <1e-6 (-SNP_pval 1e-6) of being monomorphic were

considered SNPs. Allele frequencies were calculated using fixed major and minor alleles (-doMAF) and only sites with a minor allele frequency below <0.05 (-minMAF 0.05) were retained. Unless otherwise stated, major and minor alleles were inferred from genotype likelihoods (-doMajorMinor 1). A maximum depth threshold of 238 was set (-setMaxDepth 238). This threshold was calculated as the total sample number multiplied by the average read depth per sample. Sites were also excluded if they were only present in less than 50% individuals (-minInd). The remaining sites were filtered for a quality score of > 20 (-minQ 20) and a minimum mapping quality > 30 (-minMapQ 30). Additionally, reads that were not mapped uniquely were discarded (-uniqueOnly). After global SNP calling, SNPs were called per population, per phenotype with the same filters and conditions as above with the following exceptions. Only sites that had been identified as SNPs were included (-sites) and minimum individual (-minInd) was set to include 50% of the sample number per phenotype per population.

3.3.4 Population genetics

3.3.4.1 PCA and Admixture

To visualize population structure, we performed both PCA and admixture on a linkage pruned data set. Variant sites were called globally as detailed above using ANGSD. Linked sites were identified using plink (v. v1.90b6.7)(Purcell *et al.*, 2007) with a window size of 100, a SNP number of 5, and a variance influence factor of 2. A linkage site list was then produced using the R script make_site_list_pruned.r (https://github.com/clairemerot/angsd_pipeline) (Mérot *et al.*, 2021). Genotype likelihoods were then calculated for the remaining sites using ANGSD, and a covariance matrix was generated using PCAngsd (v.1.10)(Meisner and Albrechtsen, 2018). Eigenvalues were calculated from this matrix using the function eigen() in base R. Individual admixture proportions were also estimated using PCAngsd through the-admix option and utilising the automatic ancestral cluster (K) estimation (Meisner and Albrechtsen, 2018).

3.3.4.2 Pairwise F_{ST}

Pairwise F_{ST} (Weir and Cockerham, 1984) was calculated between the Blotch/Stripe and Stripe phenotypes in each wild population and between the two populations. This was achieved using the ANGSD .saf files from per phenotype per population and global population SNP calling respectively. The ANGSD utility Realsfs fst index was then used to generate the weighted F_{ST} estimates. The blotch only phenotype was excluded from this analysis due to low sample sizes.

3.3.4.3 Genomewide F_{ST}

To estimate F_{ST} across the genome, we called SNPs independently for phenotypes within each population, as described above, but with-doGeno 2. This designated the most probable genotype as the genotype, as required for downstream analysis. The resulting bcf files were converted into vcf format using BCFtools convert and then compressed and indexed using BCFtools index. Subsequently,

parseVCFs.py was used to convert the vcf files into the correct format for popgenWindows.py (both available at https://github.com/simonhmartin/genomics_general). The popgenWindows.py script estimates windowed F_{ST}, with windows of a fixed SNP number. This reduces the overestimation of F_{ST} in windows with few SNPs, which can result in erroneous peaks. For this analysis, windows were set to those of 150 SNPs or 50kb windows, depending on which criteria was met first. Outlier windows were then calculated for each population. Windows were considered outliers if they were in the top 99.99th percentile, representing the top 0.01% of windows. This was determined using the base R quantile() function.

3.3.4.4 Linkage Decay

Candidate gene identification can be improved with a better understanding of linkage by allowing only genes likely to be linked to candidate SNPs to be retained. We estimated the rate of linkage decay in the wild specimens and ethanol preserved samples that clustered with either wild population. We calculated the decay of linkage disequilibrium (LD) between loci with ngsLD (v. 1.2.1)(Fox *et al.*, 2019), using genotype likelihoods generated in ANGSD. Only linkage between loci within 1mb (--max_kb_dist 100) and further than 100 SNPs apart (-max_snp_dist 100) was estimated. The average LD across genomic intervals was then estimated using the ld_decay_calc.py script (https://github.com/speciationgenomics/scripts/blob/master/ld_decay_calc.py) and visualized in R.

3.3.4.5 Candidate gene identification

To investigate whether outlier windows were shared between populations, indicating shared origin of colouration or genotypic convergence, we used the findOverlaps() function from the GenomicRanges (v. 1.50.2)(Lawrence *et al.*, 2013). No overlaps were identified, so all outlier regions were retained. Adjacent outlier windows were combined and outlier windows with less than 50 SNPs were combined with the adjacent non-outlier windows. This was achieved using custom R scripts (available at https://github.com/EmilyPhelps/CorydoradinaeMimicry2024). The PCAdapt pcadapt() (v. 4.3.5)(Privé *et al.*, 2020) function was then run on each region in R and the first two principal components were plotted using ggplot()(v.3.5.1)(Wickham, 2016), coloured by phenotype, including the geom_convexhull()(v.0.1.0)(Martin, 2017) layer. For each population, the resulting p-values were integrated with position information for each SNP. All SNPs were then combined into a single per population dataframe. Next Bonferroni correction was applied to the SNP p-values using p.adjust() in base R with the method “bonferroni”. SNPs were then filtered to include only SNPs in regions where the PCA plots separated by phenotype and were non-overlapping. Subsequently, SNPs were filtered to include only those associated with the principal component by which the phenotypes were split. This was achieved using the get.PC() function from PCAdapt. Finally, only SNPs with an adjusted p.value <0.05 were retained as candidate SNPs.

To identify candidate genes associated with the presence or absence of a blotch, we used the *C. fulleri* genome annotation (corydoras_fulleri_v1, Chapter 2). Genes were considered candidates if they were within 40kb of a candidate SNP. Annotated genes were allocated functions and a UniProt accession via a NCBI blastp against the UniProt vertebrate database as part of the functional annotation process outlined in chapter 2(Camacho *et al.*, 2009; UniProt Consortium, 2021). A gene name, protein name and biological process gene ontology (GO) term was provided using the id mapping function on uniprot.com(UniProt Consortium, 2021). There were no shared candidate genes between populations.

GOterm and KEGG pathway enrichment was performed on the candidate genes in each population using GOprofiler web version (Kolberg *et al.*, 2023), against all *Ictalurus punctatus* genes using a Benjamini-Hochberg FDR to correct the p.values for multiple testing. To identify genes with known pigmentation functions, we screened our candidate genes for those in a database assembled from Baxter *et al.,* (2019) and Lorin *et al*., (2018), using filter() from the dplyr package (v. 2.0.0)(Wickham *et al.*, 2019). Additionally, biological process GO terms were searched for terms relating to pigmentation including melanocyte, iridophore, xanthophore, melanin, melanosome, keratinocyte, pigment, pteridine, carotenoid, purine, fibroblast, melanoblast, albinism, colour, color, erythrophore, leucophore and cyanophore. These search terms were based on the methodology of Baxter *et al.,* (2019).

3.3.4.6 Tajima's D

Site allele frequency likelihoods were estimated from genotype likelihoods assuming Hardy–Weinberg equilibrium (-doSaf 1) for Blotch/Stripe and Stripe phenotypes in ANGSD, using only the samples at the phenotypic extremes, as in the genomewide F_{ST} analysis. Per population, per phenotype folded (-fold 1) SFS files were generated using RealSFS. Theta was then calculated from the SFS files using the RealSFS saf2theta. Finally, thetaStat do_stat from ANGSD, was used to calculate Tajima's D in non-overlapping 2kb windows.

3.3.4.7 D_{XY}

To include both variant and invariant sites we ran ANGSD twice. This methodology ensures that sites that are variants within populations but are not variant within phenotypes are captured. Again, we only used samples that reflected the phenotypic extremes, as described above. First, we ran ANGSD within populations across phenotypes with the previously described quality filters and a more stringent minimum individual filter of 80% of the individuals, as D_{XY} can be biased by missing data (Korunes and Samuk, 2021). No SNP or minimum allele filters were included. A sites file for each population was created via this initial run and indexed using sites index (ANGSD)(Korneliussen, Albrechtsen and Nielsen, 2014). Next, we re-ran ANGSD per phenotype within each population

including the same options as the initial run, as well as the indexed sites file (-sites). Per-site D_{XY} was then calculated from the minimum allele frequency of the Blotch/Stripe and Stripe phenotypes of both populations using calcDxy.R (https://github.com/mfumagalli/ngsPopGen/blob/master/scripts/calcDxy.R) written by Joshua Penalba and associated with the ngsPopGen toolkit (Fumagalli *et al.*, 2014). Windowed D_{XY} was then calculated using the script angsd_dxy_step3_processoutputoverwindows_04242020.pl (https://github.com/ksamuk/pixy_analysis/blob/ab90cc1e95513901010a15fb260febfbc5604b9a/data_generation/angsd/angsd_dxy_step3_processoutputoverwindows_04242020.pl#L4) (Korunes and Samuk, 2021) for non-overlapping 5Kb windows.

3.4 Results

An average of 36569002 reads (range = 16906869 - 53807414, median= 33811196) were sequenced with an average GC of 41.62% (range = 40.58 - 42.12, median = 41.68). Trimming and alignment resulted in a mapping rate of 98.31% (median= 99.00%, range= 93.86% - 99.29%), with a mean of 98.29% of the filtered reads being uniquely mapped. Reads were mapped to an average depth of 5.00 ± 12.08 (median= 3.22 ± 11.85, range=3.22-7.00), covering 91.72% of the genome, on average. Global variant calling identified a total of 11,289,938 SNPs.

3.4.1 mtDNA haplotypes

Haplotypes resolved from mitochondrial genes revealed a mean number of 6 (range=2- 9) haplotypes per gene among wild and aquarium samples (Figure 3.2). Most genes resolved the Blotch morph as having a unique haplotype, separated by 2- 4 mutations from the most common Blotch/Stripe and

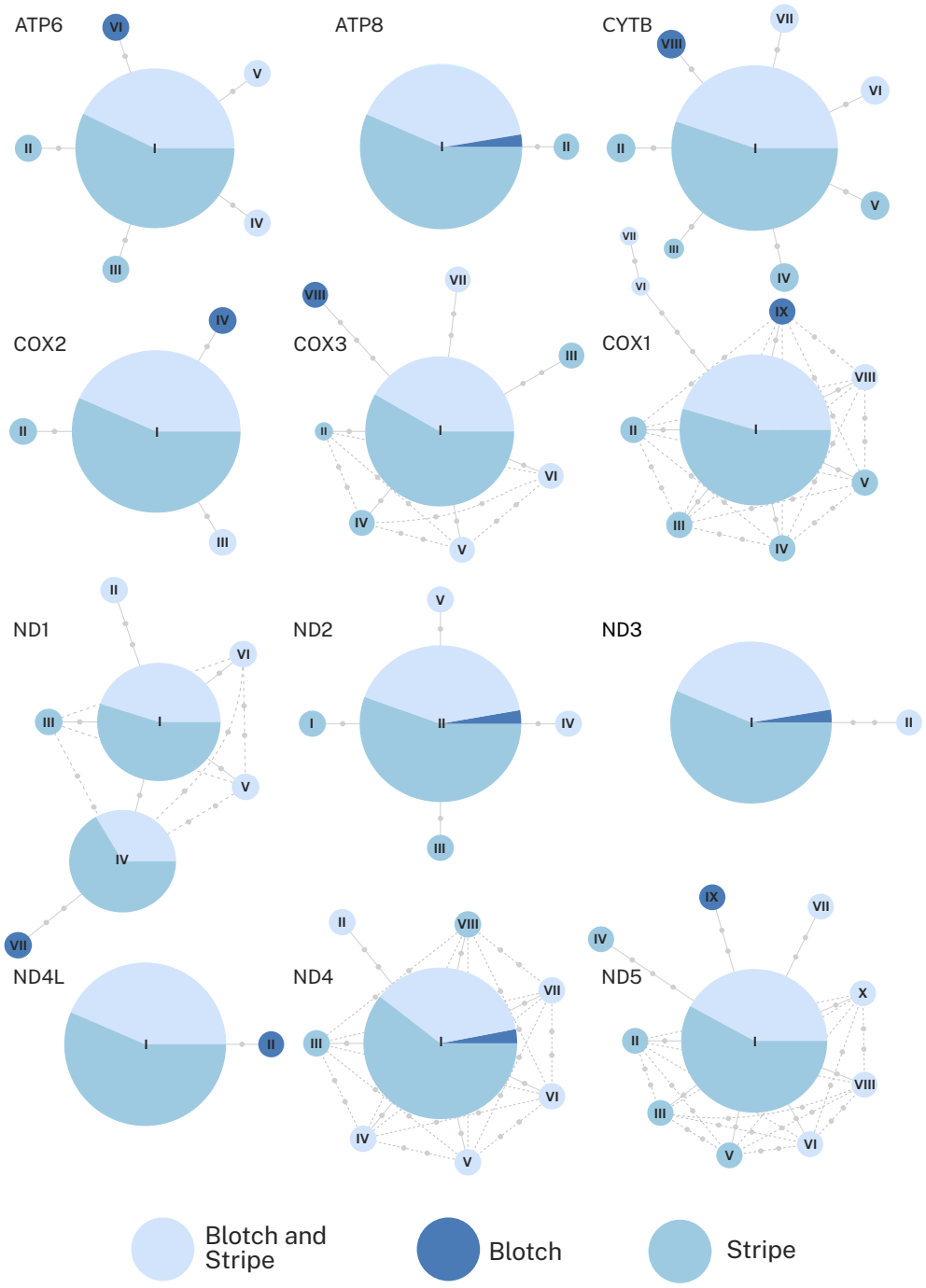


Figure 3.2 The mitochondrial haplotypes of *Corydoras fulleri.* Haplotypes were generated from both wild and aquarium individuals, characterized by the presence and absence of a blotch on the flank of the individual and/ or a stripe along the caudal peduncle. The Roman numerals indicate the number of haplotypes present, while the pie charts show the proportion of each haplotype represented by different phenotypes. Grey dots represent the number of mutations between haplotypes.

Stripe morphs. However, samples size was very low for the blotch morph and other individuals of either Blotch/Stripe or Stripe had similarly divergent haplotypes. The genes *atp8*, *nd3*, *nd2* and *nd4* showed the Blotch morph as sharing a haplotype with the majority of other individuals, which were either stripe or Blotch/Stripe morphs. In all genes, the majority of Blotch/Stripe and Stripe individuals, shared a haplotype. These data suggest that the different morphs represent a single polymorphic taxon.

3.4.2 Pairwise F_{ST}

The pairwise F_{ST}, estimated from nuclear SNPs, supported the mtDNA analyses (Table 3.1). F_{ST} was lower within populations between the Blotch/Stripe and Stripe morphs than it was between populations. This indicates lower genetic diversity within than between populations (Table 3.1). If separate taxa had hybridized, we would expect the opposite with lower F_{ST} among the same morph at different sites than between morphs at the same site.

Table 3.1 Pairwise estimates of Weir and Cockerham's F_{ST} between *Corydoras fulleri* populations, and between colour morphs within populations.

Comparison	Population	F_{ST}
Between Populations		
Cheese Creek- Alegría	-	0.073
Between Phenotypes		
Stripe - Blotch / Stripe	Cheese Creek	0.004
Stripe – Blotch / Stripe	Alegría	0.021

3.4.3 Principal Components Analysis

A total of 28 wild individuals were used to explore population structuring using PCA, with 11 from Cheese Creek (Blotch = 1, Striped = 3, Blotch / Stripe = 7), 12 from Alegría (Stripe = 9, Blotch / Stripe = 3) and 5 from unknown origins (Stripe = 2, Blotch=1, Blotch / Stripe = 2). The PCA resolved two clear clusters that were reflective of the two sampling sites with limited clustering among phenotypes within each site (Supplementary Figure 3.1). One sample was not assigned to either cluster but

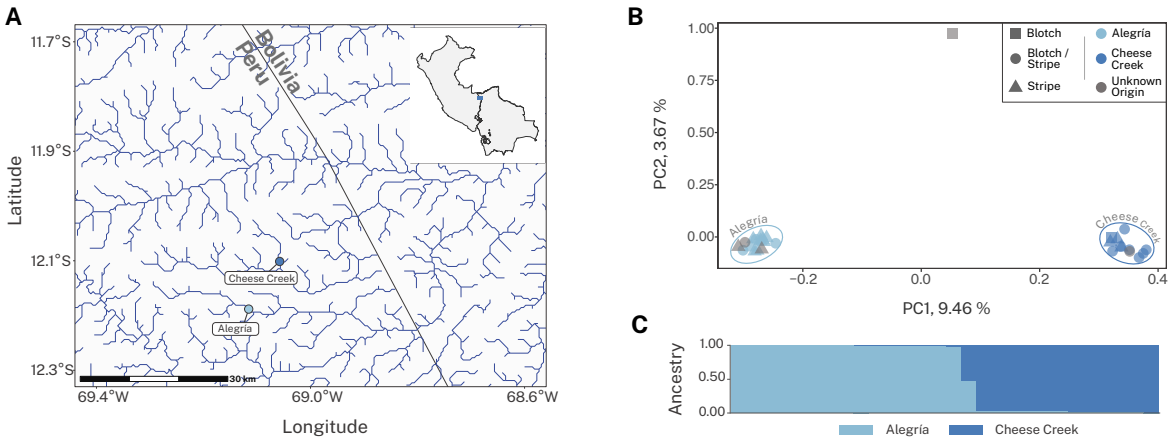


Figure 3.3 The population structuring between the sampling sites of *Corydoras fulleri*. **A** Map of the sampling sites Cheese Creek and Alegría, in the Madre de Dios region of Peru. **B** Principal Component Analysis showing the clustering of individuals between the two sampling populations, including samples with missing sampling location. Shape denotes the phenotype of the individual, categorised by the presence or absence of a blotch on the flank of the individual and/or a stripe along the caudal peduncle. **C** Shared ancestry of wild samples, with clustering K=2.

separated on both PC1 and PC2 (Figure 3.3B). This individual was likely to represent an F1 aquarium hybrid having been bred from individuals collected from the two sampled populations.

3.4.4 Model based clustering (Admixture)

A cluster (K) value of two was identified as best fitting the data, using the automatic selection tool in PCAngsd, and individuals showed clear separation into two clusters representing the two populations. There was no evidence of structure within a site representing colour morphs. Mirroring the PCA, the admixture plot identified a single individual as likely to be a hybrid between the two populations.

3.4.5 Identifying regions of the genome associated with pigmentation

To identify regions of genetic differentiation within the genome, we used a reduced sample of individuals representing the extremes of Stripe or Blotch / Stripe phenotypes rather than a continuous range of blotch variation. Samples of unknown origin that clustered with a specific population were incorporated into the population they clustered with. This included 13 individuals from Alegría (Blotch / Stripe = 3, Stripe = 10) and 8 individuals from Cheese Creek (Blotch / Stripe = 5, Stripe = 3). Windowed estimates of F_{ST} identified 42 outlier regions in Alegría and 33 in Cheese Creek. There were no shared regions between the two populations. Of these regions, 15 were associated with the presence or absence of a blotch in Alegría and 23 in Cheese Creek, after filtering using PCAdapt. Within these windows, 174 SNPs were found to be significant ($p < 0.05$) in Alegría and 269

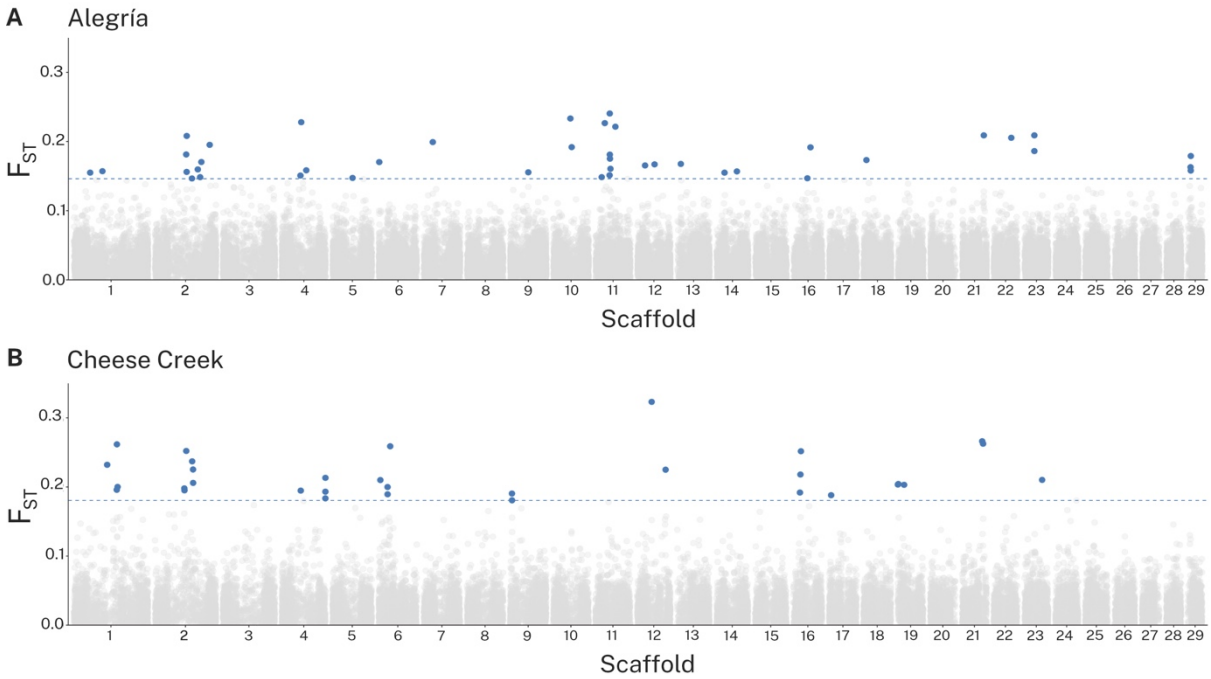


Figure 3.4 Windowed estimates of F_{ST} across the *Corydoras fulleri* genome. SNPs are in windows of 150 SNPs or 50kb, depending on which criteria is met first. Outlier windows are in the top 99.9th percentile, denoted by the dashed blue line. Outlier windows are shown in blue. **A** Alegría (Blotch / Stripe = 3, Stripe = 10) **B** Cheese Creek (Blotch / Stripe = 5, Stripe = 3).

Table 3.2 Candidate genes within 40kb of a SNP associated with the presence or absence of a blotch in *Corydoras fulleri*. Genes are either already known to play a role in pigmentation or represent likely candidates based on known functions.

Scaffold	Population	Protein	Gene	Biological Process Gene Ontology (GO)
Scaffold 2	Alegría	Phosphodiesterase	*pde11al*	cAMP-mediated signalling (GO:0019933) | negative regulation of cAMP-mediated signalling (GO:0043951) | negative regulation of cGMP-mediated signalling (GO:0010754)
Scaffold 11	Alegría	Integrin alpha-11	*itga11b*	cell adhesion mediated by integrin (GO:0033627) | cell-cell adhesion (GO:0098609) | cell-matrix adhesion (GO:0007160) | integrin-mediated signalling pathway (GO:0007229)
Scaffold 14	Alegría	ATPase, H+ transporting, lysosomal accessory protein 1-like precursor	*atp6ap1lb*	regulation of cellular pH (GO:0030641)
Scaffold 2	Cheese Creek	adenylate cyclase	*adcy3a*	adenylate cyclase-activating G protein-coupled receptor signalling pathway (GO:0007189) | cAMP biosynthetic process (GO:0006171) | intracellular signal transduction (GO:0035556) | melanosome localization (GO:0032400)
Scaffold 4	Cheese Creek	Serine/threonine-protein kinase receptor	*acvr2aa*	activin receptor signalling pathway (GO:0032924) | cartilage development (GO:0051216) | cellular response to growth factor stimulus (GO:0071363) | hindbrain development (GO:0030902) | neural crest cell migration (GO:0001755) | skeletal system development (GO:0001501)
Scaffold 6	Cheese Creek	Cyclic AMP-responsive element-binding protein 3-like protein 2	*creb3l2*	cartilage development (GO:0051216) | chondrocyte differentiation (GO:0002062) | regulation of ER to Golgi vesicle-mediated transport (GO:0060628) | regulation of extracellular matrix constituent secretion (GO:0003330) | regulation of transcription by RNA polymerase II (GO:0006357) | response to endoplasmic reticulum stress (GO:0034976) | response to unfolded protein (GO:0006986)
Scaffold 6	Cheese Creek	Rab-3A-interacting protein	*rab3ip*	exocytosis (GO:0006887) | Kupffer's vesicle development (GO:0070121) | melanosome transport (GO:0032402)
Scaffold 16	Cheese Creek	Unconventional myosin-Va isoform X3	*myo5aa*	actin filament organization (GO:0007015) | system development (GO:0048731) | vesicle transport along actin filament (GO:0030050)
Scaffold 19	Cheese Creek	Mast/stem cell growth factor receptor	*kitb*	B cell differentiation (GO:0030183) | embryonic hemopoiesis (GO:0035162) | Fc receptor signalling pathway (GO:0038093) | hematopoietic progenitor cell differentiation (GO:0002244) | Kit signalling pathway (GO:0038109) | positive regulation of cell migration (GO:0030335) | positive regulation of receptor signalling pathway via JAK-STAT (GO:0046427)

1

in Cheese Creek. This resulted in 29 candidate genes in Alegría and 57 in Cheese Creek (Table 3.2, Supplementary Table 3.1). Linkage was shown to decay to a background level of 0.05 at 40kb distance, so detecting genes a maximum of 40Kb from candidate SNPs is appropriate (Supplementary Figure 3.2). Of these genes, six have known functions in pigmentation. Additionally, two genes, *adcy3a* and *rab3ip* had biological processes GO terms associated with pigmentation. The gene adcy3a is currently not in the known pigmentation gene databases but has the GO term "melanosome localization (GO:0032400)".

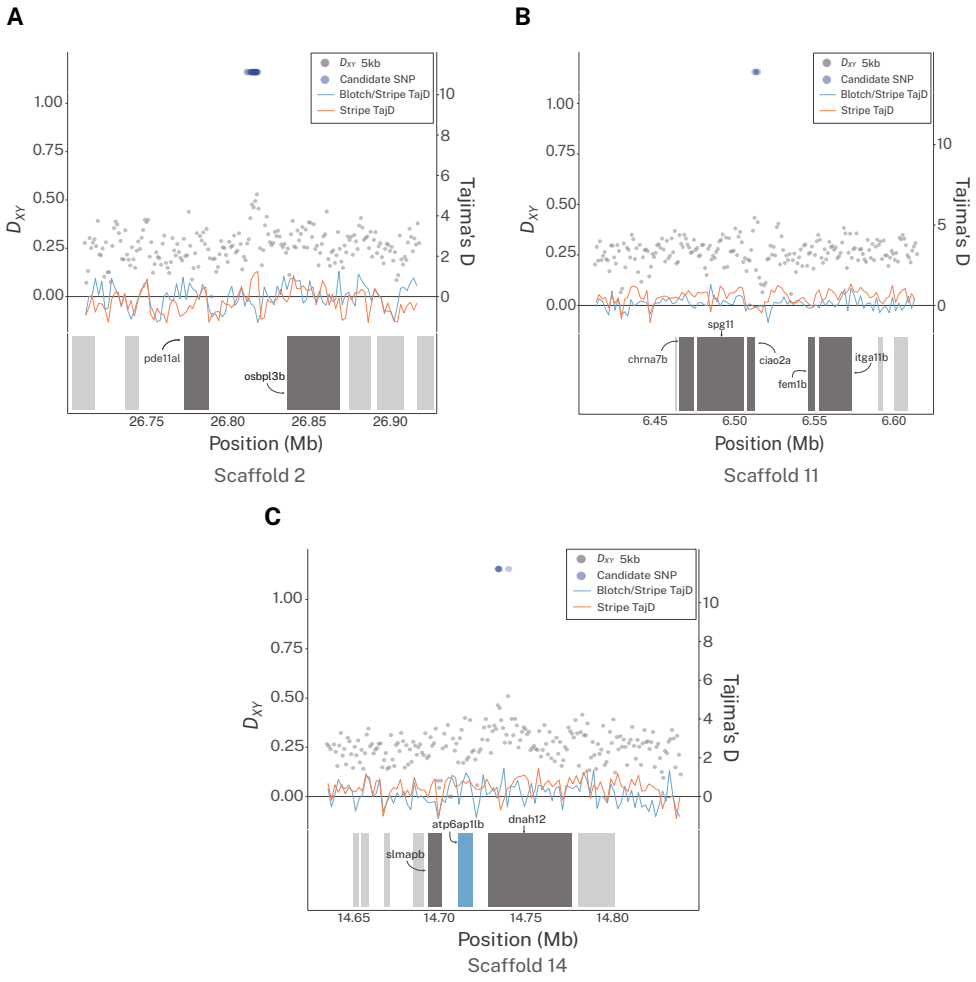


Figure 3.5 Estimates of D_{XY} across 5kb windows between Blotch / Stripe and Stripe morphs of *Corydoras fulleri* from Alegría. Estimates of Tajima's D in 2kb windows for each morph. Blue points denote candidate SNPs associated with the presence or absence of a blotch. The light grey bars are additional genes > 40kb from a candidate SNP. Dark grey bars denote genes < 40kb from a candidate SNP and blue bars represent pigment genes. **A** Scaffold 2, 2.6- 2.7 Mb **B** Scaffold 11, 6.4 – 6.65 Mb **C** Scaffold 14, 14.6 -14.9 Mb

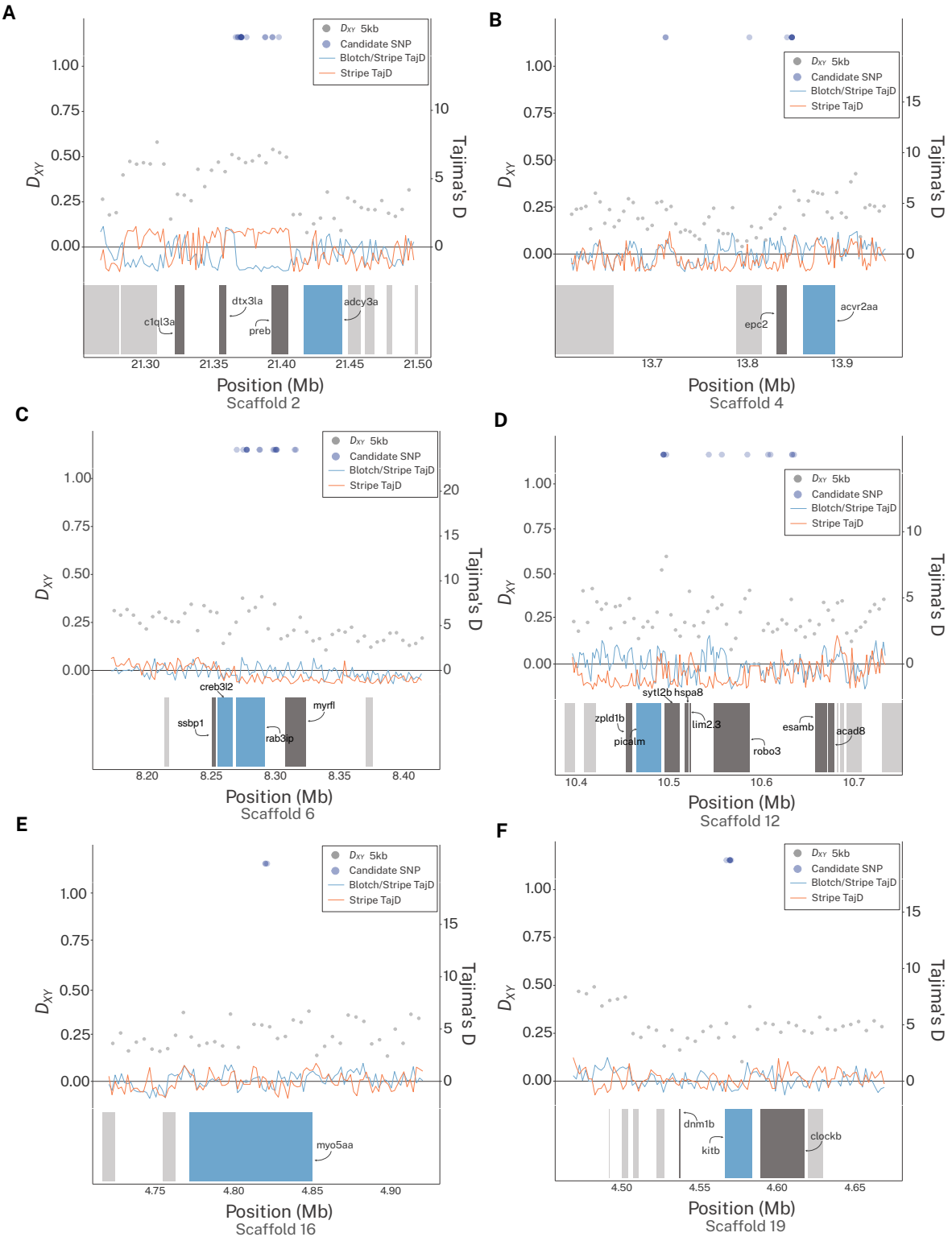


Figure 3.6 Estimates of D_{XY} across 5kb windows between Blotch/Stripe and Stripe morphs of *Corydoras fulleri* from Cheese Creek. Estimates of Tajima's D in 2kb windows for each morph. Blue points denote candidate SNPs associated with the presence or absence of a blotch. The light grey bars are additional genes > 40kb from a candidate SNP. Dark grey bars denote genes < 40kb from a candidate SNP and blue bars are pigment genes. **A** Scaffold 2, 21.25- 21.51Mb **B** Scaffold 4, 13.57- 14 Mb **C** Scaffold 6, 8.15-8.45 Mb **D** Scaffold 12, 10.35-10.8 Mb **E** Scaffold 16, 4.7-4.95 Mb **F** Scaffold 19, 4.45- 4.7 Mb.

All the pigment candidates, apart from *atp6ap1*, were identified in the Cheese Creek population. Gene ontology and KEGG pathway enrichment analysis revealed no significantly enriched terms in the candidate genes from Alegría but found the KEGG pathways for "Melanogenesis" (KEGG:04916) and "MAPK Signalling pathway" (KEGG:04010) were significantly enriched ($p < 0.05$) in the Cheese Creek candidate genes. Additionally, we estimated absolute genetic differentiation (D_{XY}) in non-overlapping 5kb windows across reduced regions in which our candidate genes occurred. In some regions, an increased genetic differentiation ($D_{XY} > 0.4$) close to candidate SNPs (Figure 3.5, Figure 3.6A, D) was observed. However, this was not universal, and some regions lacked elevated D_{XY} (Figure 3.6B-C, E-F). We also calculated Tajima's D in 2kb non-overlapping windows to indicate patterns and modes of selection. Whilst the overall patterns were complex, a negative Tajima's D is seen in one or both phenotypes in regions containing candidate SNPs (Figure 3.5, Figure 3.6). Where negative Tajima's D values are present in both phenotypes, this may indicate positive selection at a population scale but could also be indicative of population expansion. In regions where only the Blotch / Stripe or Stripe morphs have a negative Tajima's D could be indicative of positive selection occurring in one morph or changes in morph frequency within the population.

3.5 Discussion

Despite longstanding interest in the field of mimicry, our knowledge of the genetics underpinning the striking diversity in warning colouration in vertebrates is limited. In this study we introduce a new vertebrate model system, *C. fulleri,* to further understand the evolution of mimicry, in particular, polymorphic mimicry. This system can be utilised to explore whether patterns observed in Lepidoptera mimetic communities are widespread. To that end, firstly, using mtDNA sequencing we confirm that *C. fulleri* is indeed a single polymorphic mimetic species rather than a hybrid zone between two different taxa. Secondly, using low coverage resequencing and a genome scanning approach, we show regions associated with mimetic patterning are spread across the genome rather than being found in restricted tightly linked regions, as in Lepidoptera systems (Jiggins, 2016). Finally, we provide evidence to suggests this trait utilises established pigmentation pathways, specifically relating to the distribution and density of melanin.

3.5.1 Single polymorphic taxon or contact zone?

In this study, we show that *C. fulleri* Blotch / Stripe and Stripe morphs share mtDNA haplotypes. There was evidence to suggest the Blotch morph possessed some unique haplotypes. However, we had few representatives of the Blotch morph. Additionally, if Blotch and Stripe individuals did represent distinct taxa, and Blotch / Stripe individuals the intermediate hybrids, rare Blotch haplotypes would be

present within the Blotch / Stripe individuals. Thus, it is most likely that the rare blotch haplotype is a sampling artefact. Many other less common haplotypes are also separated by 2-5 mutations from the most common haplotypes, further supporting this hypothesis. These findings are reinforced by evidence from the pairwise F_{ST}. Genetic differentiation was low between the Stripe and Blotch / Stripe phenotypes. If the Blotch / Stripe phenotype had a hybrid origin, the pairwise F_{ST} would be elevated reflecting an equal contribution from both the Blotch and Stripe parental lineages. Moreover, we would expect differentiation into phenotypic groups on the PCA and Admixture analyses.

3.5.2 Nature of mimicry in the *C. fulleri* system

In the wild, *C. fulleri* is found in sympatry with three other Corydoradinae species, with whom one of the morphs share a resemblance. The Stripe morph shares a resemblance with the sympatric CW62 (*Hoplisoma sp*.), CW040 (*Brochis sp*.) and *Ho. cf. acrensis* (Tencatt *et al.*, 2021). Another taxon, *O. aeneus,* is also found at both sites. *Osteogaster aeneus* possess a large blotch on the flank and may represent a co-mimic with the *C. fulleri* Blotch morph. Locally stable polymorphic mimicry, where a single species diversified to participate in two mimicry rings in a single locality, has previously been identified in the Lepidopteran species *He. numata* (Merrill *et al.*, 2015; Jay *et al.*, 2018). Here intermediate forms between the two morphs are selected against, which is maintained via a supergene genetic architecture (Joron *et al.*, 2011; Jay *et al.*, 2018). Within the supergene, loci for colour patterns are closely linked and recombination across the region is reduced, via an inversion (Joron *et al.*, 2011; Jay *et al.*, 2018). Within *C. fulleri* Blotch / Stripe individuals there is a range of intermediate blotch sizes and differing degrees of melanism, suggesting that selection is not acting strongly against intermediates (Tencatt *et al.*, 2021). Additionally, the evidence in this study shows sites across the genome are associated with pigmentation, which is not indicative of a supergene (Jay *et al.*, 2018). However, this could also represent a very early stage of a complex mimetic interaction, perhaps due to a recent range expansion of a co-mimic. Alternatively, the frequency of co-mimics may change seasonally, which could result in changing dynamics of frequency dependent selection and could prevent the formation the genetic architecture required to prevent intermediate forms, although this remains to be seen.

In the butterfly *He. melpomene*, mimicry loci are tightly linked but not characterised by a supergene architecture (Joron *et al.*, 2006; Kronforst and Papa, 2015). This species, and its co-mimics *He. erato* or *He. timareta,* show polymorphism over a large geographic scale but are regionally monomorphic separated by small transitional regions where intermediate forms can be found (Mallet, 1999; Joron *et al.*, 2006; Mérot *et al.*, 2013). These hybrid zones are also observed in the poison arrow frog *Ranitomeya imitator*, which is also geographically polymorphic, forming mimetic interactions with

additional *Ranitomeya sp.* (Twomey *et al.*, 2016). Perhaps, the observed phenotypic variation in *C. fulleri* represents a hybrid zone. The observed genetic architecture of the blotch trait supports this hypothesis, with unlinked genes allowing for intermediate variation. This is congruent with recent work uncovering the genetic architecture of the *R. imitator* polymorphism, where many known pigment genes were found to be differentially expressed between morphs, although the genomic location of these genes was not noted (Rubio *et al.*, 2024; Stuckert *et al.*, 2024). A caveat to this hypothesis, is that hybrid zones tend to be narrow, as small as 10km in *Heliconius sp.* (Mallet *et al.*, 1990). All morphs are present in both Cheese Creek and Alegría, which are separated by 135 km of waterways (Figure 3.3A), indicating a much wider potential hybrid region. However, to fully reject this hypothesis, the range of *C. fulleri* needs to be better understood. Another explanation for the variation seen at a single site, could be that *C. fulleri* is not mimicking both striped and blotched taxa, but that one morph represents the ancestral patterning, and positive selection is driving the advergence on a single mimetic phenotype. The presence of a blotch is seen in sister taxa in neighbouring regions, indicating the blotch phenotype is more common than a stripe phenotype, for example CW080 (Fuller and Evers, 2005). This, however, could only be truly resolved by understanding ancestral *Corydoras* patterns and the timing of the *C. fulleri* arrival to the region, relative to its co-mimics (Wright, 2011; Muell *et al.*, 2022).

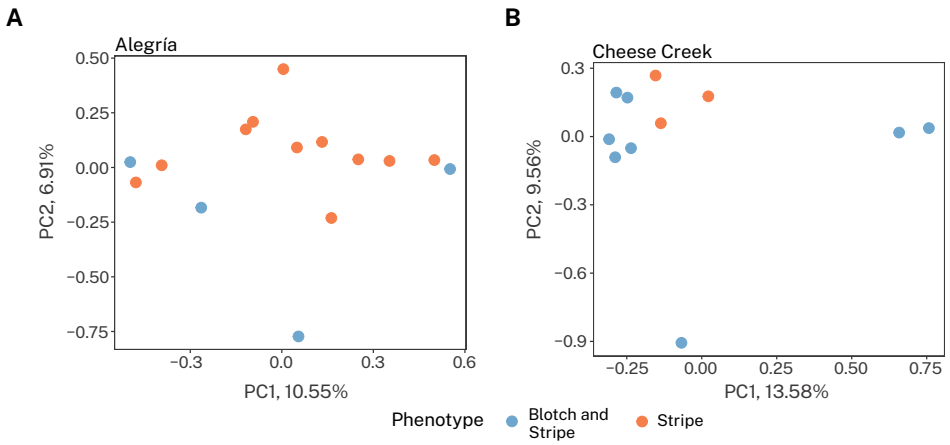
Within both populations we identified known pigment genes associated with increased relative genetic differentiation (F_{ST}), indicating variation in aposematic patterning within *C. fulleri* is utilising known pigmentation pathways. The melanophores are the melanin, black pigment, containing cells in teleosts (Parichy, 2021). All the pigment genes identified in this study were found to influence melanin pigmentation, primarily focusing on the amount of melanin and movement of melanosomes (Rawls and Johnson, 2000; O'Sullivan *et al.*, 2004; Baxter *et al.*, 2019). This indicates that both the amount of pigment and the location of the melanosomes are being targeted, resulting in either the presence or absence of the blotch phenotype. *Corydoras* fry possess a large melanic saddle shaped area on the flank at seven days old, which may either persist throughout development or disappear (Fuller, 1999). This "saddle" is also observed in *C. fulleri* (Tencatt *et al.*, 2021). Perhaps changes in the localization of melanin at this early stage could result in a reduction of the blotch in order to more closely resemble striped co-mimics. Within the Alegría population, the genes *itga11b* and *pde11al* may represent novel pigment genes. The gene *itag11b* encodes an integrin protein, a family that have been shown to play a role in melanocyte adhesion and migration (Pinon and Wehrle-Haller, 2011). Within the Cheese Creek population, the gene *itgb4* (Supplementary Table 3.1) has also been highlighted as a candidate, indicating that perhaps the integrin pathway is being targeted in both populations. Alternatively, *pde11al* is associated with negative regulation of the cAMP signalling

pathway. This pathway has been demonstrated to play a major role in pigmentation, with many genes in this pathway having already been implicated with pigmentation phenotypes (Parichy, 2021). Whilst these genes, as well as the other candidate pigment genes, show variable patterns of absolute genetic differentiation (D_{XY}) and deviation from neutral evolution (Tajima's D), there is some indication that selection could be occurring. However, to confirm the true nature of these patterns, demographic modelling and selection scans, to detect positive selection, would be required to make firm inferences (Fijarczyk and Babik, 2015; Koropoulis, Alachiotis and Pavlidis, 2020).

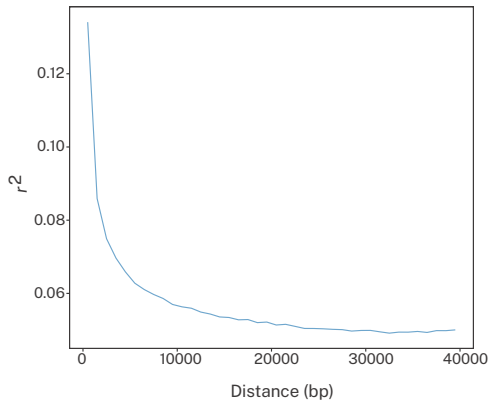
The identification of genes involved in melanin variation in both populations indicates some parallelism at a pathway level, however, we found no overlapping regions associated with the presence or absence of a blotch between the populations. This could be due to several factors. Firstly, we lacked the overalls sample sizes to utilise genome wide association approaches (GWAS) (Santure and Garant, 2018). These can account for population structuring and traits that are continuous in nature allowing for an overall more powerful approach (Santure and Garant, 2018; Lou *et al.*, 2021). The continuous variation and population structuring further reduced the number of individuals present to make inferences. Additionally, association approaches can be used to detect alleles associated with polygenic traits and detect the effect size of these traits (Barghi, Hermisson and Schlötterer, 2020; Fagny and Austerlitz, 2021). A more complex genetic architecture produces a lack of parallelism between populations (Barghi, Hermisson and Schlötterer, 2020). However, the cumulative allele frequencies can be used to demonstrate association (Cheng and Kirkpatrick, 2016, 2020). This represents a future direction for this work to better understand the architecture of the blotch trait, which is vital to understanding maintenance of polymorphism in this system.

Here we show the presence or absence of a blotch in the polymorphic *C. fulleri* is likely mediated by variation in genes related to melanin density and melanosome distribution. The identification of candidate genes in *C. fulleri* provides evidence for various hypotheses explaining the origin and maintenance of polymorphism in this system. Additionally, we show this is likely to be a polygenic trait involving many genes within the melanin related pathways. This is unlike evidence found in the *Heliconius* system, where few genes of large effect underpin aposematic coloration. The architecture in *C. fulleri* resembles that emerging from the *Ranitomeya* system highlighting the need to explore the genetics of mimicry in vertebrate systems, making our work on *C. fulleri* a very timely addition to our understanding of the genetics of mimicry.

3.6 Supplementary Materials



Supplementary Figure 3.1 Principal components one and two, showing structuring of Blotch/Stripe and Stripe *Corydoras fulleri* individuals within the **A** Alegría and **B** Cheese Creek populations. Estimated using low coverage whole genome sequence data.



Supplementary Figure 3.2 The average linkage decay across distances (bp) between two sites within the *Corydoras fulleri* genome. Estimated using low coverage whole genome sequence data.

Supplementary Table 3.1 Additional candidate genes within 40kb of a SNP associated with the presence or absence of a blotch in *Corydoras fulleri. G*enes do not play a known role in pigmentation.

Scaffold	Population	Protein	Gene
Scaffold 2	Alegría	AT-rich interactive domain-containing protein 1A	*arid1aa*
Scaffold 2	Alegría	Axonemal dynein light intermediate polypeptide 1	*dnali1*
Scaffold 2	Alegría	Nucleolar GTP-binding protein 2	*gnl2*
Scaffold 2	Alegría	GPN-loop GTPase 2	*gpn2*
Scaffold 2	Alegría	Potassium voltage-gated channel subfamily H member 2	*kcnh2b*
Scaffold 2	Alegría	Microtubule-associated protein	*LOC100001114*
Scaffold 2	Alegría	Oxysterol-binding protein	*osbpl3b*
Scaffold 2	Alegría	Smad nuclear-interacting protein	*snip1*
Scaffold 9	Alegría	Retinitis pigmentosa 9 protein	*rp9*
Scaffold 9	Alegría	UPF0711 protein C18orf21 homolog	*si:dkey-184p18.2*
Scaffold 11	Alegría	Neuronal acetylcholine receptor subunit alpha-7	*chrna7b*
Scaffold 11	Alegría	MIP18 family protein FAM96A	*ciao2a*
Scaffold 11	Alegría	Fem-1 homolog b	*fem1b*
Scaffold 11	Alegría	Protein kinase C-binding protein NELL2a	*nell2a*
Scaffold 11	Alegría	Spatacsin variant 1	*spg11*
Scaffold 12	Alegría	ADP-ribosylation factor interacting protein 2a	*arfip2a*
Scaffold 12	Alegría	Complement C4-B	*c4*
Scaffold 12	Alegría	FH2 domain-containing 4	*fhdc4*
Scaffold 12	Alegría	RING-type E3 ubiquitin transferase	*trim3b*
Scaffold 14	Alegría	Dynein heavy chain 12, axonemal isoform X2	*dnah12*
Scaffold 14	Alegría	Sarcolemma associated protein (Slmap protein)	*slmapb*
Scaffold 18	Alegría	C-type lectin lectoxin-Lio2-like	*LOC103911305*
Scaffold 21	Alegría	Cathepsin S, ortholog 1	*ctss1*
Scaffold 22	Alegría	Acetylcholine receptor subunit delta	*chrnd*
Scaffold 23	Alegría	Prostate stem cell antigen-like	*LOC103911143*
Scaffold 23	Alegría	Uncharacterized protein LOC110439392	*LOC110439392*
Scaffold 1	Cheese Creek	Interleukin 4 receptor, tandem duplicate 1	*il4r.1*
Scaffold 1	Cheese Creek	Protein FAM26E-like	*LOC110437808*
Scaffold 1	Cheese Creek	Methyl methanesulfonate-sensitivity protein 22-like	*mms22l*
Scaffold 1	Cheese Creek	Uncharacterized protein si:dkey-181m9.8	*si:dkey-181m9.8*
Scaffold 1	Cheese Creek	Uroporphyrinogen decarboxylase	*urod*
Scaffold 1	Cheese Creek	Zinc finger SWIM domain-containing protein 5	*zswim5*
Scaffold 2	Cheese Creek	Complement C1q-like protein 3	*c1ql3a*
Scaffold 2	Cheese Creek	C-myc promoter-binding protein	*dennd4a*

Scaffold 2	Cheese Creek	E3 ubiquitin-protein ligase	*dtx3la*
Scaffold 2	Cheese Creek	Gonadotropin releasing hormone receptor 1	*gnrhr1*
Scaffold 2	Cheese Creek	Ras GTPase-activating-like protein IQGAP3	*iqgap3*
Scaffold 2	Cheese Creek	Tetratricopeptide repeat protein 24	*LOC103910918*
Scaffold 2	Cheese Creek	Maestro heat-like repeat family member 1	*mroh1*
Scaffold 2	Cheese Creek	Prolactin regulatory element-binding	*preb*
Scaffold 2	Cheese Creek	Protogenin A	*prtga*
Scaffold 2	Cheese Creek	Scratch family zinc finger 1b	*scrt1b*
Scaffold 2	Cheese Creek	Basic helix-loop-helix transcription factor scleraxis-like	*scxb*
Scaffold 2	Cheese Creek	U3 small nucleolar RNA-associated protein 11	*utp11*
Scaffold 4	Cheese Creek	Enhancer of polycomb homolog	*epc2*
Scaffold 4	Cheese Creek	C-type lectin lectoxin-Lio2-like	*LOC103911305*
Scaffold 4	Cheese Creek	Secretory phospholipase A2 receptor-like isoform X1	*si:cabz01083838.1*
Scaffold 4	Cheese Creek	Si:dkey-88n24.7 (Uncharacterized protein LOC555676 isoform 1 precursor)	*si:dkey-88n24.7*
Scaffold 6	Cheese Creek	Calcium-activated potassium channel beta 2 subunit	*kcnmb2a*
Scaffold 6	Cheese Creek	Myelin regulatory factor-like protein	*myrfl*
Scaffold 6	Cheese Creek	protein-tyrosine-phosphatase	*ptprb*
Scaffold 6	Cheese Creek	Single-stranded DNA-binding protein, mitochondrial	*ssbp1*
Scaffold 9	Cheese Creek	C-type lectin lectoxin-Lio2-like	*LOC103911305*
Scaffold 9	Cheese Creek	Gastrula zinc finger protein XlCGF28.1-like	*LOC110439387*
Scaffold 12	Cheese Creek	Acyl-CoA dehydrogenase family, member 8	*acad8*
Scaffold 12	Cheese Creek	Endothelial cell-selective adhesion molecule	*esamb*
Scaffold 12	Cheese Creek	Heat shock cognate 71 kDa protein (Heat shock protein 8)	*hspa8*
Scaffold 12	Cheese Creek	Lens intrinsic membrane protein 2.3	*lim2.3*
Scaffold 12	Cheese Creek	phosphatidylinositol-binding clathrin assembly protein b	*picalmb*
Scaffold 12	Cheese Creek	Roundabout, axon guidance receptor, homolog 3	*robo3*
Scaffold 12	Cheese Creek	Synaptotagmin-like 2b	*sytl2b*
Scaffold 12	Cheese Creek	Zona pellucida-like domain-containing protein 1	*zpld1b*
Scaffold 16	Cheese Creek	Otogelin-like protein isoform	*otogl*
Scaffold 16	Cheese Creek	protein-tyrosine-phosphatase	*ptprq*
Scaffold 19	Cheese Creek	Circadian locomoter output cycles protein kaput	*clockb*
Scaffold 19	Cheese Creek	dynamin GTPase	*dnm1b*
Scaffold 19	Cheese Creek	Uncharacterized protein LOC108190180	*LOC108190180*
Scaffold 21	Cheese Creek	NLR family CARD domain-containing protein 3-like	*LOC100536939*
Scaffold 21	Cheese Creek	Protein NLRC5	*LOC568664*

Scaffold 23	Cheese Creek	Evolutionarily conserved signalling intermediate in Toll pathway, mitochondrial	*ecsit*
Scaffold 23	Cheese Creek	Intraflagellar transport protein 25 homolog	*hspb11*
Scaffold 23	Cheese Creek	Integrin beta	*itgb4*
Scaffold 23	Cheese Creek	methylcrotonoyl-CoA carboxylase	*mccc2*
Scaffold 23	Cheese Creek	Disabled homolog 2 isoform X3	*si:ch211-204c21.1*
Scaffold 23	Cheese Creek	Survival motor neuron protein 1	*smn1*

3.7 Attribution Statement

Project conception by Martin Taylor and Emily Phelps. Sample collection by Martin Taylor, Emily Phelps and Chris Butler. All wet laboratory work and analysis was performed by Emily Phelps. Supervision by Martin Taylor.

3.8 References

Alexandrou, M.A., Oliveira, C., Maillard, M., *et al.* (2011) Competition and phylogeny determine community structure in Mullerian co-mimics, *Nature*, 469(7328), pp. 84–88. https://doi.org/10.1038/nature09660.

Al-Nakeeb, K., Petersen, T.N. and Sicheritz-Pontén, T. (2017) Norgal: extraction and de novo assembly of mitochondrial DNA from whole-genome sequencing data, *BMC Bioinformatics*, 18(1). https://doi.org/10.1186/s12859-017-1927-y.

Andrews, S. (2010) *FastQC: A Quality Control Tool for High Throughput Sequence*. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Barghi, N., Hermisson, J. and Schlötterer, C. (2020) Polygenic adaptation: a unifying framework to understand positive selection, *Nature Reviews. Genetics*, 21(12), pp. 769–781. https://doi.org/10.1038/s41576-020-0250-z.

Bates, H.W. (1862) Contributions to an insect fauna of the Amazon valley. Lepidoptera: heliconinae, *Journal Of The Proceedings Of The Linnean Society Of London Zoology*, 6(22), pp. 73–77. https://doi.org/10.1111/j.1096-3642.1862.tb00932.x.

Baxter, L.L., Watkins-Chow, D.E., Pavan, W.J., *et al.* (2019) A curated gene list for expanding the horizons of pigmentation biology, *Pigment Cell & Melanoma Research*, 32(3), pp. 348–358. https://doi.org/10.1111/pcmr.12743.

Camacho, C., Coulouris, G., Avagyan, V., *et al.* (2009) BLAST+: architecture and applications, *BMC Bioinformatics*, 10. https://doi.org/10.1186/1471-2105-10-421.

Caro, T. (2017) Wallace on coloration: Contemporary perspective and unresolved insights, *Trends In Ecology & Evolution*, 32(1), pp. 23–30. https://doi.org/10.1016/j.tree.2016.10.003.

Chatelain, P., Elias, M., Fontaine, C., *et al.* (2023) Müllerian mimicry among bees and wasps: a review of current knowledge and future avenues of research, *Biological Reviews Of The Cambridge Philosophical Society*, 98(4), pp. 1310–1328. https://doi.org/10.1111/brv.12955.

Cheng, C. and Kirkpatrick, M. (2016) Sex-specific selection and sex-biased gene expression in humans and flies, *PLoS Genetics*, 12(9). https://doi.org/10.1371/journal.pgen.1006170.

Cheng, C. and Kirkpatrick, M. (2020) The signal of sex-specific selection in humans is not an artefact: Reply to Mank et al, *Molecular Ecology*, pp. 1406–1407. https://doi.org/10.1111/mec.15420.

Chouteau, M., Llaurens, V., Piron-Prunier, F., *et al.* (2017) Polymorphism at a mimicry supergene maintained by opposing frequency-dependent selection pressures, *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 114(31), pp. 8325–8329. https://doi.org/10.1073/pnas.1702482114.

Danecek, P., Bonfield, J.K., Liddle, J., *et al.* (2021) Twelve years of SAMtools and BCFtools, *GigaScience*, 10(2). https://doi.org/10.1093/gigascience/giab008.

Darwin, C. (1859) *On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life*. London : John Murray, 1859.

Dias, A.C., Tencatt, L.F.C., Roxo, F.F., *et al.* (2024) Phylogenomic analyses in the complex Neotropical subfamily Corydoradinae (Siluriformes: Callichthyidae) with a new classification based on morphological and molecular data, *Zoological Journal Of The Linnean Society* .

Ezray, B.D., Wham, D.C., Hill, C.E., *et al.* (2019) Unsupervised machine learning reveals mimicry complexes in bumblebees occur along a perceptual continuum, *Proceedings Biological Sciences*, 286(1910). https://doi.org/10.1098/rspb.2019.1501.

Fagny, M. and Austerlitz, F. (2021) Polygenic Adaptation: Integrating Population Genetics and Gene Regulatory Networks, *Trends In Genetics*, 37(7), pp. 631–638. https://doi.org/10.1016/j.tig.2021.03.005.

Fijarczyk, A. and Babik, W. (2015) Detecting balancing selection in genomes: limits and prospects, *Molecular Ecology*, 24(14), pp. 3529–3545. https://doi.org/10.1111/mec.13226.

Fox, E.A., Wright, A.E., Fumagalli, M., *et al.* (2019) ngsLD: evaluating linkage disequilibrium using genotype likelihoods, *Bioinformatics*35(19), pp. 3855–3856. https://doi.org/10.1093/bioinformatics/btz200.

Fuller, I. (1999) The fry patterns in the South-American catfish genus Corydoras (Pisces, Siluriformes, Callichthyidae), *Bulletin Zoologisch Museum* .

Fuller, I.A.M. and Evers, H.E. (2005) *Identifying Corydoradinae Catfishe*. Kidderminster, England: Ian Fuller Enterprises.

Fumagalli, M., Vieira, F.G., Linderoth, T., *et al.* (2014) ngsTools: methods for population genetics analyses from next-generation sequencing data, *Bioinformatics* , 30(10), pp. 1486–1487. https://doi.org/10.1093/bioinformatics/btu041.

Garrison, E. and Marth, G. (2012) Haplotype-based variant detection from short-read sequencing, 1207. http://dx.doi.org/

Greven, H., Flasbeck, T. and Passia, D. (2006) Axillary glands in the armoured catfish Corydoras aeneus (Callichthyidae, Siluriformes), *Verhandlungen Der Gesellschaft Für Ichthyologie*, 5, pp. 65–69.

Holmes, I.A., Grundler, M.R. and Davis Rabosky, A.R. (2017) Predator perspective drives geographic variation in frequency-dependent polymorphism, *The American Naturalist*, 190(4), pp. 78–93. https://doi.org/10.1086/693159.

Jay, P., Whibley, A., Frézal, L., *et al.* (2018) Supergene evolution triggered by the introgression of a chromosomal inversion, *Current Biology*, 28(11), pp. 1839-1845 https://doi.org/10.1016/j.cub.2018.04.072.

Jiggins, C.D. (2016) Genes on the wing: colour pattern genetics, in *The Ecology And Evolution Of Heliconius Butterflies*. Oxford University Press, pp. 112–137. https://doi.org/10.1093/acprof:oso/9780199566570.003.0008.

Joron, M., Papa, R., Beltrán, M., *et al.* (2006) A conserved supergene locus controls colour pattern diversity in Heliconius butterflies, *PLoS Biology*, 4(10). https://doi.org/10.1371/journal.pbio.0040303.

Joron, M., Frezal, L., Jones, R.T., *et al.* (2011) Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry, *Nature*, 477(7363), pp. 203–206. https://doi.org/10.1038/nature10341.

Jun, G., Wing, M.K., Abecasis, G.R., *et al.* (2015) An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data, *Genome Research*, 25(6), pp. 918–925. https://doi.org/10.1101/gr.176552.114.

Katoh, K. and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability, *Molecular Biology And Evolution*, 30(4), pp. 772–780. https://doi.org/10.1093/molbev/mst010.

Kolberg, L., Raudvere, U., Kuzmin, I., *et al.* (2023) g:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update), *Nucleic Acids Research*, 51(W1), pp. 207–212. https://doi.org/10.1093/nar/gkad347.

Korneliussen, T.S., Albrechtsen, A. and Nielsen, R. (2014) ANGSD: Analysis of Next Generation Sequencing Data, *BMC Bioinformatics*, 15(356). https://doi.org/10.1186/s12859-014-0356-4.

Koropoulis, A., Alachiotis, N. and Pavlidis, P. (2020) Detecting Positive Selection in Populations Using Genetic Data, in J.Y. Dutheil (ed.) *Statistical Population Genomics*. New York, NY: Springer US, pp. 87–123. https://doi.org/10.1007/978-1-0716-0199-0_5.

Korunes, K.L. and Samuk, K. (2021) pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data, *Molecular Ecology Resources*, 21(4), pp. 1359–1368. https://doi.org/10.1111/1755-0998.13326.

Kronforst, M.R. and Papa, R. (2015) The functional basis of wing patterning in Heliconius butterflies: the molecules behind mimicry, *Genetics*, 200(1), pp. 1–19. https://doi.org/10.1534/genetics.114.172387.

Lawrence, M., Huber, W., Pagès, H., *et al.* (2013) Software for computing and annotating genomic ranges, *PLoS Computational Biology*, 9(8). https://doi.org/10.1371/journal.pcbi.1003118.

Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform, *Bioinformatics*, 25(14), pp. 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Lorin, T., Brunet, F.G., Laudet, V., *et al.* (2018) 1 Teleost fish-specific preferential retention of pigmentation gene-containing families after 2 whole genome duplications in vertebrates, *G3* , 8(5), pp. 1795–1806. https://doi.org/10.1534/g3.118.200201.

Lou, R.N., Jacobs, A., Wilder, A.P., *et al.* (2021) A beginner's guide to low-coverage whole genome sequencing for population genomics, *Molecular Ecology*, 30(23), pp. 5966–5993. https://doi.org/10.1111/mec.16077.

Mallet, J., Barton, N., Lamas, G., *et al.* (1990) Estimates of selection and gene flow from measures of cline width and linkage disequilibrium in heliconius hybrid zones, *Genetics*, 124(4), pp. 921–936. https://doi.org/10.1093/genetics/124.4.921.

Mallet, J. (1999) Causes and Consequences of a Lack of Coevolution in Müllerian mimicry, *Evolutionary Ecology*, 13(7), pp. 777–806. https://doi.org/10.1023/A:1011060330515.

Mallet, J. and Joron, M. (1999) Evolution of Diversity in Warning Color and Mimicry: Polymorphisms, Shifting Balance, and Speciation, *Annual Review Of Ecology And Systematics*, 30(1), pp. 201–233. https://doi.org/10.1146/annurev.ecolsys.30.1.201.

Marek, P.E. and Bond, J.E. (2009) A Müllerian mimicry ring in Appalachian millipedes, *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 106(24), pp. 9755–9760. https://doi.org/10.1073/pnas.0810408106.

Martin, C.A. (2017) ggConvexHull: Add a convex hull geom to ggplot2. Available at: http://github.com/cmartin/ggConvexHull.

Martin, M., Ebert, P. and Marschall, T. (2023) Read-based phasing and analysis of phased variants with WhatsHap, *Methods In Molecular Biology*, 2590, pp. 127–138. https://doi.org/10.1007/978-1-0716-2819-5_8.

Meisner, J. and Albrechtsen, A. (2018) Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data, *Genetics*, 210(2), pp. 719–731. https://doi.org/10.1534/genetics.118.301336.

Mérot, C., Mavárez, J., Evin, A., *et al.* (2013) Genetic differentiation without mimicry shift in a pair of hybridizing*Heliconius*species (Lepidoptera: Nymphalidae): Mimicry and Hybridization in Butterflies, *Biological Journal Of The Linnean Society. Linnean Society Of London*, 109(4), pp. 830–847. https://doi.org/10.1111/bij.12091.

Mérot, C., Berdan, E.L., Cayuela, H., *et al.* (2021) Locally adaptive inversions modulate genetic variation at different geographic scales in a seaweed fly, *Molecular Biology And Evolution*, 38(9), pp. 3953–3971. https://doi.org/10.1093/molbev/msab143.

Merrill, R.M., Dasmahapatra, K.K., Davey, J.W., *et al.* (2015) The diversification of Heliconius butterflies: what have we learned in 150 years?, *Journal Of Evolutionary Biology*, 28(8), pp. 1417–1438. https://doi.org/10.1111/jeb.12672.

Motyka, M., Kampova, L. and Bocak, L. (2018) Phylogeny and evolution of Müllerian mimicry in aposematic Dilophotes: evidence for advergence and size-constraints in evolution of mimetic sexual dimorphism, *Scientific Reports*, 8(1), pp. 1–10. https://doi.org/10.1038/s41598-018-22155-6.

Muell, M.R., Chávez, G., Prates, I., *et al.* (2022) Phylogenomic analysis of evolutionary relationships in Ranitomeya poison frogs (Family Dendrobatidae) using ultraconserved elements, *Molecular Phylogenetics And Evolution*, 168(107389). https://doi.org/10.1016/j.ympev.2022.107389.

Müller, F. (1878) Über die vortheile der mimicry bei schmetterlingen, *Zoologischer Anzeiger*, 1, pp. 4–55.

Nishikawa, H., Iijima, T., Kajitani, R., *et al.* (2015) A genetic mechanism for female-limited Batesian mimicry in Papilio butterfly, *Nature Genetics*, 47(4), pp. 405–409. https://doi.org/10.1038/ng.3241.

O'Sullivan, T.N., Wu, X.S., Rachel, R.A., *et al.* (2004) dsu functions in a MYO5A-independent pathway to suppress the coat color of dilute mice, *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 101(48), pp. 16831–16836. https://doi.org/10.1073/pnas.0407339101.

Paradis, E. (2010) pegas: an R package for population genetics with an integrated-modular approach, *Bioinformatics* , 26(3), pp. 419–420. https://doi.org/10.1093/bioinformatics/btp696.

Paradis, E. and Schliep, K. (2019) ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R, *Bioinformatics* , 35(3), pp. 526–528. https://doi.org/10.1093/bioinformatics/bty633.

Parichy, D.M. (2021) Evolution of pigment cells and patterns: recent insights from teleost fishes, *Current Opinion In Genetics & Development*, 69, pp. 88–96. https://doi.org/10.1016/j.gde.2021.02.006.

Pasteur, G. (1982) A classificatory review of mimicry systems, *Annual Review Of Ecology And Systematics*, 13(1), pp. 169–199. https://doi.org/10.1146/annurev.es.13.110182.001125.

Pinon, P. and Wehrle-Haller, B. (2011) Integrins: versatile receptors controlling melanocyte adhesion, migration and proliferation: Integrin function in melanocytes, *Pigment Cell & Melanoma Research*, 24(2), pp. 282–294. https://doi.org/10.1111/j.1755-148X.2010.00806.x.

Privé, F., Luu, K., Vilhjálmsson, B.J., *et al.* (2020) Performing Highly Efficient Genome Scans for Local Adaptation with R Package pcadapt Version 4, *Molecular Biology And Evolution*, 37(7), pp. 2153–2154. https://doi.org/10.1093/molbev/msaa053.

Purcell, S., Neale, B., Todd-Brown, K., *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses, *American Journal Of Human Genetics*, 81(3), pp. 559–575. https://doi.org/10.1086/519795.

Rawls, J.F. and Johnson, S.L. (2000) Zebrafish kit mutation reveals primary and secondary regulation of melanocyte development during fin stripe regeneration, *Development* , 127(17), pp. 3715–3724. Available at: https://www.ncbi.nlm.nih.gov/pubmed/10934016.

Rubio, A.O., Stuckert, A.M.M., Geralds, B., *et al.* (2024) What makes a mimic? Orange, red, and black color production in the mimic poison frog (Ranitomeya imitator), *Genome Biology And Evolution*, 16(7). https://doi.org/10.1093/gbe/evae123.

Ruxton, G.D., Allen, W.L., Sherratt, T.N., *et al.* (2018) *Avoiding attack : the evolutionary ecology of crypsis, warning signals, and mimicry*. Oxford, United Kingdom: Oxford University Press.

Santure, A.W. and Garant, D. (2018) Wild GWAS—association mapping in natural populations, *Molecular Ecology Resources*, 18(4), pp. 729–738. https://doi.org/10.1111/1755-0998.12901.

Stuckert, A.M.M., Chouteau, M., McClure, M., *et al.* (2024) The genomics of mimicry: Gene expression throughout development provides insights into convergent and divergent phenotypes in a Müllerian mimicry system, *Molecular Ecology*, 33(14). https://doi.org/10.1111/mec.17438.

Symula, R., Schulte, R. and Summers, K. (2001) Molecular phylogenetic evidence for a mimetic radiation in Peruvian poison frogs supports a Müllerian mimicry hypothesis, *Proceedings Of The Royal Society Of London. Series B*, 268(1484), pp. 2415–2421. https://doi.org/10.1098/rspb.2001.1812.

Tencatt, L.F.C., Dos Santos, S.A., Evers, H.-G., *et al.* (2021) Corydoras fulleri (Siluriformes: Callichthyidae), a new catfish species from the rio Madeira basin, Peru, *Journal Of Fish Biology*, 99(2), pp. 614–628. https://doi.org/10.1111/jfb.14750.

Tencatt, L.F.C., do Couto, O.L.P., Santos, S.A., *et al.* (2024) A new long-snouted Corydoras (Siluriformes: Callichthyidae) from the rio Xingu and rio Tapajós basins, Brazilian Amazon, *Neotropical Ichthyology*. https://doi.org/10.1590/1982-0224-2023-0112.

Tencatt, L.F.C. and Ohara, W.M. (2016) Two new species of Corydoras Lacépède, 1803 (Siluriformes: Callichthyidae) from the rio Madeira basin, Brazil, *Neotropical Ichthyology: Official Journal Of The Sociedade Brasileira De Ictiologia*, 14(1). https://doi.org/10.1590/1982-0224-20150063.

Twomey, E., Vestergaard, J.S., Venegas, P.J., *et al.* (2016) Mimetic divergence and the speciation continuum in the mimic poison frog Ranitomeya imitator, *The American Naturalist*, 187(2), pp. 205–224. https://doi.org/10.1086/684439.

Twomey, E., Kain, M., Claeys, M., *et al.* (2020) Mechanisms for Color Convergence in a Mimetic Radiation of Poison Frogs, *The American Naturalist*, 195(5), pp. 132–149. https://doi.org/10.1086/708157.

Twomey, E., Vestergaard, J.S. and Summers, K. (2014) Reproductive isolation related to mimetic divergence in the poison frog Ranitomeya imitator, *Nature Communications*, 5(1), pp. 1–8. https://doi.org/10.1038/ncomms5749.

UniProt Consortium (2021) UniProt: the universal protein knowledgebase in 2021, *Nucleic Acids Research*, 49(1). https://doi.org/10.1093/nar/gkaa1100.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., *et al.* (2013) From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline, *Current Protocols In Bioinformatics*, 43(1). https://doi.org/10.1002/0471250953.bi1110s43.

Wallace, A.R. (1877) The colors of animals and plants, *The American Naturalist*, 11(11), pp. 641–662. https://doi.org/10.1086/271979.

Weir, B.S. and Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure, *Evolution*, 38(6), pp. 1358–1370. https://doi.org/10.2307/2408641.

Wickham, H. (2016) ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. Available at: https://ggplot2.tidyverse.org.

Wickham, H., Averick, M., Bryan, J., *et al.* (2019) Welcome to the tidyverse, *Journal Of Open Source Software*, 4(43). https://doi.org/10.21105/joss.01686.

Wilson, J.S., Williams, K.A., Forister, M.L., *et al.* (2012) Repeated evolution in overlapping mimicry rings among North American velvet ants, *Nature Communications*, 3(1), pp. 1–7. https://doi.org/10.1038/ncomms2275.

Wilson, J.S., Pan, A.D., Limb, E.S., *et al.* (2018) Comparison of African and North American velvet ant mimicry complexes: Another example of Africa as the 'odd man out', *PloS One*, 13(1). https://doi.org/10.1371/journal.pone.0189482.

Wright, J.J. (2011) Conservative coevolution of müllerian mimicry in a group of rift lake catfish, *Evolution*, 65(2), pp. 395–407. https://doi.org/10.1111/j.1558-5646.2010.01149.x.

4 Quantifying venom strength and venom candidate gene expression to better understand the evolution of mimicry in the Corydoradinae.

4.1 Abstract

Different types of defensive colour pattern mimicry are distinguished by the relative unpalatability of participating taxa. Batesian mimicry is characterised by a palatable taxon evolving to resemble an unpalatable taxon whereas Müllerian mimicry is characterised by two unpalatable taxa evolving to share a resemblance. Only by understanding the relative toxicity of participating taxa, is it possible to fully understand the nature of the mimetic interaction. Here we explore venom potency between two genera of Corydoradinae, *Corydoras* and *Hoplisoma,* which commonly form mimetic communities in South American rivers. We found a significant difference between the toxicity of venom gland tissue extracts in comparison to muscle tissue extracts but found no significant difference in venom potency between the genera using brine shrimp toxicity assays, suggesting that they are Müllerian mimics. We also used RNA sequencing to identify candidate venom genes and venom housekeeping genes in the Corydoradinae, which allowed a comparison between the *Corydoras sp* and *Hoplisoma sp* venom gene expression. We did not identify large logfold changes between the two genera in venom gene expression which supports the venom potency assay. This work contributes to the current understanding of the genetics of mimicry, focusing specifically on unpalatability, which is key to understanding the ecological interaction between members of mimetic communities.

4.2 Introduction

The protective mimicry phenotype is underpinned by a degree of unpalatability, often toxicity (ingestion) or venom (injection). As such, unpalatability is a key but understudied component of mimicry (Ruxton *et al.*, 2018). It is palatability that distinguishes Batesian Mimicry, where a palatable species mimics a non-palatable species, from Müllerian mimicry, where both species are unpalatable (Bates, 1862; Mallet and Joron, 1999; Ruxton *et al.*, 2018). Batesian mimicry is considered to be a parasitic interaction, where the palatable species benefits from the protection of the aposematic warning colouration of the unpalatable model (Bates, 1862; Mallet and Joron, 1999; Ruxton *et al.*, 2018). In contrast, Müllerian mimicry is considered to be a mutualistic interaction where both taxa

benefit through reduced cost of predator education (Müller, 1878; Ruxton *et al.*, 2018). Under Müller's original theory, co-mimics were considered equally unpalatable (Müller, 1878; Ruxton *et al.*, 2018). However, one co-mimic may be more toxic/unpalatable than the other and this theoretically changes the mutualistic relationship to a parasitic one (quasi-Batesian mimicry), as the less toxic species erodes the protection of the more toxic species (Speed, 1993; Aubier, Joron and Sherratt, 2017; Ruxton *et al.*, 2018). This suggests protective mimicry may be a continuum between Batesian and Müllerian mimicry rather than a strict dichotomy. In support of this, in 93% of theoretical predator aversion learning models, the more palatable species was shown to be a parasite of the less palatable species (Speed and Turner, 1999). Currently, examples of this phenomenon are lacking in nature, although they are somewhat supported by lab-based experiments (Rowland *et al.*, 2010). For example, artificial prey experiments, using Great tits as predators, showed as unequally defended mimics increased, relative per capita predation on the mimic-model pair increased (Rowland *et al.*, 2010). Conversely, when mimics were rare, they posed no evolutionary pressure on the model (Rowland *et al.*, 2010). Additionally, the level of unpalatability is also influenced by the predator's condition, including starvation level or existing toxin burden of the individual (Barnett, Bateson and Rowe, 2007; Sherratt, Roberts and Kassen, 2009; Rowland *et al.*, 2010). Due to this state dependence, quasi-Batesian mimicry can be considered as intermittently mutualistic or parasitic (Ruxton *et al.*, 2018).

Mimetic interactions are proposed to evolve via two routes: i) advergence- where one taxon evolves to resemble another or ii) convergence- where both taxa evolve a shared phenotype (Ruxton *et al.*, 2018). While Batesian (and similarly quasi-Batesian) mimicry are thought to evolve through advergence, the steps in the evolution of Müllerian mimicry are harder to decipher (Ruxton *et al.*, 2018). Most evidence supports advergence but an initial advergence followed by convergence could also result in Müllerian resemblance (Symula, Schulte and Summers, 2001; Marek and Bond, 2009; Merrill *et al.*, 2015; Motyka, Kampova and Bocak, 2018; Ruxton *et al.*, 2018). The nature of the mimetic interaction, or location on the mimicry continuum, could provide an important insight to the evolution of a mimetic system. Therefore, an important step in understanding the nature of a mimetic interaction is investigating differences in unpalatability between or among co-mimics.

Unpalatability in fish is primarily through venom (but see Caley and Schulter (2003)), which is a substance secreted from a specialised gland in one animal and delivered to a target through the infliction of a wound (Smith and Wheeler, 2006; Casewell *et al.*, 2013). The evolution of piscine venom is thought to have occurred via the co-option of ichthyocrinotoxins, present in the mucus secretions of the skin, which have immunity functions and the venom glands from epidermal cells that

secrete these toxins (Cameron and Endean, 1973; Gratzer *et al.*, 2015; Harris and Jenner, 2019). Over 2,900 cartilaginous and ray-finned fish species are considered venomous and venom is thought to have evolved 18/19 times in the group (Smith and Wheeler, 2006; Wright, 2009). Despite the number of venomous taxa within fish, piscine venoms are largely underexplored. This is partially due to the lack of defined venom gland structure and subsequently the difficulties with extracting venom, often requiring dissection of venom glands, or due to contamination with body mucus (Harris and Jenner, 2019; von Reumont *et al.*, 2022). Utilising RNA sequencing approaches reduces some of these logistical issues as small samples of the venom gland can provide a lot of information at relatively low cost (von Reumont *et al.*, 2022). As such, this approach can be utilised to elucidate the evolution of venom and venom composition for taxa with little prior information, as is the case with piscine venoms (Magalhães *et al.*, 2006; de Oliveira Júnior *et al.*, 2016). Additionally, venoms often evolve from the same gene families, despite considerable phylogenetic distances among venomous taxa (Casewell *et al.*, 2013). For example, the phospholipase A_2 proteins are part of the venom arsenal of squamate reptiles, cephalopods and insects (Fry *et al.*, 2009). Thus, using the domains of known venom proteins offers a robust framework for the identification of novel venom genes (Fry *et al.*, 2009; Xie *et al.*, 2016).

Corydoradinae represent an excellent system to investigate the evolution of mimicry. They are a species rich group of Neotropical armoured catfish found across South America, consisting of nine monophyletic lineages, across seven genera (Marburger *et al.*, 2018; Dias *et al.*, 2024). Members of the Corydoradinae are defended by lockable pectoral and dorsal spines, with a basal axillary gland containing venom (Greven, Flasbecl and Passia, 2006). In the group, 27 mimicry rings have been identified, with sympatric species belonging to different lineages (Alexandrou *et al.*, 2011). Not all species are equally likely to participate in mimicry rings, with those from *Corydoras*, *Brochis* and *Hoplisoma* being overrepresented (Alexandrou *et al.*, 2011). Additionally, members of *Corydoras* and *Brochis* are only co-mimics when *Hoplisoma sp* is present, suggesting *Hoplisoma* species have biological characteristics which increase the likelihood of mimicry evolving when they are present in the community (Alexandrou *et al.*, 2011). Anecdotal evidence suggests *Hoplisoma* are more toxic than other genera, based on pain inflicted on researchers after envenomation during capture. Interestingly, a whole genome duplication (WGD) is thought to have occurred at the base of *Hoplisoma* (Marburger *et al.*, 2018). Gene duplication events have been shown to increase venom dosage and replenishment as well as increase the functional diversity of the venom (Wong and Belov, 2012). As such, a whole genome duplication (or tandemly duplicated genes) could affect venom potency and complexity in this group. By investigating differences in toxicity between lineages and the genetic underpinning of these differences, we will be able to shed light on the evolution of mimicry in this system.

In this chapter we use complementary analyses to better understand the nature of venom production and potency in the Corydoradinae. Specifically,

1) we utilise a brine shrimp cytotoxicity assay to explore differences in venom potency between members of *Hoplisoma* and *Corydoras.*
2) using RNAseq within a single *Corydoras* species (*C. simulatus*), in conjunction with hidden Markov models (HMM), we identify putative venom genes and housekeeping genes.
3) Using a cross-genus approach, we then explore differences in the putative venom gene expression between the *Corydoras* and *Hoplisoma.*

These complimentary approaches will assist in elucidating the relative potency of venoms in the two genera and consequently, the nature of the mimetic interaction between Corydoradinae co-mimics.

4.3 Methods

4.3.1 Brine cytotoxicity assay

To test for a difference in venom potency between *Corydoras* and *Hoplisoma* species we performed a brine shrimp cytotoxicity assay. Brine shrimp have been shown to be good indicators of venom toxicity, often being used to assess the lethal dose (LD50) of snake venoms (Hahn and O'Connor, 2000; Okumu *et al.*, 2020). They have also proven to be a good indicator of toxicity to fish (Chan *et al.*, 2021). Predatory fish, e.g. *Hoplias sp*, are likely key predators to the Corydoradinae, making brine shrimp cytotoxicity an ecologically relevant assay (Fuller and Evers, 2005). We were not able to calculate the LD50 of Corydoradinae venom in this study as the venom produced from each individual is small and requires dissection of the venom gland from a euthanised individual. Therefore, the number of individuals required to calculate the LD50 would be too large.

4.3.1.1 Venom extract preparation

Extracts were prepared using tissues (venom/axillary gland or muscle) from two Corydoradinae individuals per extract. Corydoradinae individuals were acquired from the aquarium trade. Muscle tissue was chosen for the potency assay due to the volume that could be obtained, and the reduction of contamination from the epidermal mucus. Crude extracts from the venom gland and muscle tissue were prepared following Wright (2009). Briefly, venom gland and muscle tissue were excised from euthanised individuals, and the tissue was then homogenised in 8ml of 1X physiological saline (0.15 mol/L) per 1g of tissue using a sterile disposable micro-pestle. The homogenate was then centrifuged

at 6000 rpm at 4°C for 20 minutes before collecting the supernatant, before an additional centrifuge step for 5 minutes at 4°C at 6000 rpm to create the final raw extract.

A power analysis based on preliminary assay results, using lm() function in base R (v. 4.2.2)(R Core Team, 2023) and pwr.t.test() from the package pwr (v.1.3.0)(Champely, 2020), estimated a minimum of 16 repeats, per treatment. A total number of 20 repeats, per treatment were prepared to further improve power. The same individuals were used to create the venom gland extract and the muscle extract to control for individual differences. Each individual was measured from snout to the base of the caudal fin and the ventral side to dorsal fin using manual callipers to the nearest 0.5mm. An average length for both individuals in an extract was calculated. Additionally, the weight of the extract tissue was recorded for each pair of individuals. As well as being required to ensure consistent concentration between repeats, this allowed us to test for relationships between gland size and individual length.

4.3.1.2 Brine Shrimp Assay

To perform the brine shrimp cytotoxicity assay, 10 (± 1-2) 24-hour old live brine shrimp nauplii were added to each well of a 96 multi-well plate (0.3 mL wells). Brine shrimp cysts were purchased from ZM Fish Food and Equipment (Winchester, UK) and hatched in artificial seawater, made using Instant Ocean Aquarium salt (ZM Fish Food and Equipment, Winchester, UK), at 26°C in 14/10 Light/Dark cycle. Any saline moved with the shrimp was removed and replaced with 10µl of extract with a micro-pipette. A saline only control was used with 10 (± 1-2) 24-hour old live brine shrimp nauplii in 10µl of saline. The wells around the perimeter of the plate were filled with 100µl of saline solution to reduce edge affects. Treatment extracts were replicated over three plates and placement was randomised to prevent confounding effects of plate position, using Well Plate Maker (Borges *et al.*, 2021). Empty wells were filled with 100µl of saline solution and the plate was covered with clear microplate sealing film. Finally, shrimp were left in the plate for 24hrs under constant light conditions at room temperature. After 24 hours, the deceased brine shrimp were counted before euthanising all brine shrimp by adding 100µl of ethanol to each well. The total number of brine shrimp were then counted. The proportion of brine shrimp mortality over the 24 hour was calculated as brine shrimp deaths divided by the total number of brine shrimp.

A generalized linear mixed model (GLMM) was performed using the lme4 (v.1.1.35.1)(Bates *et al.*, 2015) R package, via glmer() with mortality as the dependent variable and tissue type and genera included as fixed effects. The ID of the individuals used to create each tissue extraction was included as a random effect. Mortality was assessed as a binomial trait, with total number of nauplii in the well

included as a weight to account for the data being proportional. The distribution of residuals of our model showed no significant deviation or dispersion, as assessed using the R package DHARMa (v.0.4.6)(Hartig, 2022). Additional models were run including average venom gland size and average body length as a fixed factor. These non-significant factors were subsequently dropped due to increased Akaike information Criterion (AIC), indicating poorer model fit. We performed post-hoc testing to generate pairwise comparisons, using emmeans() from the package emmeans (v. 1.10.3)(Lenth, 2024) which uses Tukey's tests to adjust for multiple testing. To explore the effect of tissue type on mortality (independently of Corydoradinae genera), compared to our physiological saline controls, we ran another GLMM with mortality as the dependent variable and tissue type as a fixed effect. Again, we included ID as a random effect.

We also tested for differences in gland size per lineage. The *Corydoras* individuals appeared to have a larger body size than *Hoplisoma* individuals, so a preliminary t-test was performed in base R. We subsequently tested for differences in gland size with Corydoradinae genera as a fixed effect and body size as a random effect, using the lmer() function from lme4. We used the package lmerTest (v. 3.1.3)(Kuznetsova, Brockhoff and Christensen, 2017) to obtain exact p-values.

4.3.2 Transcriptome Sample Preparation and Sequencing

4.3.2.1 Sample Acquisition, RNA extraction and Sequencing

Species included in this study were chosen to represent three independent mimicry rings, each with a *Corydoradinae* species (*C. narcissus, C. desana, C. simulatus*) and a *Hoplisoma* species (*H. arcuatus, H. tukano, H. metae*). Individuals were euthanised and the axillary (venom) gland tissue was excised and immediately flash frozen using liquid nitrogen. Additionally, scute tissue (the thick armoured skin of *the Corydoradinae*) was taken from *C. simulatus* to provide a between tissue comparison. Scute tissue was chosen from *C. simulatus* as it was available from another study (See Chapter 5). RNA was extracted using an adapted RNAeasy micro or mini kit (Qiagen, Germany) protocol with a prior Trizol step as detailed (http://www.untergasser.de/lab/protocols/rna_prep_comb_trizol_v1_0.htm). A volume of 20μl at a minimum concentration of 20 ng/μl was sent to be sequenced. Quality control and library preparation for Illumina sequencing were performed at Novogene Cambridge, using NovaSeq 6000 and each sample sequenced to 40 million reads.

4.3.2.2 Transcriptome assembly and Annotation

Cross-species differential expression can be complicated by differences in mapping quality if using a single assembly reference. This is more pertinent when comparing expression levels in diploids / polyploid pairs. To reduce mapping biases, we created species-specific master transcriptomes, with

each sequenced individual mapped to a transcriptome of its own species. This was achieved by combining raw sequence files from individuals of the same species before trimming reads, removing adapters, and filtering using a phred score of 20 in cutadapt (v. 2.10)(Martin, 2011). The processed sequences were then visually inspected using FastQC (v.0.11.8)(Andrews, 2010). Next, the sequences were assembled into *de novo* transcriptomes using Trinity (v. 2.11.0)(Grabherr *et al.*, 2011) with the default settings. We then used BUSCO (v. 5.3.2)(Simão *et al.*, 2015) to evaluate the quality of the transcriptomes, against the actinopterygii_odb10 (teleost) database. The transcriptomes were annotated using Trinnotate (v.4.0.2)(Bryant *et al.*, 2017) with UniProt (UniProt Consortium, 2021) and Pfam (Mistry *et al.*, 2021) databases using tblastx and blastp from BLAST(v. 2.10.1+)(Camacho *et al.*, 2009) and HMMER (v.3.3)(http://hmmer.org/).

4.3.3 Venom gene excavation and differential expression

4.3.3.1 Hidden Markov Models

Genes with roles in venom maintenance and production are only known in taxa distantly related to the *Corydoradinae*. To identify candidate genes containing domains with roles in venom production and maintenance, we built HMMs using venom proteins from VenomZone (https://venomzone.expasy.org/). These were grouped by protein families and HMMs were created per group using hmmbuild by HMMER. We then used TransDecoder (v. 5.5.0)(https://github.com/TransDecoder/TransDecoder) to extract the longest isoform of each *C. simulatus* transcript and translate these into protein sequences. The HMMs were then used to identify candidates with known venom domains in the *C. simulatus* transcriptome using hmmsearch by HMMER.

4.3.3.2 Differential Expression analysis

Transcript abundance was quantified from trimmed and filtered sequences using Salmon (v.1.10.0)(Patro *et al.*, 2017), implemented through the align_andestimate_abundance.pl script from Trinity, for each individual using the species-specific transcriptome. Transcript abundance was normalised by gene length using TXImport (v. 1.26.1)(Soneson, Love and Robinson, 2015) in R. Differential expression analysis was performed using DESeq2 (v. 1.38.3)(Love, Huber and Anders, 2014). For between tissue comparisons, both the venom gland and scute tissue were pseudo aligned to the *C. simulatus* transcriptome (using Salmon). To compare between genera, we utilised the transcriptome annotations. We assigned annotations to each transcript in R using the SwissProt blastp hit, or if missing, the top SwissProt tblastx hit. When multiple hits had been assigned, these were filtered to leave annotations with the lowest evalue and the highest percentage identity. Transcripts

that had annotations belonging to prokaryotes were removed. The gene id was then replaced in the gene transcript map file provided by Trinity with the annotation for that transcript. Gene level inferences were estimated using the tx2gene() function from TXImport, creating species specific matrices. To run the differential analysis between species, the species-specific matrices were combined using the annotation via reduce() from purrr (v. 1.0.2)(Wickham and Henry, 2023) and left_join() from base R. To view clustering between samples, we performed normalisation via vst() from DESeq2 (v. 1.38.3)(Love, Huber and Anders, 2014).

4.3.3.3 Gene network analysis

Weighted gene network analysis was performed to identify genes with correlated expression, to further identify venom candidates as well as housekeeping genes (Haney *et al.*, 2019). We performed comparative gene network expression on the venom gland and scute tissue from *C. simulatus.* To construct gene expression modules, we normalized the read data using vst() from DESeq2 (v. 1.38.3)(Love, Huber and Anders, 2014). To estimate expression modules, we utilised the pickSoftThreshold() and blockwiseModules() functions from WGCNA (v.1.72.5)(Langfelder and Horvath, 2008, 2012). To explore differential expression of modules between venom and scute tissue, we created a design matrix using the function model.matrix(), before fitting a linear model using lmFit(), both from limma (v.3.54.2)(Ritchie *et al.*, 2015). To stabilize variability in the estimates, due to small sample size, we applied empirical Bayes moderation, via the eBayes() function from limma (Love, Huber and Anders, 2014). To correct for multiple testing using the Benjamini-Hochberg (BH) method we used the function topTable(), from limma, and then calculated the mean expression across the significantly differentially expressed modules so that one expression value for each sample could be visualised.

4.4 Results

4.4.1 Potency assay

We explored venom potency differences between a *Corydoras sp* and a *Hoplisoma sp*, which are co-mimics in the Rio Meta in Columbia. We identified a significant effect of tissue type on the mortality of brine shrimp (estimate = 2.67, SE = 0.43, z = 6.16, p < 0.0001) with venom gland causing

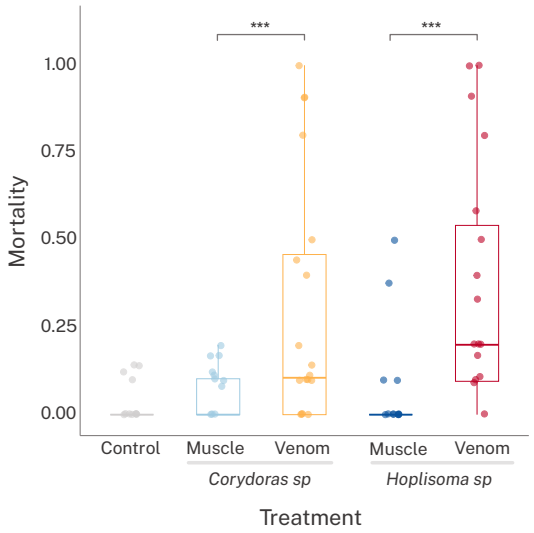


Figure 4.1 The effect of different tissue extracts and *Corydoradinae* genera on mortality of 24-hour old brine shrimp, after a 24-hour incubation period. Asterisks denote significant differences (p < 0.0001).

significantly higher mortality than muscle tissue (Figure 4.1). However, we did not find a significant effect for genera (estimate = -0.21, SE = 0.80, z =-0.27, p = 0.79) nor the interaction between genera and tissue (estimate=0.48, SE = 0.65, z = 0.74, p = 0.46). There was a significant effect of tissue within both *Corydoras sp* and *Hoplisoma sp* but not between genera (Table 4.1). We also explored the effect of tissue – independently of genera- compared to saline controls. As previously, venom was significantly more toxic to nauplii than muscle extracts (estimate = -2.88, SE = 0.33, z.ratio = -8.80, p < 0.0001). The mortality within the saline control was significantly lower than the venom extract (estimate = -3.50, SE = 0.76, z.ratio =-4.62, p < 0.0001) but not significantly different from the muscle extract (estimate = -0.62, SE = 0.79, z.ratio = -0.79, p = 0.71). Gland size could impact the amount of protein produced (de Roodt *et al.*, 2016). The *Corydoras sp* individuals used in this study were significantly larger than the *Hoplisoma sp* (t = 17.61, df = 31.92, p-value < 0.0001). Controlling for

body size, we found no significant effect of genera on gland size (estimate=-0.00, SE=0.00, df=17.40, t=-1.67, p = 0.11).

Table 4.1 Post-hoc testing exploring the effect of tissue extract and Corydoradinae genera on mortality in 24-hour old brine shrimp after a 24-hour incubation period. Asterisks denote significant differences.

Contrast	Tissue / Species	Estimate	SE	df	z.ratio	p.value
Between Genera						
Corydoras sp - Hoplisoma sp	Muscle tissue	0.21	0.80	Inf	0.27	0.79
Corydoras sp - Hoplisoma sp	Venom tissue	-0.26	0.61	Inf	-0.43	0.67
Within Genera						
Muscle- Venom	*Corydoras sp*	-2.67	0.43	Inf	-6.16	<.0001***
Muscle- Venom	*Hoplisoma sp*	-3.15	0.49	Inf	-6.42	<.0001***

4.4.2 Transcriptome Assembly

To identify candidate venom genes and quantify potential differences in expression between co-mimics, we sequenced the venom gland (n = 6) and scute tissue (n = 6) of *C. simulatus*. Additionally, we sequenced the venom glands of *H. arcuatus* (n = 4), *H. metae* (n = 5), *H. tukano* (n = 2), *C. narcissus* (n = 2) and *C. desana* (n = 3). By including additional species, we can look at changes between genera rather than species-specific changes in expression. The mean number of reads was 47163373 (range= 39776276 - 58749872), and the mean read number was similar between scute tissue and venom tissue. The *de novo* assembled transcriptomes had an average BUSCO score of 79.73% complete gene models, indicating a good assembly and low fragmentation of gene models (Supplementary Table 4.1). The transcriptome with the highest BUSCO score was *C. simulatus*, with 88.1% complete gene models. This may be due to this transcriptome being assembled from two tissues, rather than just the venom gland.

4.4.3 Venom gene excavation

We used HMM to identify putative venom genes, based on known venom domains and confirmed the role of these candidates using a within species comparison of RNAseq in *C. simulatus,* between the venom gland and scute tissue. The HMMs were built from known venom protein families and used to identify candidate venom genes within the *C. simulatus* transcriptome. The models identified a total of 1,054 transcripts with domains similar to those of known venoms. Within the identified transcripts,

the highest represented family was the True venom lectin (C–type lectins) and the snaclecs, which are also a type of C-type lectins found in snakes (Clemetson, Morita and Kini, 2009).

To further identify candidate venom genes, we performed differential expression analysis between tissues in *C. simulatus.* Within *C. simulatus*, samples clustered by tissue type (Figure 4.2). A total of 1,451 genes were differentially expressed ($p < 0.05$) between tissue type, representing 5.7% of gene models. Of these genes, 763 were upregulated in the venom gland (Figure 4.3A). Of these genes, 33 had been identified by the HMM as candidate venom genes (Figure 4.3A). The domains of the true venom (C-lectins), the ohanin/vespryn and snaclec families were most highly represented in the putative venom genes (Figure 4.3B). One co-expression module contained 25 venom candidate genes, representing our final venom candidates (Table 4.2). Additionally, the module contained 462 genes that had been significantly differentially expressed between the venom gland and the scute tissue. These could therefore represent venom housekeeping genes. Furthermore, this module was

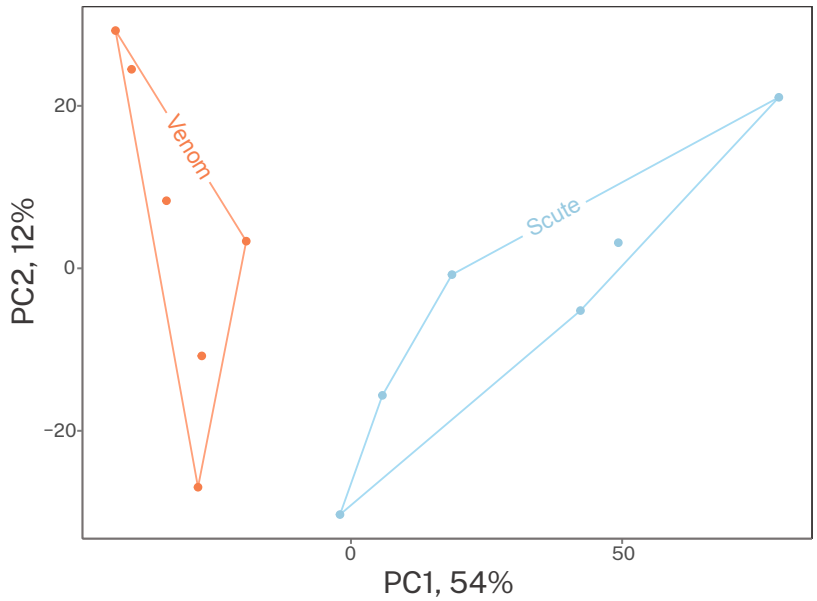


Figure 4.2 Principal component analysis of RNAseq data from two tissues, venom gland (axillary tissue) and scute tissue from *Corydoras simulatus.*

differentially expressed between the two tissues and upregulated in the venom gland ($p < 0.05$) (Figure 4.4).

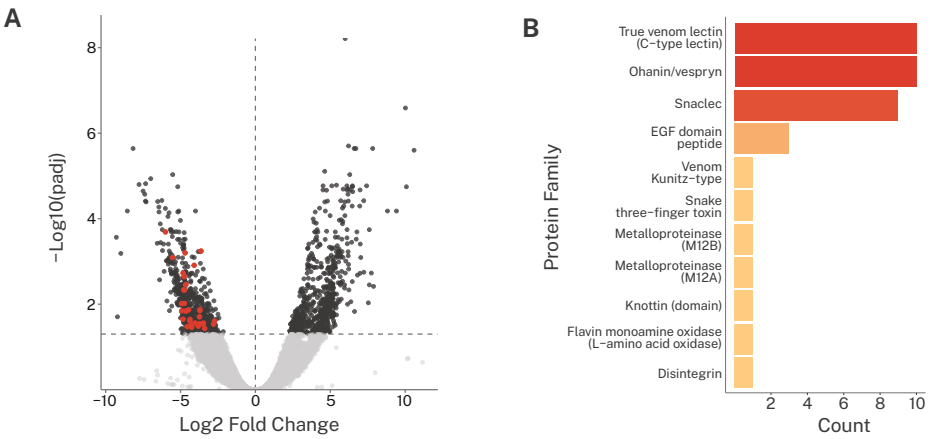


Figure 4.3A Volcano plot showing the differential expression between *Corydoras simulatus* venom gland and scute tissue. The putative venom genes are highlighted in red. **B** The frequency of venom domains from known families represented in putative venom genes.

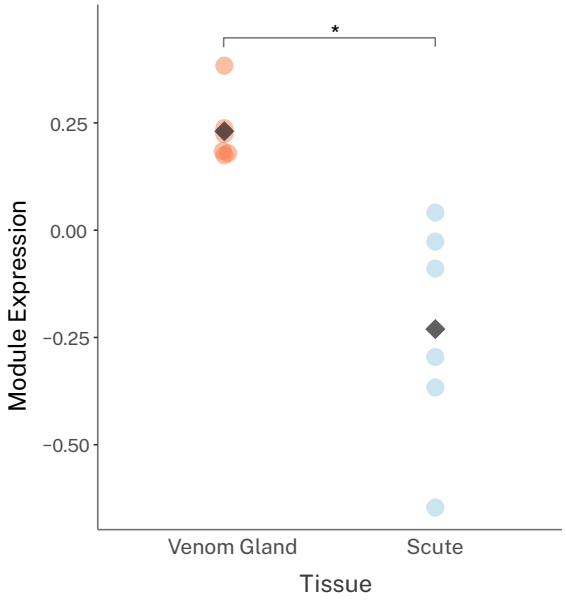


Figure 4.4 Difference in module expression between the venom gland and scute tissue of *Corydoras simulatus.* Module contains 25 venom candidate genes, identified in this study. Significant difference is denoted by the asterisk ($p < 0.05$).

Table 4.2 Putative venom genes identified via differential expression of the venom gland and scute tissues of *Corydoras simulatus,* which have correlated expression and domains similar to those of known venom genes.

Gene	Protein	Venom Domains	Transcript ID
mrc1	Macrophage mannose receptor 1	Snaclec | True venom lectin (C-type lectin)	TRINITY_DN441_c0_g1
trim35	E3 ubiquitin-protein ligase TRIM35	Ohanin/vespryn	TRINITY_DN2477_c0_g1
trim35	E3 ubiquitin-protein ligase TRIM35	Ohanin/vespryn	TRINITY_DN2477_c0_g2
trim25	E3 ubiquitin/ISG15 ligase TRIM25	Ohanin/vespryn	TRINITY_DN4518_c7_g1
adam9	Disintegrin and metalloproteinase domain-containing protein 9	Disintegrin | Metalloproteinase (M12B) | Metalloproteinase (M12A)	TRINITY_DN1528_c2_g1
pkd1l2	Polycystin-1-like protein 2	EGF domain peptide | True venom lectin (C-type lectin)	TRINITY_DN12556_c0_g1
spink1	Serine protease inhibitor Kazal-type 1	EGF domain peptide | Knottin (domain)	TRINITY_DN4489_c0_g1
trim16	Tripartite motif-containing protein 16	Ohanin/vespryn	TRINITY_DN1264_c2_g4
smox	Spermine oxidase	Flavin monoamine oxidase (L-amino acid oxidase)	TRINITY_DN12006_c0_g1
tpm1	Tropomyosin alpha-1 chain	Snake three-finger toxin	TRINITY_DN79_c1_g1
cd209d	CD209 antigen-like protein D	Snaclec | True venom lectin (C-type lectin)	TRINITY_DN128_c0_g1
trim16	Tripartite motif-containing protein 16-like protein	Ohanin/vespryn	TRINITY_DN37114_c0_g2
ly75	Putative C-type lectin domain family 20 member A	Snaclec | True venom lectin (C-type lectin)	TRINITY_DN5431_c0_g2
ly75	Putative C-type lectin domain family 20 member A	Snaclec | True venom lectin (C-type lectin)	TRINITY_DN19294_c0_g1
spint1	Kunitz-type protease inhibitor 1	EGF domain peptide | Venom Kunitz-type	TRINITY_DN940_c3_g1
mrc1	Macrophage mannose receptor 1	Snaclec | True venom lectin (C-type lectin)	TRINITY_DN38136_c0_g1
mrc1	Macrophage mannose receptor 1	Snaclec | True venom lectin (C-type lectin)	TRINITY_DN35291_c0_g1
spsb2	B box and SPRY domain-containing protein	Ohanin/vespryn	TRINITY_DN1393_c17_g1
-	Nuclear factor 7, ovary	Ohanin/vespryn	TRINITY_DN4427_c0_g1
trim16	Tripartite motif-containing protein 16	Ohanin/vespryn	TRINITY_DN1157_c1_g1
clec4m	C-type lectin domain family 4 member M	Snaclec | True venom lectin (C-type lectin)	TRINITY_DN73_c0_g1
trim39	E3 ubiquitin-protein ligase TRIM39	Ohanin/vespryn	TRINITY_DN1539_c0_g1
znf33a	Zinc-binding protein A33	Ohanin/vespryn	TRINITY_DN959_c1_g1
mrc1	Macrophage mannose receptor 1	Snaclec | True venom lectin (C-type lectin)	TRINITY_DN1684_c1_g1
ly75	Putative C-type lectin domain family 20 member A	Snaclec | True venom lectin (C-type lectin)	TRINITY_DN2761_c0_g1

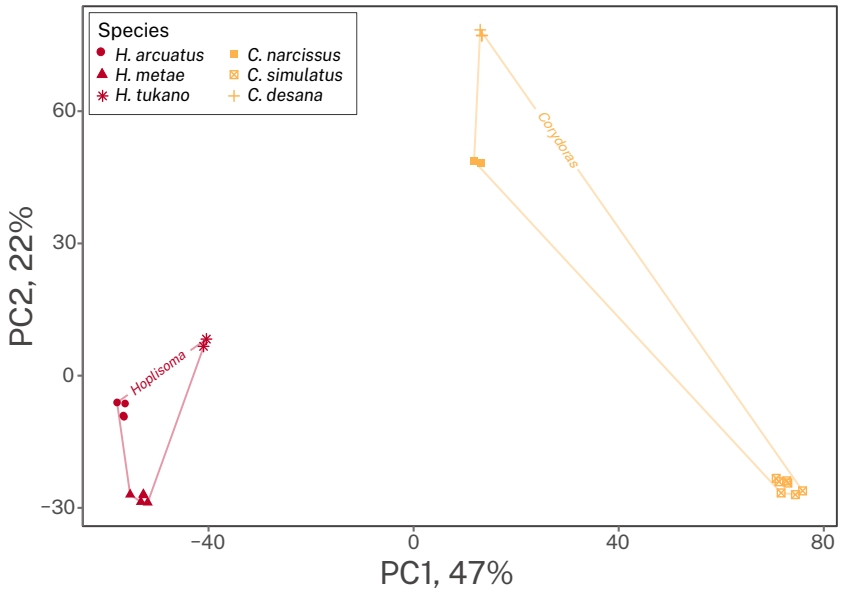


Figure 4.5 Principal component analysis of RNAseq data from the venom glands of six species in two Corydoradinae genera, *Hoplisoma* and *Corydoras*. These are *H. arcuatus, H. metae, H. tukano* from and *C. narcissus, C. simulatus, C. desana.*

4.4.4 Differential Expression in *Hoplisoma* and *Corydoras*

We investigated the relative expression of the identified venom candidate genes between *Corydoras sp* and *Hoplisoma sp*, which are often co-mimics. Samples clustered within species and genera (Figure 4.5). A total of 5,669 gene models were differentially expressed between the two genera. Of these, 11 had annotations matching those that were identified as venom candidates (Figure 4.6). The venom ortholog with the highest log-fold expression was Tropomyosin alpha-1 chain which had a log2fold change of 3.14, being significantly upregulated in the *Hoplisoma sp* ($p < 0.0001$). Of the 462 housekeeping venom genes, 90 were differentially expressed between *Hoplisoma sp* and *Corydoras sp.* The differentially expressed housekeeping orthologs were enriched in Gene Ontology (GO) terms associated with protein transport and cell to cell signalling (Figure 4.7)

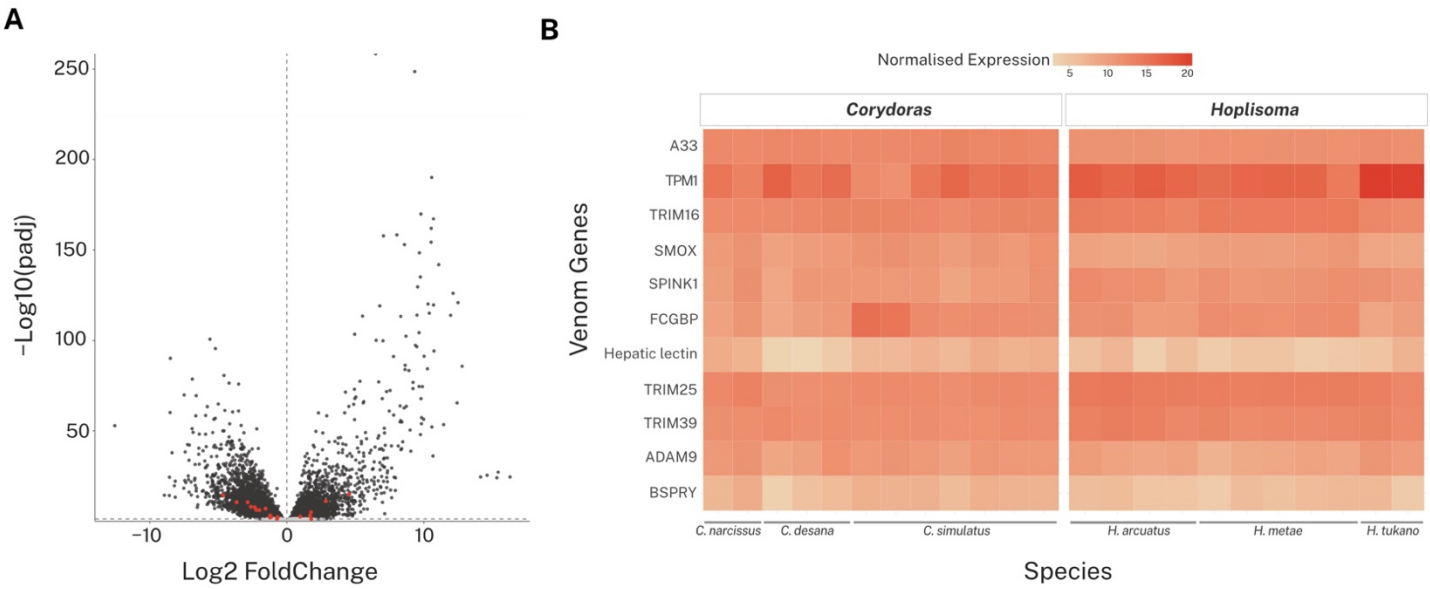


Figure 4.6 A Volcano plot showing significantly differentially expressed genes($p < 0.05$) between *Corydoras* and *Hoplisoma* venom glands. Venom orthologs are highlighted in red. **B** Heatmap of significantly differentially expressed venom orthologs *Corydoras* and *Hoplisoma* species. Each column represents a single individual. Expression data has been normalized and species are given at the base of the heatmap.

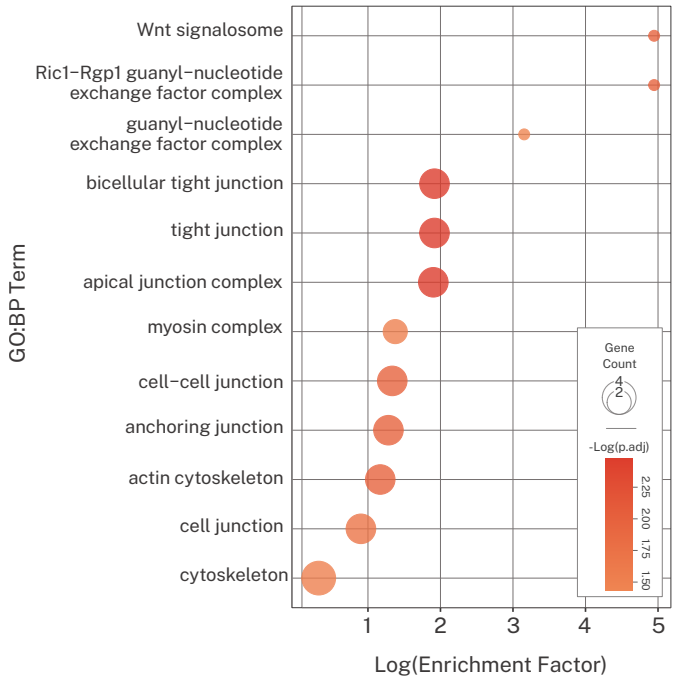


Figure 4.7 Gene ontology (GO) enrichment of differentially expressed venom housekeeping genes between the venom (axillary) glands of *Corydoras* and *Hoplisoma* species. The size of the circle represents the number of venom housekeeping genes with each GO term.

4.5 Discussion

Protective mimicry relies on two components, unpalatability and the advertisement of unpalatability via warning colouration. The former is often understudied in mimetic communities but is vital to understanding the evolution of such systems. To better understand the nature of ecological interactions between co-mimics, we investigated the venom potency of *Corydoras sp* and *Hoplisoma sp.* These Corydoradinae genera frequently co-occur in Müllerian mimicry rings, but evidence from envenomation of researchers and hobbyists suggests that *Hoplisoma sp* envenomations are more painful than other Corydoradinae genera (Alexandrou *et al.*, 2011). Using a brine shrimp cytotoxicity assay, we demonstrated venom gland extracts were more toxic to brine shrimp than muscle tissue, but *Hoplisoma sp* do not have more potent venom than *Corydoras sp* (Figure 4.1). We also explored the genetic basis for venom within the group, identifying candidate genes with domains similar to those of known venoms (Table 4.2, Figure 4.3). These genes had correlated expression, providing further evidence to their importance as venom candidates. Additional co-expressed genes which are significantly differentially expressed between tissue types are thought to represent venom housekeeping genes (Barua and Mikheyev, 2021). Some venom genes were significantly differentially expressed between *Corydoras* and *Hoplisoma*, but the logfold change was low (Figure 4.6). Many housekeeping genes were also significantly differentially expressed between the genera. Gene ontology enrichment analysis revealed that these genes were enriched in terms relating to cell-cell signalling and protein movement (Figure 4.7).

The evolution of mimetic interactions occurs through either advergence, where one taxon evolves to resemble the other, or convergence, where both taxa evolve to share a resemblance (Ruxton *et al.*, 2018). In Batesian mimicry, selection acts on the palatable taxon to gain protection from predators by more closely resembling the unpalatable taxon (Ruxton *et al.*, 2018). In Müllerian mimicry, co-mimics may either adverge or converge. Müller's original model suggested taxa in Müllerian mimicry rings should converge on a shared phenotype, but this is rarely seen in nature and empirical evidence largely supports advergence (Müller, 1878; Symula, Schulte and Summers, 2001; Marek and Bond, 2009; Ruxton *et al.*, 2018). A proposed explanation for this is quasi-Batesian mimicry, where unequal unpalatabilities result in a parasitic (Batesian-like) interaction (Speed, 1993; Mallet, 1999; Speed and Turner, 1999). Theoretical models and experimental systems demonstrate the plausibility of this hypothesis, but empirical evidence is lacking (Mallet, 1999; Mallet and Joron, 1999; Rowland *et al.*, 2010). The evidence we present here suggests that the *Corydoradinae* co-mimics are likely to be true Müllerian mimics with regards to venom potency, following Müllers original assumptions. Although, differences in orientation and shape of spine serrations could result in differences in unpalatability,

e.g. *Hoplisoma* are considered to have sharper spines (Dias *et al.*, 2024). Additionally, venom gland size has been associated with a diversity in venom toxins and potency in spiders (Pekár *et al.*, 2018). However, we found a non-significant effect of genera on gland size, when controlling for body size ($p=0.11$). Whilst brine shrimp are not the recipients of Corydoradinae venom in nature, they are often used to assess potency in snake venoms and show a strong correlation with both mouse lethality and toxicity to fish (Okumu *et al.*, 2020; Chan *et al.*, 2021; Araya *et al.*, 2024). The significant difference ($p < 0.0001$) between venom extract and muscle extract control indicates the efficacy of this technique and presents a methodology for testing relative venom potency in piscine systems, which remain largely unexplored (Smith and Wheeler, 2006; Harris and Jenner, 2019).

The venom genes identified in this study have domains predominantly matching the true venom lectin (C-type lectin) and the Ohanin/ vespryn families (Figure 4.3A). Additionally, the Snaclec domains were also identified in nine candidate transcripts, although these were often in conjunction with the true venom lectin domains. Both Snaclec and True venom lectins are C-type lectins, so this pairing is unsurprising (Clemetson, Morita and Kini, 2009). C-type lectins have been identified as components of venom in the Scorpaeniformes (the Scorpion fishes), which also use venom for defence rather than predation, as in the Corydoradinae (Nakagawa *et al.*, 2015; Campos *et al.*, 2016; Ziegman *et al.*, 2019). Although venom evolved independently in both the Scorpaeniformes and Corydoradinae, this convergence could represent co-option of similar proteins into venoms in both groups, e.g. ichthyocrinotoxins (Smith and Wheeler, 2006; Wright, 2009; Harris and Jenner, 2019).

Of our venom candidates, multiple gene models were annotated as *trim35*, *mrc1* and *ly75* (Table 4.2). This could represent gene duplication events. Gene duplication is widely considered to play a role in venom evolution and complexity in snakes, cone snails and spiders, sea anemones and Scorpaenoid fish (Chang and Duda, 2012; Chuang and Shiao, 2014; Gendreau *et al.*, 2017; Smith *et al.*, 2023). Although these occur predominantly through tandem duplication (Wong and Belov, 2012). The Corydoradinae could represent an opportunity to study the role of both whole genome duplication and tandem duplication in venom complexity, as evidence indicates a whole genome duplication event has occurred in *Hoplisoma* (Marburger *et al.*, 2018)(Chapter 2).

Whilst our comparison of *Hoplisoma* and *Corydoras* venom genes largely corroborate our venom potency assay, showing small log foldchanges in expression of significantly differentially expressed genes, we explored this on an orthogroup level, based on transcriptome annotation and thus, there may still be some differences in expression at the transcript level. For example, evidence in the house spider demonstrated alternative splicing has contributed to venom complexity (Haney *et al.*, 2019). However, delimiting paralogs from isoforms in short read transcriptomes is challenging, particularly in

recent polyploids, but this could be reduced by utilising long read sequencing (Oh, Lee and Park, 2021). Despite these limitations, we do see some differences in the expression of *tpm1*, particularly in *H. tukano* (Figure 4.6). Whilst being a member of *Hoplisoma*, *H. tukano,* is a member of a different sub-group (deriving from Lineage 6) from the other *Hoplisoma spp.* included here (Alexandrou *et al.*, 2011; Dias *et al.*, 2024). This could represent a future line of investigation into venom complexity within *Hoplisoma.* Additionally, some of the candidate housekeeping venom genes are significantly differentially expressed. We have shown that these housekeeping genes are enriched for both intracellular and protein transport (Figure 4.7). This is similar to the findings in the common house spider (*Parasteatoda tepidariorum*), where venom module genes were enriched for GO terms relating to intracellular signalling, transport and signal transduction pathways (Zhu *et al.*, 2023). However, this study, and those on the snake venom gene networks, also identified genes associated with protein modification, which plays a key role in venom production, which we did not identify in our data (Haney *et al.*, 2019; Barua and Mikheyev, 2021).

In conclusion, we have provided further evidence that mimetic *Corydoras sp* and *Hoplisoma sp* are true Müllerian mimics with regards to venom potency, and thus, are likely to have a mutualistic interaction. Additionally, we have shown that the venom of the *Corydoras* is a complex cocktail, likely to contain both C-type lectins and Ohanin/ vespryn genes, which is in line with other findings of piscine venoms. Significant but small log2fold changes between *Corydoras sp* and *Hoplisoma sp* are concordant with the potency assay, indicating that venom potency and complexity is conserved between the two groups. This work is an important contribution to the understanding of the genetics of mimicry, particularly of unpalatability, which is overlooked in favour of warning colouration.

4.6 Supplementary Materials

Supplementary Table 4.1 The BUSCO scores for both the *Corydoras* and *Hoplisoma de novo* transcriptome assemblies, indicating quality of assembly.

Species	Tissue	Complete	Single	Duplicated	Fragmented	Missing
Hoplisoma arcuatus	Venom gland	75.2	38.4	36.8	5.1	19.7
Hoplisoma metae	Venom gland	83.2	27	56.2	3.3	3.3
Hoplisoma tukano	Venom gland	69.4	41.9	27.5	5.9	24.7
Corydoras narcissus	Venom gland	81.2	33.3	47.9	3.7	15.1
Corydoras desena	Venom gland	81.3	37.4	43.9	3.3	15.4
Corydoras simulatus	Venom gland and Scute tissue	88.1	28.2	59.9	4.1	7.8

4.7 Attribution Statement

Project conception by Emily Phelps and Martin Taylor. All wet laboratory work and analysis was performed by Emily Phelps. Supervision by Martin Taylor.

4.8 References

Alexandrou, M.A., Oliveira, C., Maillard, M., *et al.* (2011) Competition and phylogeny determine community structure in Mullerian co-mimics, *Nature*, 469(7328), pp. 84–88. https://doi.org/10.1038/nature09660.

Andrews, S. (2010) *FastQC: A Quality Control Tool for High Throughput Sequence*. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Araya, X., Okumu, M., Durán, G., *et al.* (2024) Assessment of the Artemia salina toxicity assay as a substitute of the mouse lethality assay in the determination of venom-induced toxicity and preclinical efficacy of antivenom, *Toxicon: X*, 22(100195). https://doi.org/10.1016/j.toxcx.2024.100195.

Aubier, T.G., Joron, M. and Sherratt, T.N. (2017) Mimicry among unequally defended prey should be mutualistic when predators sample optimally, *The American Naturalist*, 189(3), pp. 267–282. https://doi.org/10.1086/690121.

Barnett, C.A., Bateson, M. and Rowe, C. (2007) State-dependent decision making: educated predators strategically trade off the costs and benefits of consuming aposematic prey, *Behavioral Ecology*, 18(4), pp. 645–651. https://doi.org/10.1093/beheco/arm027.

Barua, A. and Mikheyev, A.S. (2021) An ancient, conserved gene regulatory network led to the rise of oral venom systems, *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 118(14). https://doi.org/10.1073/pnas.2021311118.

Bates, D., Mächler, M., Bolker, B., *et al.* (2015) Fitting Linear Mixed-Effects Models Using lme4, *Journal Of Statistical Software*, pp. 1–48. https://doi.org/10.18637/jss.v067.i01.

Bates, H.W. (1862) Contributions to an insect fauna of the Amazon valley. Lepidoptera: heliconinae, *Journal Of The Proceedings Of The Linnean Society Of London Zoology*, 6(22), pp. 73–77. https://doi.org/10.1111/j.1096-3642.1862.tb00932.x.

Borges, H., Hesse, A.-M., Kraut, A., *et al.* (2021) Well Plate Maker: a user-friendly randomized block design application to limit batch effects in large-scale biomedical studies, *Bioinformatics* , 37(17), pp. 2770–2771. https://doi.org/10.1093/bioinformatics/btab065.

Bryant, D.M., Johnson, K., DiTommaso, T., *et al.* (2017) A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors, *Cell Reports*, 18(3), pp. 762–776. https://doi.org/10.1016/j.celrep.2016.12.063.

Caley, M.J. and Schluter, D. (2003) Predators favour mimicry in a tropical reef fish, *Proceedings. Biological Sciences*, 270(1516), pp. 667–672. https://doi.org/10.1098/rspb.2002.2263.

Camacho, C., Coulouris, G., Avagyan, V., *et al.* (2009) BLAST+: architecture and applications, *BMC Bioinformatics*, 10. https://doi.org/10.1186/1471-2105-10-421.

Cameron, A.M. and Endean, R. (1973) Epidermal secretions and the evolution of venom glands in fishes, *Toxicon*, 11(5), pp. 401–410. https://doi.org/10.1016/0041-0101(73)90115-3.

Campos, F.V., Menezes, T.N., Malacarne, P.F., *et al.* (2016) A review on the Scorpaena plumieri fish venom and its bioactive compounds, *The Journal Of Venomous Animals And Toxins Including Tropical Diseases*, 22(1). https://doi.org/10.1186/s40409-016-0090-7.

Casewell, N.R., Wüster, W., Vonk, F.J., *et al.* (2013) Complex cocktails: the evolutionary novelty of venoms, *Trends In Ecology & Evolution*, 28(4), pp. 219–229. https://doi.org/10.1016/j.tree.2012.10.020.

Chan, W., Shaughnessy, A.E.P., van den Berg, C.P., *et al.* (2021) The validity of brine shrimp (Artemia Sp.) toxicity assays to assess the ecological function of marine natural products, *Journal Of Chemical Ecology*, 47(10–11), pp. 834–846. https://doi.org/10.1007/s10886-021-01264-z.

Chang, D. and Duda, T.F., Jr (2012) Extensive and continuous duplication facilitates rapid evolution and diversification of gene families, *Molecular Biology And Evolution*, 29(8), pp. 2019–2029. https://doi.org/10.1093/molbev/mss068.

Chuang, P.-S. and Shiao, J.-C. (2014) Toxin gene determination and evolution in scorpaenoid fish, *Toxicon*, 88, pp. 21–33. https://doi.org/10.1016/j.toxicon.2014.06.013.

Clemetson, K.J., Morita, T. and Kini, R.M. (2009) Classification and nomenclature of snake venom C-type lectins and related proteins, *Toxicon*, 54(1). https://doi.org/10.1016/j.toxicon.2009.04.001.

Dias, A.C., Tencatt, L.F.C., Roxo, F.F., *et al.* (2024) Phylogenomic analyses in the complex Neotropical subfamily Corydoradinae (Siluriformes: Callichthyidae) with a new classification based on morphological and molecular data, *Zoological Journal Of The Linnean Society* .

Fry, B.G., Roelants, K., Champagne, D.E., *et al.* (2009) The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms, *Annual Review Of Genomics And Human Genetics*, 10(1), pp. 483–511. https://doi.org/10.1146/annurev.genom.9.081307.164356.

Fuller, I.A.M. and Evers, H.E. (2005) *Identifying Corydoradinae Catfishe*. Kidderminster, England: Ian Fuller Enterprises.

Gendreau, K.L., Haney, R.A., Schwager, E.E., *et al.* (2017) House spider genome uncovers evolutionary shifts in the diversity and expression of black widow venom proteins associated with extreme toxicity, *BMC Genomics* , (1). https://doi.org/10.1186/s12864-017-3551-7.

Grabherr, M.G., Haas, B.J., Yassour, M., *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome, *Nature Biotechnology*, 29(7), pp. 644–652. https://doi.org/10.1038/nbt.1883.

Gratzer, B., Millesi, E., Walzl, M., *et al.* (2015) Skin toxins in coral-associated Gobiodon species (Teleostei: Gobiidae) affect predator preference and prey survival, *Marine Ecology*, 36(1), pp. 67–76. https://doi.org/10.1111/maec.12117.

Greven, H., Flasbecl, T. and Passia, D. (2006) Axillary glands in the armoured catfish Corydoras aeneus (Callichthyidae, Siluriformes), *Verhandlungen Der Gesellschaft Für Ichthyologie*, 5, pp. 65–69.

Hahn, S.T. and O'Connor, J.M. (2000) An investigation of the biological activity of bullrout (Notesthes robusta) venom, *Toxicon*, 38(1), pp. 79–89. https://doi.org/10.1016/s0041-0101(99)00135-x.

Haney, R.A., Matte, T., Forsyth, F.S., *et al.* (2019) Alternative transcription at venom genes and its role as a complementary mechanism for the generation of venom complexity in the common house spider, *Frontiers In Ecology And Evolution*, 7. https://doi.org/10.3389/fevo.2019.00085.

Harris, R.J. and Jenner, R.A. (2019) Evolutionary ecology of fish venom: Adaptations and consequences of evolving a venom system, *Toxins*, 11(2). https://doi.org/10.3390/toxins11020060.

Hartig, F. (2022) DHARMa: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models. https://CRAN.R-project.org/package=DHARMa.

Kuznetsova, A., Brockhoff, P.B. and Christensen, R.H.B. (2017) lmerTest Package: Tests in Linear Mixed Effects Models, *Journal Of Statistical Software*, pp. 1–26. https://doi.org/10.18637/jss.v082.i13.

Lenth, R.V. (2024) emmeans: Estimated Marginal Means, aka Least-Squares Means. https://CRAN.R-project.org/package=emmeans.

Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, *Genome Biology*, 15(12). https://doi.org/10.1186/s13059-014-0550-8.

Magalhães, G.S., Junqueira-de-Azevedo, I.L.M., Lopes-Ferreira, M., *et al.* (2006) Transcriptome analysis of expressed sequence tags from the venom glands of the fish Thalassophryne nattereri, *Biochimie*, 88(6), pp. 693–699. https://doi.org/10.1016/j.biochi.2005.12.008.

Mallet, J. (1999) Causes and consequences of a lack of coevolution in Müllerian mimicry, *Evolutionary Ecology*, 13(7), pp. 777–806. https://doi.org/10.1023/A:1011060330515.

Mallet, J. and Joron, M. (1999) Evolution of diversity in warning color and mimicry: Polymorphisms, shifting balance, and speciation, *Annual Review Of Ecology And Systematics*, 30(1), pp. 201–233. https://doi.org/10.1146/annurev.ecolsys.30.1.201.

Marburger, S., Alexandrou, M.A., Taggart, J.B., *et al.* (2018) Whole genome duplication and transposable element proliferation drive genome expansion in Corydoradinae catfishes, *Proceedings Of The Royal Society B*, 285(1872). https://doi.org/10.1098/rspb.2017.2732.

Marek, P.E. and Bond, J.E. (2009) A Müllerian mimicry ring in Appalachian millipedes, *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 106(24), pp. 9755–9760. https://doi.org/10.1073/pnas.0810408106.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads, *EMBnet.Journal*, 17(1), pp. 10–12. https://doi.org/10.14806/ej.17.1.200.

Merrill, R.M., Dasmahapatra, K.K., Davey, J.W., *et al.* (2015) The diversification of Heliconius butterflies: what have we learned in 150 years?, *Journal Of Evolutionary Biology*, 28(8), pp. 1417–1438. https://doi.org/10.1111/jeb.12672.

Mistry, J., Chuguransky, S., Williams, L., *et al.* (2021) Pfam: The protein families database in 2021, *Nucleic Acids Research*, 49(1), pp. 412–419. https://doi.org/10.1093/nar/gkaa913.

Motyka, M., Kampova, L. and Bocak, L. (2018) Phylogeny and evolution of Müllerian mimicry in aposematic Dilophotes: evidence for advergence and size-constraints in evolution of mimetic sexual dimorphism, *Scientific Reports*, 8(1), pp. 1–10. https://doi.org/10.1038/s41598-018-22155-6.

Müller, F. (1878) Über die vortheile der mimicry bei schmetterlingen, *Zoologischer Anzeiger*, 1, pp. 4–55.

Nakagawa, H., Nagasaka, K., Sakai, H., *et al.* (2015) Isolation of a novel lectin from the dorsal spines of the devil stinger, Inimicus japonicus, *International Aquatic Research*, 7(2), pp. 143–150. https://doi.org/10.1007/s40071-015-0101-2.

Oh, J., Lee, S.-G. and Park, C. (2021) PIC-Me: paralogs and isoforms classifier based on machine-learning approaches, *BMC Bioinformatics*, 22. https://doi.org/10.1186/s12859-021-04229-x.

Okumu, M.O., Mbaria, J.M., Gikunju, J.K., *et al.* (2020) Enzymatic activity and brine shrimp lethality of venom from the large brown spitting cobra (Naja ashei) and its neutralization by antivenom, *BMC Research Notes*, 13(1), p. 325. https://doi.org/10.1186/s13104-020-05167-2.

de Oliveira Júnior, N.G., Fernandes, G. da R., Cardoso, M.H., *et al.* (2016) Venom gland transcriptome analyses of two freshwater stingrays (Myliobatiformes: Potamotrygonidae) from Brazil, *Scientific Reports*, 6(1). https://doi.org/10.1038/srep21935.

Patro, R., Duggal, G., Love, M.I., *et al.* (2017) Salmon provides fast and bias-aware quantification of transcript expression, *Nature Methods*, 14(4), pp. 417–419. https://doi.org/10.1038/nmeth.4197.

Pekár, S., Bočánek, O., Michálek, O., *et al.* (2018) Venom gland size and venom complexity-essential trophic adaptations of venomous predators: A case study using spiders, *Molecular Ecology*, 27(21), pp. 4257–4269. https://doi.org/10.1111/mec.14859.

R Core Team (2023) R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

von Reumont, B.M., Anderluh, G., Antunes, A., *et al.* (2022) Modern venomics-Current insights, novel methods, and future perspectives in biological and applied animal venom research, *GigaScience*, 11. https://doi.org/10.1093/gigascience/giac048.

Ritchie, M.E., Phipson, B., Wu, D., *et al.* (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies, *Nucleic Acids Research*, 43(7). https://doi.org/10.1093/nar/gkv007.

de Roodt, A.R., Boyer, L.V., Lanari, L.C., *et al.* (2016) Venom yield and its relationship with body size and fang separation of pit vipers from Argentina, *Toxicon*, 121, pp. 22–29. https://doi.org/10.1016/j.toxicon.2016.08.013.

Rowland, H.M., Mappes, J., Ruxton, G.D., *et al.* (2010) Mimicry between unequally defended prey can be parasitic: evidence for quasi-Batesian mimicry, *Ecology Letters*, 13(12), pp. 1494–1502. https://doi.org/10.1111/j.1461-0248.2010.01539.x.

Ruxton, G.D., Allen, W.L., Sherratt, T.N., *et al.* (2018) *Avoiding attack : the evolutionary ecology of crypsis, warning signals, and mimicry*. Oxford, United Kingdom: Oxford University Press.

Sherratt, T.N., Roberts, G. and Kassen, R. (2009) Evolutionary stable investment in products that confer both an individual benefit and a public good, *Front Biosci*, 14(12), pp. 4557–4564. https://doi.org/10.2741/3548.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., *et al.* (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, *Bioinformatics* , 31(19), pp. 3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

Smith, E.G., Surm, J.M., Macrander, J., *et al.* (2023) Micro and macroevolution of sea anemone venom phenotype, *Nature Communications*, 14(1). https://doi.org/10.1038/s41467-023-35794-9.

Smith, W.L. and Wheeler, W.C. (2006) Venom evolution widespread in fishes: a phylogenetic road map for the bioprospecting of piscine venoms, *The Journal Of Heredity*, 97(3), pp. 206–217. https://doi.org/10.1093/jhered/esj034.

Soneson, C., Love, M.I. and Robinson, M.D. (2015) Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences, *F1000Research*, 4, p. 1521. https://doi.org/10.12688/f1000research.7563.2.

Speed, M.P. (1993) Müllerian mimicry and the psychology of predation, *Animal Behaviour*, 45(3), pp. 571–580. https://doi.org/10.1006/anbe.1993.1067.

Speed, M.P. and Turner, J.R.G. (1999) Learning and memory in mimicry: II. Do we understand the mimicry spectrum?, *Biological Journal Of The Linnean Society*, 67(3), pp. 281–312.

Symula, R., Schulte, R. and Summers, K. (2001) Molecular phylogenetic evidence for a mimetic radiation in Peruvian poison frogs supports a Müllerian mimicry hypothesis, *Proc Biol Sci*, 268(1484), pp. 2415–2421. https://doi.org/10.1098/rspb.2001.1812.

Wickham, H. and Henry, L. (2023) purrr: Functional Programming Tools. https://CRAN.R-project.org/package=purrr.

Wong, E.S.W. and Belov, K. (2012) Venom evolution through gene duplications, *Gene*, 496(1), pp. 1–7. https://doi.org/10.1016/j.gene.2012.01.009.

Wright, J.J. (2009) Diversity, phylogenetic distribution, and origins of venomous catfishes, *BMC Evolutionary Biology*, 9. https://doi.org/10.1186/1471-2148-9-282.

Xie, B., Li, X., Lin, Z., *et al.* (2016) Prediction of toxin genes from Chinese yellow catfish based on transcriptomic and proteomic sequencing, *International Journal Of Molecular Sciences*, 17(4). https://doi.org/10.3390/ijms17040556.

Zhu, B., Jin, P., Zhang, Y., *et al.* (2023) Genomic and transcriptomic analyses support a silk gland origin of spider venom glands, *BMC Biology*, 21(1). https://doi.org/10.1186/s12915-023-01581-7.

Ziegman, R., Undheim, E.A.B., Baillie, G., *et al.* (2019) Investigation of the estuarine stonefish (Synanceia horrida) venom composition, *Journal Of Proteomics*, 201, pp. 12–26. https://doi.org/10.1016/j.jprot.2019.04.002.

5 The molecular basis of sex-limited Müllerian mimicry in *Corydoras simulatus.*

5.1 Abstract

Sex-specific mimicry allows the role of predation and sexually antagonistic selection on a phenotype to be explored in a single system. This study investigates male-limited mimicry in *Corydoras simulatus*, where the males are likely to be Müllerian mimics of another Corydoradinae species. We used both low-coverage whole genome resequencing and differential expression methodologies to identify genes associated with sex-specific pigmentation patterns. A set of 4 genes were identified as pigment gene candidates. Although the candidates were not previously known to play a role in teleost pigmentation, one is involved cAMP signalling, whilst others have cell movement functions, and therefore, represent novel pigment genes. Some known pigmentation related genes (*mc1r* and *tyr1*) were found to be differently expressed in melanic vs non-melanic skin tissue in female *C. simulatus,* perhaps indicating the interaction of our candidates with known pigment pathways. Additionally, two genes related to sex differentiation were identified in the genomic data, although neither were differentially expressed between males and females. These represent candidates for sex determination in the Corydoradinae. This study lays the groundwork for the first vertebrate system which can be utilised to investigate the role of sexually antagonistic selection in contrasting anti-predation strategies at the genetic level.

5.2 Introduction

Sexual dimorphism is widespread in the animal kingdom, causing dramatic phenotypic variation within the same species (Eyer, Blumenfeld and Vargo, 2019). It was the observation of this variation that inspired Darwin's theory of sexual selection (Darwin, 1859, 1871; Petrie, 2021). Sexual dimorphism occurs due to differential selection between sexes within the same species, known as sexually antagonistic (SA) selection (Rowe, Chenoweth and Agrawal, 2018). Sexually antagonistic selection arises when a phenotypic trait has a shared genetic basis but a different optimum in each sex. Intersexual selection can be driven by sexual selection, e.g. traits relating to mate choice preferences, or natural selection due to disparate ecological pressures between the sexes (Tosto *et al.*, 2023). Unequal predation pressure has been shown to drive sexual dimorphism, either in concert with sexual selection or independently. For example, female three-spined sticklebacks (*Gasterosteus aculeatus*) experience higher avian predation than males, which has selected for additional dorsal and pelvic spines (Reimchen and Nosil, 2004). Secondary sexual characteristics can also drive differential

predation rates, thus demonstrating a trade-off of increased mating opportunities (sexual selection), or reduced predation risk (natural selection). This trade-off has been well explored in guppies (*Poecilia reticulata*) where low predation streams are characterised by more conspicuous colouration, which has been shown to change on introduction of a predator (Endler, 1978, 1983).

Colour pattern mimicry comes in different forms being broadly defined as 'protective', 'reproductive' and 'aggressive' (Pasteur, 1982). Reproductive mimicry, where the mimic benefits from increased reproduction opportunities, commonly drives sexual dimorphism (Pasteur, 1982). In the Bluehead wrasses (*Thalassoma bifasciatum*), large territorial males secure the most reproductive opportunities (Todd *et al.*, 2018). However, by mimicking females, smaller males are able to sneak copulation, increasing their reproductive output (Todd *et al.*, 2018). Additionally, protective mimicry has been shown to result in intersexual phenotypic variation (Krebs and West, 1988; Maisonneuve, Smadi and Llaurens, 2022). There are two primary forms of protective mimicry, Batesian and Müllerian mimicry. Batesian mimicry occurs when a palatable species adverges on the appearance of an unpalatable model (Bates, 1862; Ruxton *et al.*, 2018). Müllerian mimicry, however, occurs when two unpalatable species share a resemblance and benefit from the reduced predator learning cost associated with the shared phenotype (Sanders, Malhotra and Thorpe, 2006; Ruxton *et al.*, 2018). Both types of mimicry are characterised by conspicuous warning colouration, advertising either honest or dishonest signals of unpalatability (Caro and Ruxton, 2019).

Female-limited Batesian mimicry is a phenomenon commonly seen in Lepidoptera (Joron and Mallet, 1998; Ohsaki, 2005; Maisonneuve, Smadi and Llaurens, 2022). Batesian mimicry is a parasitic interaction between the model and the mimic, where the palatable mimic corrodes the protection afford by the defended model (Ruxton *et al.*, 2018; Font, 2019). As only one sex is mimetic, the census population size of the mimic is halved and therefore, the burden on the model is reduced (Joron and Mallet, 1998; Ohsaki, 2005). However, this does not explain the lack of male-limited Batesian mimetic systems observed in nature. Explanations for this in Lepidoptera include female preference for non-mimetic males and ecological characteristics of females, including behaviours associated with ovipositing and egg production (e.g. flying speeds and increased stationary time) (Ohsaki, 2005; Low and Monteiro, 2018). In the female-limited Batesian mimic, Eastern tiger swallowtail butterfly, *Papilio glaucus,* females prefer males with painted ancestral wing patterns to those with a painted mimetic wing pattern, suggesting the lack of male mimicry is due to the loss of potential mating opportunities (Krebs and West, 1988; Maisonneuve, Smadi and Llaurens, 2022).

In comparison, sex-limited Müllerian mimicry is rare. This may, in part, be due to the mutualistic nature of Müllerian mimicry systems, where both taxa sharing a resemblance benefit from reduced predator learning costs (Müller, 1878; Joron and Mallet, 1998). However, sex-limited Müllerian mimicry is not entirely absent. It has been hypothesised in some butterfly systems, including female-limited mimicry in *Argyreus hyperbius* and *Atrophaneura alcinous* (Nishida, 2017). Both species were initially thought to be Batesian mimics, but were found to possess toxins (Nishida, 2017). Whilst the drivers of female-limited mimicry in these systems are not known, they may be similar to those outlined above *Papilio* systems (Nishida, 2017; Maisonneuve, Smadi and Llaurens, 2022). Alternatively, in the Asian pit vipers (*Trimeresurus spp*), mimicry is male-limited, which is thought to be driven by higher predation pressure on males (Sanders, Malhotra and Thorpe, 2006).

Sexual dimorphism occurs due to differing selection pressures between sexes of the same species, known as sexually antagonistic (SA) selection (Rowe, Chenoweth and Agrawal, 2018). Sexually antagonistic selection arises when a phenotypic trait has a shared genetic basis but a different optimum in each sex (Ruzicka *et al.*, 2020). As such, SA can promote genetic diversity within natural populations. This can occur between sexes at the same locus (intralocus), or between sexes at different loci (interlocus)(Göran Arnqvist and Locke Rowe, 2005). Interlocus conflict occurs when there is a disparity in the desired outcome of a male-female interaction (Chapman *et al.*, 2003; Göran Arnqvist and Locke Rowe, 2005). Conversely, intralocus conflict occurs when fitness optima differ for a trait that is expressed in both sexes (Chapman *et al.*, 2003; Rowe, Chenoweth and Agrawal, 2018). Resolutions to intralocus sexual conflict include sex-specific phenotypic expression and linkage to sex determining regions, resulting in dimorphism between the sexes (Cox and Calsbeek, 2009). To achieve this, loci often become located in a region unique to the sex for whom they benefit (Kottler and Schartl, 2018). For example, linkage to a novel sex determiner resulted in a female-limited cryptic orange blotch phenotype in some Lake Malawi cichlid species (Roberts, Ser and Kocher, 2009).

The Corydoradinae are a speciose group of freshwater catfish, distributed across South America, with over 200 described species (Dias *et al.*, 2024). All Corydoradinae are equipped with venom glands accompanied by sharp lockable pectoral and dorsal spines (Greven, Flasbecl and Passia, 2006). Colour patterns vary drastically across the subfamily, containing both cryptic and conspicuous elements (Alexandrou *et al.*, 2011). Additionally, many species of Corydoradinae participate in Müllerian mimicry rings, in which species shoal together (Alexandrou *et al.*, 2011). Sexual dimorphism in the Corydoradinae is largely restricted to differences in body size, with females being larger and having a rounder and deeper body (Kohda *et al.*, 2002; Fuller, 2012). However, a few species of Corydoradinae show sexual dimorphism regarding colour pattern (Weitzman and Nijssen, 1970; Fuller, 2012; Bono *et*

al., 2019). *Corydoras simulatus* is a polymorphic species found in the upper Meta river basin in Colombia, showing clear sexual dimorphic pigmentation (Weitzman and Nijssen, 1970; Fuller and Evers, 2005). The males display a mimetic phenotype (Figure 5.1A), resembling the sympatric *Hoplisoma metae* (Figure 5.1B)(Alexandrou *et al.*, 2011). Females have, however, a cryptic phenotype, representing an alternative anti-predation strategy (Figure 5.1C)(Sands, 1994; Ruxton *et al.*, 2018).

The adoption of different anti-predation strategies between the sexes of *C. simulatus* offers a powerful system in which to investigate the genetics of Müllerian mimicry. Additionally, it also offers a rare example of male-limited Müllerian mimicry, which can be leveraged to understand how these systems evolve and why they are not more pervasive. Here we aim to investigate the genetics underpinning the alternative phenotypic strategies adopted by *C. simulatus* males and females, using a multi-omics approach. Specifically, we aim to

1) identify candidate genes associated with elevated genetic differentiation between the sexes, using low coverage whole genome resequencing. This will also allow us to identify candidate genes for both coloration and other sex-based differences simultaneously.
2) further explore potential roles for identified candidate genes in pigmentation by investigating expression differences between melanated and non-melanated regions in cryptic females.
3) investigate sex-specific gene expression, underpinning pigmentation, by exploring differential expression in scute (skin) tissue from both males and females.

5.3 Methods

5.3.1 Sample Collection and DNA and RNA Sequencing

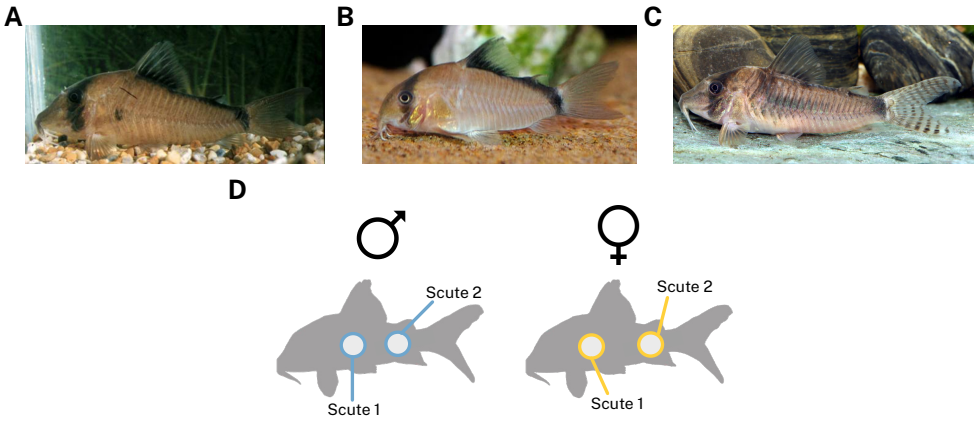


Figure 5.1 A Mimetic phenotype of *Corydoras simulatus* males. **B** Co-mimic of *C. simulatus* males, *Hoplisoma metae*. **C** Cryptic phenotype of *C. simulatus* females. **D** Schematic of the tissue locations used in RNA sequencing of *C. simulatus.* Pictures courtesy of Ian Fuller.

In preparation for RNA sequencing, two scute tissue samples (Scute 1 and Scute 2, Figure 5.1D) were excised and immediately flash frozen using liquid nitrogen. The scute is a thick dermal armour made of overlapping panels, covering the length of the body of in Corydoradinae individuals (Lowe *et al.*, 2021). The skin is attached to the internal scute tissue (Lowe *et al.*, 2021). We utilised a within individual comparison (Female Scute 1 vs Female Scute 2) as well as between sex comparison (Female Scute 1 vs Male Scute 1). RNA was extracted using an adapted RNAeasy micro or mini kit (Qiagen, Germany) protocol with a prior Trizol step as detailed (http://www.untergasser.de/lab/protocols/rna_prep_comb_trizol_v1_0.htm). The Qubit™ 4 fluorometer was used to assess RNA quality and quantity (Thermofisher Scientific, USA), following the protocol associated with the Qubit™ RNA Broad Range Quantification Assay kit (Thermofisher Scientific, USA). A volume of 20µl at a minimum concentration of 20 ng/µl was sent to be sequenced. Quality control and library preparation for Illumina sequencing were performed at Novogene Cambridge, using NovaSeq 6000 and each sample sequenced to 40 million reads.

Live, wild caught individuals *of Corydoras simulatus* were obtained from two independent aquarium sources in the UK (Batch01, Batch02). Each individual was sexed based on colour pattern in life and, after euthanasia, examination of the internal gonads. DNA was extracted from fin clips using the Qiagen DNeasy Tissue Kit (Qiagen, Germany), following the manufacturer's protocols. Quality and quantity were determined using the Qubit™ 4 fluorometer (Thermofisher

Scientific, USA), following protocol associated with the Qubit™ dsDNA Broad Range Quantification Assay kit (Thermofisher Scientific, USA). Gel electrophoresis was performed to assess DNA fragmentation, and a minimum volume of 20 µL at a concentration of 10 ng/µL was prepared for sequencing. Library preparation and sequencing of individuals was completed at Novogene, Cambridge, using NovaSeq 6000 at a desired depth of 5x per individual.

5.3.2 Genomic sequence filtering, alignment, and genotype likelihood estimation.

Raw reads were examined using FastQC (v. 0.11.9) (Andrews, 2010) and trimmed and filtered following Genome Analysis Toolkit pre-processing pipeline via a custom script (available at https://github.com/EmilyPhelps/CorydoradinaeMimicry2024). Low quality sequences and adapters were removed using the Picard (v. 2.20.2-11) (https://broadinstitute.github.io/picard/) tools MarkIlluminaAdapters and SamToFastq, before alignment to the *C. fulleri* reference genome using BWA MEM (v. 0.7.17) (Li and Durbin, 2009) and merged with unmapped bam files to retain all relevant sample and quality information using Picard MergeBamAlignment. When samples had been sequenced across different lanes, they were merged using SAMtools merge (v.1.9)(Danecek *et al.*, 2021). Subsequently, optical and sequencing duplicates were removed using Picard MarkDuplicates and overlapping reads were clipped by BamUtils (v. 1.0.15) clipOverlap (Jun *et al.*, 2015). RealignerTargetCreater and IndelRealigner from the Genome Analysis ToolKit (v. 3.8-1-0) were then used to realign regions around indels (Van der Auwera *et al.*, 2013). Finally, average read depth was calculated per sample using SAMtools depth.

Genotype likelihoods were estimated using ANGSD (v. 0.939-8)(Korneliussen, Albrechtsen and Nielsen, 2014), implementing the GATK model (-GL 2). Estimating the genotype likelihoods based on mapping and sequence quality scores accounts for a level of uncertainty associated with low coverage resequencing data (Korneliussen, Albrechtsen and Nielsen, 2014; Lou *et al.*, 2021). The *Corydoras fulleri* assembly was used as both the reference and the ancestral (-ref/anc). A maximum depth of 128 was calculated (-setMaxDepth). This threshold was estimated by multiplying the average read depth per sample by the total number of samples present. Sites were excluded if they were represented in less than half of the individuals (-minInd). The remaining sites were filtered for quality score of > 20 (-minQ 20), a minimum mapping quality (-minMapQ 30), and for sites that mapped uniquely (-uniqueOnly). SNPs were first called globally across both phenotypes and populations to obtain a site list of all variable sites. Sites were considered SNPs when they had a probability of <1e-6 (-SNP_pval 1e-6) of being monomorphic. Allele frequences were calculated using fixed major and minor alleles (-doMAF) and sites with a minor allele frequency below 0.05 (-minMAF 0.05) were retained. SNPs were then called per population, per phenotype under the same conditions, on variable sites previously

identified (-sites). Additionally, genotype likelihood was estimated without calling SNPs both globally and per phenotype and population to gain both variant and invariant sites needed for some downstream population genetic analysis.

5.3.3 PCA

To explore population structure, variant sites were called globally as detailed above using ANGSD. Linked sites were then identified using PLINK (v. v1.90b6.7)(Purcell *et al.*, 2007) with a window size of 10, a SNP number of 5 and a variance influence factor of 2. A linkage site list was then produced using the Rscript (v4.2.3) make_site_list_pruned.r (https://github.com/clairemerot/angsd_pipeline) (Mérot *et al.*, 2021; R Core Team, 2023). Genotype likelihoods were re-calculated for the remaining sites using ANGSD, and a covariance matrix was generated using PCAngsd (v.1.10)(Meisner and Albrechtsen, 2018). Eigenvalues were calculated from this matrix using the base function eigen() in R. This was then repeated for each batch, to identify patterns of differentiation within the batches.

5.3.4 Genome-wide F_{ST}

F_{ST} was estimated in genomic windows by first calling SNPs independently for phenotypes within each population, as described above, but using the -doGeno 2 option. This option allocates the most likely genotype to a SNP. The resulting bcf files were then converted into vcf format, compressed and indexed using BCFtools (v. 1.10.2)(Danecek *et al.*, 2021). Next the script parseVCFs.py was used to convert the vcf files into the correct format to be read into the popgenWindows.py script (both available at https://github.com/simonhmartin/genomics_general). This allows windows of a fixed SNP density to be estimated reducing the overestimation of F_{ST} in windows with few SNPs, resulting in erroneous peaks. Windows were set to 150 SNPs or 50kb, depending on which criteria was met first. Windows were considered outliers if they were in the top 0.3 percentile. Shared regions of genetic differentiation between the two batches were then investigated. Windows that were outliers in both populations and were within 50kb of each other were considered overlapping. This was achieved in R by decreasing the start of an outlier window by 50kb and increasing the end of a window by 50kb. Windows from each batch were then converted into a Genomic Ranges object using the makeGRangesFromDataFrame(), followed by the reduce() functions (to combined adjacent windows) provided by GenomicRanges (v. 1.50.2)(Lawrence *et al.*, 2013). Overlapping windows were then identified using the function subsetByOverlaps(), also provided by GenomicRanges. Regions that were overlapping between the two batches were then used in candidate gene identification.

5.3.5 Transcriptome assembly, annotation and pseudo-alignment of reads

A *C. simulatus* transcriptome was created using all the scute tissue samples, from both sexes along with additional *C. simulatus* auxiliary (venom) gland samples obtained from another study (see

previous chapter). All raw sequences from *C. simulatus* were concatenated, trimmed, adapters were removed and filtered for a phred score of 20 using cutadapt (v. 2.10)(Martin, 2011). The resulting sequences were visually inspected using FastQC (v.0.11.8)(Andrews, 2010) before being *de novo* assembled using Trinity (v.2.11.0) (Grabherr *et al.*, 2011) using the default settings. The resulting transcriptome was then annotated using Trinnotate (v.4.0.2)(Bryant *et al.*, 2017) with the UniProt (UniProt Consortium, 2021) and Pfam (Mistry *et al.*, 2021) databases using tblastx and blastp from BLAST(v. 2.10.1+)(Camacho *et al.*, 2009) and HMMER (v.3.3)(http://hmmer.org/).

5.3.6 Differential expression analysis

Abundance of transcripts was estimated from the trimmed and filtered scute samples using Salmon (v.1.10.0)(Patro *et al.*, 2017), via the align_andestimate_abundance.pl script from Trinity. TXimport was used to normalize transcript abundance by gene length in R (v. 1.26.1)(Soneson, Love and Robinson, 2015). Gene models were filtered to remove those where less than half of samples had a count of 50. Subsequently, DESeq2 (v. 1.38.3)(Love, Huber and Anders, 2014) was used to perform differential expression analysis between 1) Female Scute 1 and Female Scute 2, and 2) Female Scute 1 and Male Scute 1. Annotations were assigned to differentially expressed genes in R using the SwissProt blastp hit from the annotation, or if missing, the top SwissProt tblastx hit. When multiple hits were assigned, the hit with the lowest evalue and highest percentage identity were retained. Genes identified as having annotations belonging to prokaryotes were removed. Gene level inferences were estimated using the map file provided by Trinity in the assembly process and the tx2gene() function from TXImport(). Transcript counts were normalised using vst() from the DESeq2 package before plotting.

5.3.7 Identification of candidate genes

We first utilised the genome annotation for *C. fulleri* (corydoras_fulleri_v1, Chapter 3) to identify potential candidate genes. First, the annotation file was filtered in R to contain only mRNA features. These were then converted into a GenomicRanges object using makeGRangesFromDataFrame() and only genes that overlapped with the outlier regions were retained. This was achieved using the countOverlaps() function and filtering for genes with an overlap. The resulting genes were given allocated functions and a Uniprot accession via a blastp (v. 2.10.1+) (Camacho *et al.*, 2009) against the UniProt vertebrate database (UniProt Consortium, 2021) as part of the functional annotation process outlined in earlier chapters. A gene name, protein name and biological process GO term was provided using the id mapping function on uniprot.com (UniProt Consortium, 2021). Genes were considered to be good candidates if they were supported by both the genomic and expression analyses. To identify

genes that were within the outlier regions and had differential expression we used the intersect() function in base R.

To identify any candidate genes that were known pigment genes, we compiled a database from Baxter *et al.,* (2019) and Lorin *et al*., (2018). Additionally, gene ontogeny (GO) biological process terms were searched for terms relating to pigmentation including melanocyte, iridophore, xanthophore, melanin, melanosome, keratinocyte, pigment, pteridine, carotenoid, purine, fibroblast, melanoblast, albinism, colour, color, erythrophore, leucophore and cyanophore. Similarly, we curated a list of teleost sex determination genes from Kitano *et al.,* (2024) and Bhattacharya and Modi (2021) and a list of GO search terms related to sex determination/ differentiation including estrogen, progesterone, testis, gonad, sex, allosome, ovary, puberty, androgen, sperm, egg and steroidogenesis.

5.3.8 Tajima's D and D_{XY}

Candidates were identified only in the outlier region on Scaffold 9 (see results), so Tajima's D and D_{XY} was visualised only for this region. Site allele frequency likelihoods were estimated from the genotype likelihoods assuming Hardy–Weinberg equilibrium (-doSaf 1) for each sex in each batch using ANGSD. Next, a folded (-fold 1) SFS file was generated for each sex per batch using RealSFS, implemented in ANGSD. Theta was then calculated using RealSFS saf2theta. Finally, thetaStat do_stat (ANGSD) was utilised to calculate Tajima's D in sliding windows of 10kb across the outlier region on Scaffold 9.

Estimating D_{XY} required us to run ANGSD twice. This methodology ensures sites that are variants within batches but are fixed within sexes are captured. ANGSD was initially run across both batches as described above, but with a minimum individual filter of 80% of samples as D_{XY} can be biased by missing data (Korunes and Samuk, 2021). This initial run was used to create a site file, which was indexed using ANGSD sites index. A second run of ANGSD was performed on each batch on each sex separately, without the SNP calling filters but with the new site file to capture both variant and invariant sites. Next, per-site D_{XY} was calculated using the script calcDxy.R by J. Penalba (available with ngsPopGen). Windowed D_{XY} was finally then calculated using the script angsd_dxy_step3_processoutputoverwindows_04242020.pl (https://github.com/ksamuk/pixy_analysis/blob/ab90cc1e95513901010a15fb260febfbc5604b9a/data_generation/angsd/angsd_dxy_step3_processoutputoverwindows_04242020.pl#L4)(Korunes and Samuk, 2021).

5.3.9 SNP density

SNP density was calculated in non-overlapping 10kb windows. The genome was partitioned into windows using the slidingwindows() function in the GenomicRanges R package (v. 1.50.2)(Wang et al.,

2012) with both the window and step size set to 10000. Per window SNPs density was calculated using the CountOverlaps() function, also from GenomicRanges, for each sex in each batch.

5.4 Results

A total of 27 individuals were sequenced in two batches. In Batch01, the total number of individuals sequenced was 15, with 2 males and 13 females. In Batch02, the total number of individuals sequenced was 12, with 3 males and 9 females. Initial sequencing produced an average of 34,126,898 reads (median= 33,832,964, range=26,108,028-44,952,834), with an average duplicate read percentage of 17.66 % (median=17.67 %, range=12.61%-22.90%) and a GC content of 42.65% (median=43.00%, range=41.00%- 44.00%). Trimming and alignment resulted in a mapping rate of 60.71% (median=61.24%, range=56.05%-63.22%), with a mean of 97.06% of the filtered reads being uniquely mapped. Duplicate reads were reduced to 0%, through the duplicate removal filtering step. Reads were mapped to an average depth of 3.648 ± 10.806 (median= 3.681 ± 10.913, range=2.731-4.933) across both batches, covering 69.7% of the genome on average. Variant calling identified a total of 3,592,628 SNPs.

Principal component analysis of both batches (Figure 5.2) revealed clustering among individuals in each of the two batches along PC1 which explained 11.49% of the variation. Individuals in Batch01 also formed a cluster on PC2, whereas Batch02 showed more variation, although this only accounted for 3.66% of the total variation. Principal component analysis was then run on each batch independently. Both batches demonstrated some internal structuring, but Batch02 was not more structured than Batch01, despite the global PCA (Supplementary Figure 5.1). Due to the population structuring identified in the PCA, subsequent population genetic analyses were performed on each batch independently.

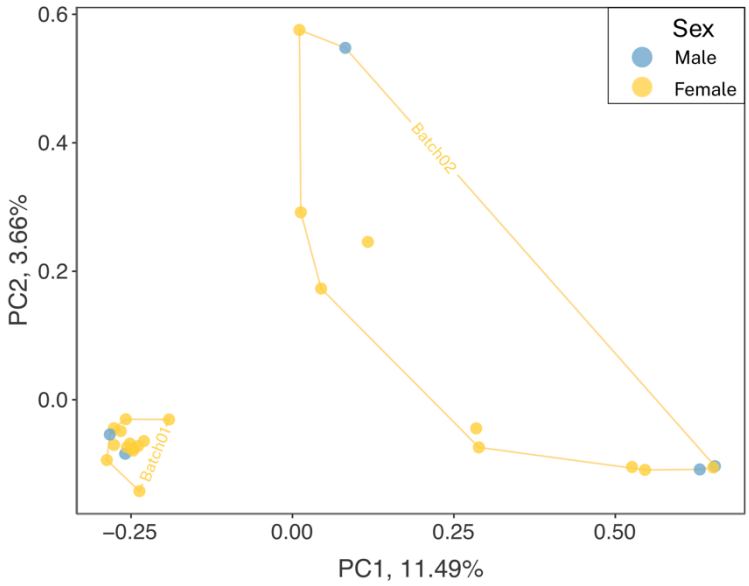


Figure 5.2 Principal component analysis of genomic variation between male and female *Corydoras simulatus* samples, from two aquarium sources. Males are highlighted in blue, whilst females are highlighted in yellow. The convex hull surrounding the two main clusters denotes the source the samples originated from.

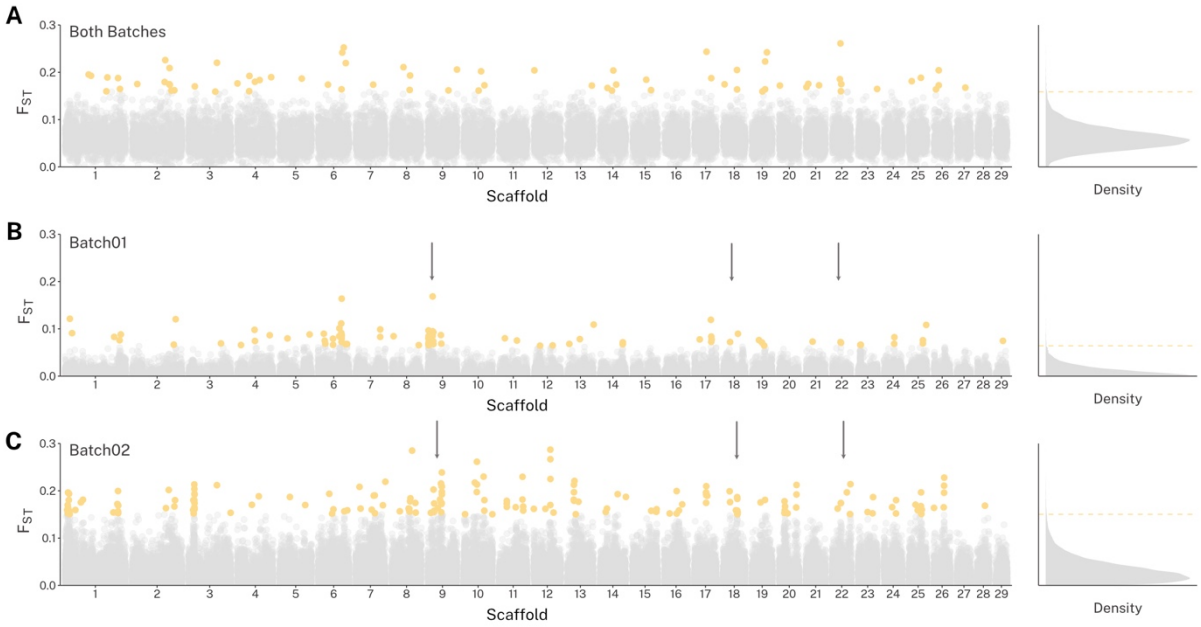


Figure 5.3 Genome-wide windowed F_{ST} between male and female *Corydoras simulatus*. Windows are of either 150 SNPs or 50kb, which ever criteria is met first. Outlier windows are denoted in yellow; these represent the top 0.3% of windows. Histograms represent the distribution of windowed $F_{ST,}$ and the hashed yellow line is the threshold above which values are considered outliers. Arrows denote the overlapping peaks between batches. **A)** F_{ST} was calculated in between sexes (male= 5, female= 22) of both batches. **B)** F_{ST} was calculated between sexes (male= 2, female= 13) in Batch01. **C)** F_{ST} was calculated between sexes (male= 3, female= 9) of Batch02.

Genome-wide estimates of F_{ST} (Figure 5.3) were concordant with the global PCA. When both batches are combined to estimate F_{ST} between sexes, average F_{ST} per window is higher than in each batch alone (Figure 5.3). As reflected by the global PCA, Batch02 had a higher level of F_{ST} between the sexes (Figure 5.3C). Shared outlier regions were identified between the sexes in the two batches on Scaffold 9, Scaffold 18 and Scaffold 22. The size of these shared regions varied, at 7,214,28bp, 18,764 bp, and 125,774bp respectively. Additionally, highly differentiated regions between the sexes were identified in Batch02 on Scaffold 1, 12, 13 and 26. Similarly, Batch01 contained an elevated region of F_{ST} on Scaffold 6. These were not explored in this study but could represent future lines of investigation.

Table 5.1 Candidate genes controlling pigmentation or sex differentiation in *Corydoras simulatus* present in the outlier region on Scaffold 9.

Gene	Protein Name	Evidence	Biological Process Gene Ontology (GO)
shmt1	Serine hydroxymethyltransferase, cytosolic	Overlapping between datasets	cellular response to leukemia inhibitory factor (GO:1990830) | folic acid metabolic process (GO:0046655) | glycine biosynthetic process from serine (GO:0019264) | glycine metabolic process (GO:0006544) | L-serine catabolic process (GO:0006565) | L-serine metabolic process (GO:0006563) | protein homotetramerization (GO:0051289) | purine nucleobase biosynthetic process (GO:0009113) | tetrahydrofolate interconversion (GO:0035999) | tetrahydrofolate metabolic process (GO:0046653)
angpt1	Angiopoietin-1	Overlapping between datasets	angiogenesis (GO:0001525) | cell differentiation (GO:0030154) | Tie signalling pathway (GO:0048014)
uqcrb	Cytochrome b-c1 complex subunit 7	Overlapping between datasets	mitochondrial electron transport, ubiquinol to cytochrome c (GO:0006122)
arpp21	cAMP-regulated phosphoprotein 21	Overlapping between datasets	
nvpf	FMRFamide related peptide 1	Sex related GO function	negative regulation of gonadotropin secretion (GO:0032277) | negative regulation of luteinizing hormone secretion (GO:0033685) | neuropeptide signaling pathway (GO:0007218) | positive regulation of circadian sleep/wake cycle, sleep (GO:0045938)
kank2	KN motif and ankyrin repeat domain-containing protein 2	Sex related GO function	apoptotic process (GO:0006915) | negative regulation of actin filament polymerization (GO:0030837) | negative regulation of cell population proliferation (GO:0008285) | negative regulation of G1/S transition of mitotic cell cycle (GO:2000134) | negative regulation of intracellular estrogen receptor signaling pathway (GO:0033147) | negative regulation of transcription by RNA polymerase II (GO:0000122) | negative regulation of vitamin D receptor signaling pathway (GO:0070563)

Shared outlier regions contained a total of 234 genes. Two genes were identified as candidate sex differentiation genes (Table 5.1). The first of the two sex differentiation candidates, *nvpf* (FMRFamide related peptide 1), has a known role in luteinizing hormone secretion and is associated with negative regulation of gonadotrophin function. A second candidate for sex differentiation is *kank2* (KN motif and ankyrin repeat domain-containing protein 2) which is associated with estrogen signalling.

To identify genes involved in increased pigmentation on the female flank, vital for crypsis, we quantified differences in gene expression patterns between the pigmented region on the female (Scute 1) and the non-pigmented region on the female (Scute 2). Additionally, we quantified differences between the pigmented region on the female (Scute 1) and the equivalent non-pigmented region in males (Scute 1) to understand sex specific differences. A total of 3 males and 5 females were

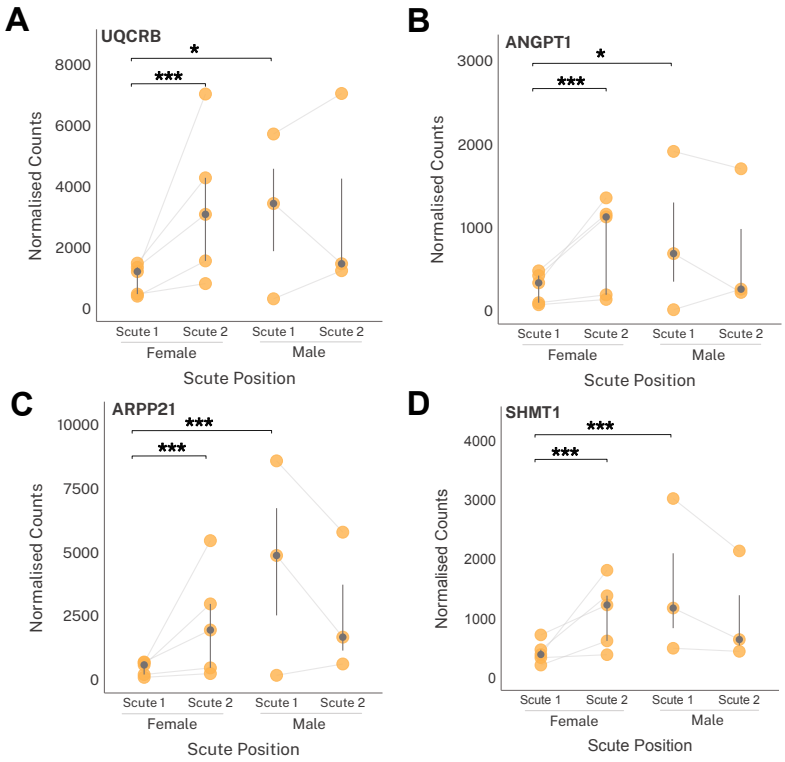


Figure 5.4 A-D Normalised gene counts per sample for differentially expressed genes between female Scute 1 and female Scute 2, female Scute 1 and male Scute 1 and were within the outlier regions identified by F_{ST} outlier analysis. Asterisks denote the degree of significance assigned by differential expression analysis. * p.value < 0.05, ** p.value < 0.01, *** p.value < 0.001

used in RNA sequencing, with a mean raw read number of 96,408,781 (range=25,064,528-313,112,608). Between the female scute tissues, a total of 1,439 genes were differentially expressed. Between sexes at the Scute 1 tissue location, 126 genes were differentially expressed. The within female comparison yielded 33 genes that have a known role in pigmentation, highlighting the legitimacy of this approach for identifying pigment genes (Supplementary Table 5.1). Four genes were identified as differentially expressed in both comparisons and occurring in the genomic outlier region (Figure 5.4, Table 1)

SNP density can be used to infer the level of degeneration of a sex-linked region in the heterogametic sex (Darolti *et al.*, 2019; Palmer *et al.*, 2019). Upstream of the *kank2* gene, males have a reduced SNP

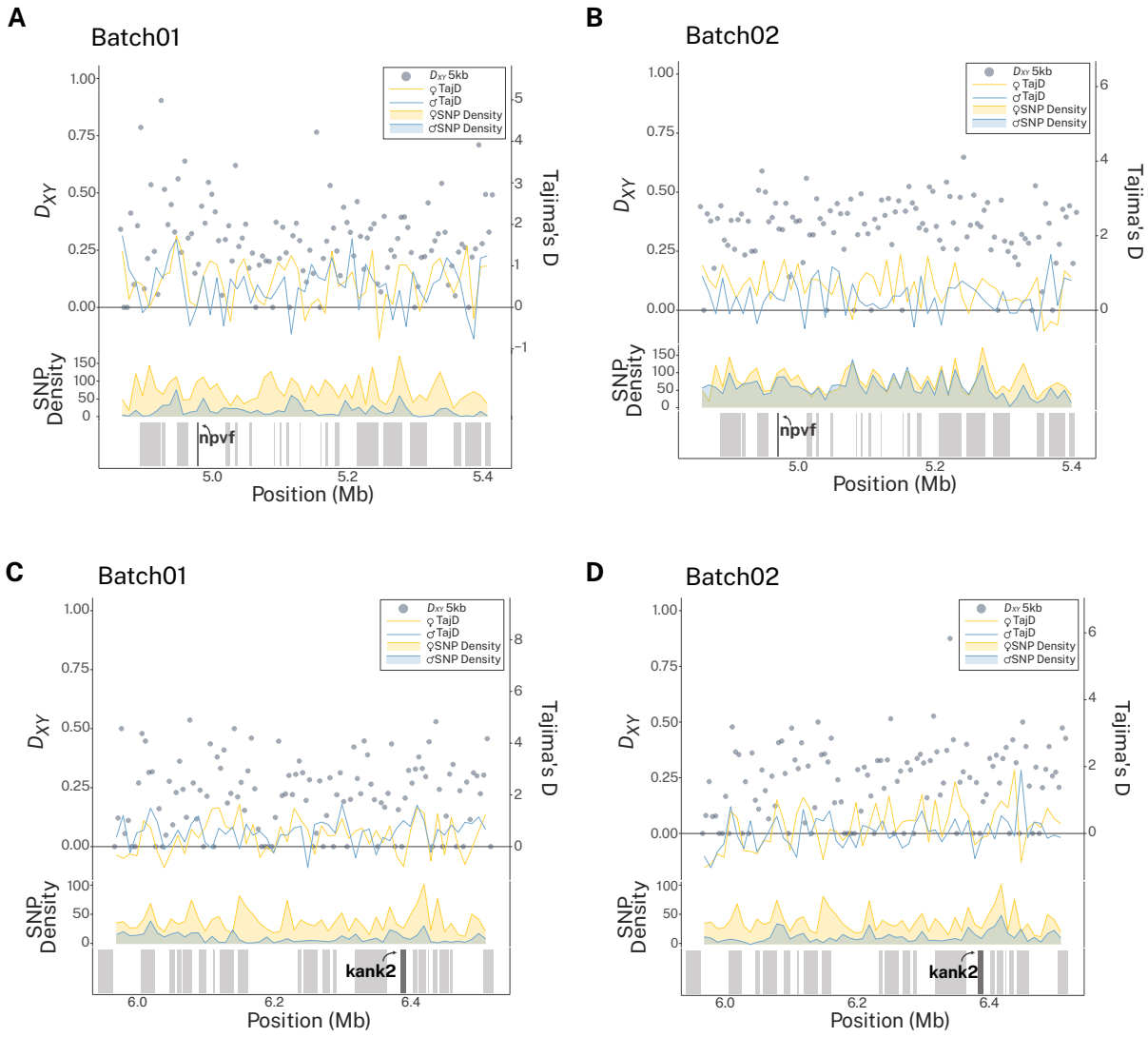


Figure 5.5 D_{XY} in 5kb windows, Tajima's D in 10kb windows, SNP density in 10kb windows and a schematic of candidate genes in the region for sex differentiation/determination from **A** and **C** Batches01, **B** and **D** Batches02.

density compared to females (Figure 5.5 C and D). This is true for both batches. This pattern is only seen in batch01 for the *nvpf* gene (Figure 5.5 A and B). We also see this lower SNP density in males in the pigment candidates, across both batches (Figure 5.6). Thus, if differences in SNP density is reflective of a sex determining region here, it may cover the whole region identified on Scaffold 9. Although we do not know the sex of our reference nor which sex is the homogametic, the reduced number of SNPs in the males indicate that there is a difference in this region that could be followed up with sex specific references and higher coverage resequencing of each sex to confirm the structure of the system.

We explored the levels of total genetic differentiation within a reduced region surrounding both our pigment and sex candidate genes in a 5Kb non-overlapping windows (Figures 5.5 and 5.6). These regions showed some elevation in D_{XY} compared to the genomic average, although overall patterns were variable. Additionally, we calculated Tajima's D in non-overlapping 10kb windows. These deviated from zero in both batches and patterns were generally complex. In the sex differentiation/ determination candidates, patterns of Tajima's D are highly variable (Figure 5.5). However, there is an elevated region of Tajima's D in both sexes upstream of *npvf* in both batches (Figure 5.5A-B). This region is also associated with an increase in D_{XY}. Similarly, there is an elevated region of Tajima's D in both sexes downstream of *kank2*, which is also accompanied by an increase in D_{XY} (Figure 5.5C-D). This is thought to be an indicator of balancing selection, which can be caused by sexually antagonistic

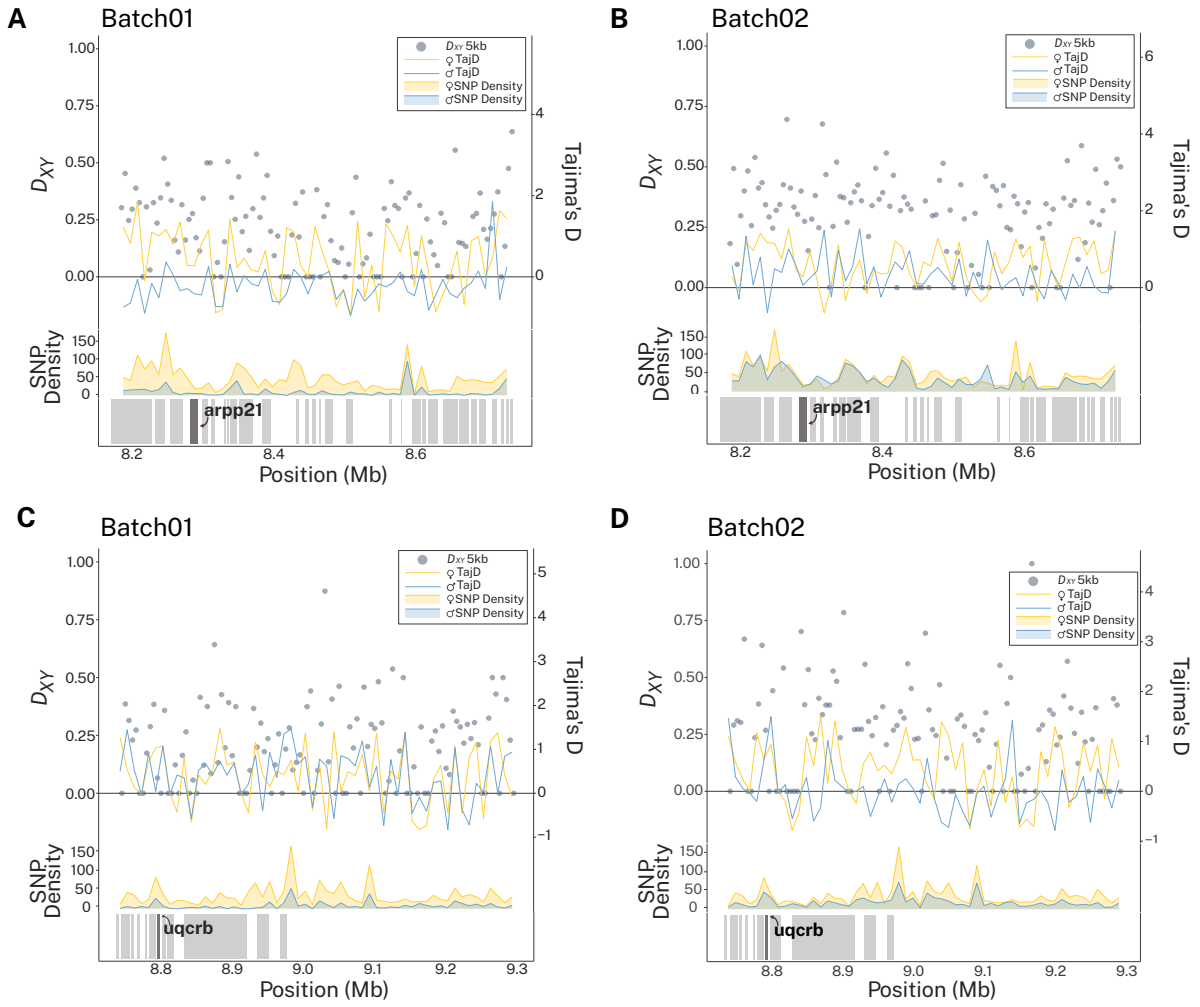


Figure 5.6 D_{XY} in 5kb windows, Tajima's D in 10Kb windows, SNP density in 10kb windows and a schematic of candidate genes in the region for pigment candidates from **A** and **C** Batch01, **B** and **D** Batch02

selection (Mank, 2017). Similar patterns of Tajima's D are seen in the pigment candidate *uqcrb* (Figure

5C-D), but these are only associated with elevated D_{XY} in batch02. In *arpp21*, a peak in Tajima's D is present in both sexes in Batch02 and associated with elevated D_{XY} but this is not the case in Batch01. This lack of concordance could be due to the lower sample size of males in Batch01.

5.5 Discussion

Here we have used a multi-omics approach to understand the genetics underpinning a rare occurrence of male-limited Müllerian mimicry in *C. simulatus*. We identified three regions with elevated genetic differentiation between males and females. Within the largest region (> 7mb) on Scaffold 9, we identified 4 genes that were differentially expressed between the melanated flank of the females, and the non-melanated region towards the caudal peduncle as well as in comparisons between the melanated flank of the females and the non-melanated flank of the males. These genes are likely to represent those that are involved in the presence or absence of melanin on the flank of an individual, which is vital to presenting either a female cryptic pattern or a male mimetic pattern. Whilst these genes are not known pigment genes, they do have some functions to suggest they could be co-opted into a pigment related role. Additionally, we identified two genes that represent candidate sex differentiation genes within the genetically differentiated region on scaffold 9, although these were not differentially expressed between males and females.

The different anti-predation strategies between male and female *C. simulatus* are evident from differences in their colouration. The female is characterised by a melanic blotch on the flank which fades into the tan body creating a low contrast camouflage pattern (Weitzman and Nijssen, 1970; Fuller and Evers, 2005; Ruxton *et al.*, 2018). Both sexes possess and eye bar and a blotch on the caudal peduncle as well as a melanated line along the dorsal perimeter of the body (Weitzman and Nijssen, 1970; Fuller and Evers, 2005). Males, however, lack the flank blotch and have much more defined edges to the eye bar, caudal peduncle blotch and melanated dorsal line (Weitzman and Nijssen, 1970; Fuller and Evers, 2005). Subsequently, the males share a resemblance to sympatric *Hoplisoma metae* (Weitzman and Nijssen, 1970; Fuller and Evers, 2005; Alexandrou *et al.*, 2011). Whilst the candidate pigment genes (Table 1) identified in this study do not have known roles in pigmentation, the down regulation of these genes in the melanated region of the female relative to the non-melanated regions in both males and females suggests these genes are involved in pigmentation in this system (Figure 5.4). As such, these genes represent novel candidate pigment genes. There is some evidence from the known roles of these genes to suggest they could be co-opted to perform pigmentation functions. For example, the gene *arpp21* is associated with cAMP signalling. cAMP is a key messenger involved in intracellular signalling and plays an important role in pigment motility and production (Parichy, 2021). The pigment gene *mc1r* induces cAMP production which

activates the melanin synthesis pathway involving the *tyrp1* protein (Buscà and Ballotti, 2000; Cal *et al.*, 2017). Whilst neither *mc1r* nor *tyrp1* were identified in our genomic region, they were identified as differentially expressed between female scute locations (Supplementary Table 5.1, *mc1r* : $p < .001$, *tryp1* : $p < 0.01$). This suggests that perhaps the novel pigment genes identified in this study are interacting with those in known pigmentary pathways.

Sex-limited Müllerian mimicry, especially male-limited mimicry, is exceptionally rare in nature, making *C. simulatus* a valuable example through which we can better understand how these systems evolve and why they are not more widespread (Joron and Mallet, 1998; Maisonneuve, Smadi and Llaurens, 2022). Differences in male and female behaviours, including those relating to reproduction, anti-predation and feeding could be driving the observed differences, as has been shown in the Asian pitvipers and *Papilio spp.*(Ohsaki, 2005; Sanders, Malhotra and Thorpe, 2006; Kunte, 2009). Work exploring the effect of pigmentation on behaviour in wild type and albino *Osteogaster aeneus* (a member of the Corydoradinae) individuals, found that in mixed shoals pigmented individuals showed greater foraging activity and sheltered more regularly (Svitačová *et al.*, 2024). In the wild, *C. simulatus* inhabits small fast flowing streams with a sandy substrate and a complex matrix of leaf litter and deadwood (Fuller and Evers, 2005). The more heavily pigmented females could spend more time sheltering and foraging among the leaf litter than the males, who may spend more time out in the open. As well as dimorphism in pigmentation in *C. simulatus*, males are smaller and less full-bodied, as is the case in most members of the Corydoradinae (Pruzsinszky and Ladich, 1998; Kohda *et al.*, 2002). Consequently, female fecundity is increased with size, whereas male fecundity is not (Kohda *et al.*, 2002). Subsequently, female *C. simulatus* may have higher energetic restrictions resulting in an increased need for foraging and time in shelter, making crypsis more advantageous. This is in contrast to the female Batesian *Papilio* systems, where females spend more time out in the open, searching for specific host plants on which to lay their eggs (Maisonneuve, Smadi and Llaurens, 2022). Female Corydoradinae species lay eggs indiscriminately among the leaf-litter and plant growth and therefore may not be as exposed as female *Papilio spp* (Fuller and Evers, 2005; Kunte, 2009; Fuller, 2012). Another feature of the *Papilio* system is that dimorphism is maintained by female choice (Krebs and West, 1988). There is little evidence for female choice in the Corydoradinae subfamily with males being non-aggressive even during mating and male fecundity is largely linked to the number of courtships an individual participates in (Kohda *et al.*, 2002).

Whilst differences in behaviour can drive sex-limited mimicry, the presence of this in a Müllerian mimicry system is puzzling (Joron and Mallet, 1998). Due to the mutualistic interaction between co-mimics, there is no benefit to halving the effective mimetic population size, as in Batesian systems

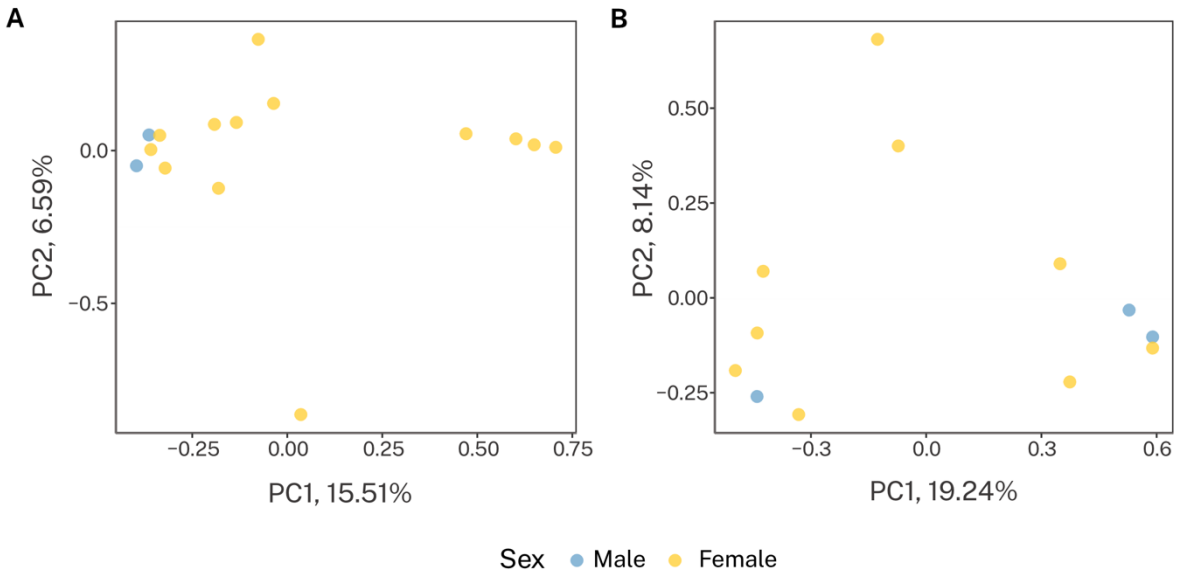
(Joron and Mallet, 1998; Ohsaki, 2005). One hypothesis to explain this is *C. simulatus* may represent a quasi-Batesian system parasitising *H. metae*, with the former being weakly defended and therefore acting as a Batesian mimic corroding the protection of the more defended *H. metae* (Speed, 1999; Rowland *et al.*, 2010). However, previous work presented in this thesis, indicates this to be unlikely due to equal potency of *Corydoras* and *Hoplisoma* venom, as well as little differences in the genetic components of the venom (Chapter 4). An alternative could be that *C. simulatus* males adverged on the *H. metae* colour pattern, perhaps moving from a cryptic phenotype, but this phenotypic change incurred higher costs for females. The evolution of a conspicuous phenotype from a cryptic one, represents a shift from one adaptive peak to another and costs of moving could have been higher for females, perhaps due to behaviours relating to reproductive restraints as discussed above (Merrill *et al.*, 2015). Whilst sex-limited Müllerian mimicry is present in *C. simulatus*, it is puzzling that is not more prevalent across the Corydoradinae. Perhaps there are unique ecological differences in environment of *C. simulatus*, such as higher density of predators compared to other Corydoradinae environments, although this has not been explored. Additionally, it could represent the initial stage of mimicry evolution, whilst other systems are more established. Ultimately, these hypotheses are yet to be explored but could benefit from a better understanding of the timing and nature (e.g. advergence/convergence, and ancestral pattern) of colour pattern change in this system.

Sexually antagonistic selection results in intralocus conflict, which is thought to be partially resolved through sexual dimorphism (Cox and Calsbeek, 2009; Kaufmann, Howie and Immonen, 2023). Resolution to sexual conflict can be in part realised at a genomic level through the differential expression of genes between the sexes or by linkage to a sex determining region (Cox and Calsbeek, 2009). Within the outlier region on Scaffold 9 we identified two genes that represent candidates for sex determination or differentiation in *C. simulatus*. Sex determination initiates the sex differentiation pathway, which activates genes associated with testis or ovary development (Rajendiran *et al.*, 2021; Kitano *et al.*, 2024). At this stage, it is not possible to identify whether our candidate genes determine the type of gonad or solely play a role in the subsequent development. To determine this, we would need to explore expression differences in embryo development (Wen *et al.*, 2023). However, whilst these genes are within the region on Scaffold 9 with elevated genetic differentiation, along with candidates for pigmentation, they are distributed > 1 mb away from the pigment genes. Previous work in a sister species showed that linkage decays around 40 kb (Chapter 3). If the region on Scaffold 9 is implicated with sex determination, recombination may be reduced in this region (Kottler and Schartl, 2018). However, further work is needed to determine the architecture of this region. Alternatively, we did see differences in expression of our pigment candidates between the sexes (Figure 5.4), indicating

that sex-specific expression on the flank of the individual could be contributing to the resolution of intralocus conflict here.

The rarity of sex-limited Müllerian mimicry makes *C. simulatus* a valuable model system, which can be utilised to further improve our understanding of the evolution of protective mimicry. Here, we have identified candidate genes within a 7mb region on Scaffold 9, that we have demonstrated to play a role in the pigmentation of the flank between males and female *C. simulatus.* This trait is one that distinguishes between the cryptic strategy of females and the conspicuous mimetic strategy of males. We hypothesise that evolution of the mimetic phenotype in this system is due to differences in behaviour relating to reproductive restraints in females, but acknowledge that this requires further work, specific to this system. Additionally, whilst we identified two candidate sex differentiation/ determination genes, it is unlikely that sex linkage is underpinning sexual dimorphism here. Alternatively, we have provided evidence to demonstrate sex-specific expression in the flank is at least partially responsible for the resolution of intralocus conflict in this system. Currently, the genetic basis of sex-limited mimicry , and mimicry in general is limited to the Neotropical butterflies (Joron and Mallet, 1998; Ohsaki, 2005; Kunte, 2009; Merrill *et al.*, 2015; Jiggins, 2016). Thus, *C. simulatus* represents an important vertebrate system, contributing to the current understanding of the evolution and maintenance of mimetic phenotypes.

5.6 Supplementary Material



Supplementary Figure 5.1 Principal component analysis of the *Corydoras simulatus* genomic data within each batch, from different aquarium sources. Blue points represent males, yellow points represent females. **A** Batch01, **B** Batch02

Supplementary Table 5.1 Known pigment genes that are differentially expressed between melanated and non-melanated scute (skin) tissue in female *Corydoras simulatus*.

Gene	Protein	p.adjust	Trinity ID
mlph	Melanophilin	0.00	TRINITY_DN1113_c0_g1
bcl2	Apoptosis regulator Bcl-2	0.00	TRINITY_DN11304_c2_g1
sash1	SAM and SH3 domain-containing protein 1	0.00	TRINITY_DN133410_c0_g1
mc1r	Melanocyte-stimulating hormone receptor	0.00	TRINITY_DN14590_c1_g1
rbpj	Suppressor of hairless protein homolog	0.05	TRINITY_DN1486_c0_g1
ccdc28b	Coiled-coil domain-containing protein 28B	0.00	TRINITY_DN1549_c1_g4
tbx15	T-box transcription factor TBX15	0.00	TRINITY_DN171628_c0_g1
gpr143	G-protein coupled receptor 143	0.05	TRINITY_DN1793_c6_g1
atp1a1	Sodium/potassium-transporting ATPase subunit alpha-1	0.00	TRINITY_DN18049_c0_g1
usp13	Ubiquitin carboxyl-terminal hydrolase 13	0.04	TRINITY_DN1861_c0_g1
ctsd	Cathepsin D	0.00	TRINITY_DN18914_c0_g1
frem2	FRAS1-related extracellular matrix protein 2	0.00	TRINITY_DN19763_c0_g2
nnt	NAD(P) transhydrogenase, mitochondrial	0.00	TRINITY_DN2340_c0_g2
tyrp1	5,6-dihydroxyindole-2-carboxylic acid oxidase	0.03	TRINITY_DN2491_c0_g1
tyrp1	5,6-dihydroxyindole-2-carboxylic acid oxidase	0.00	TRINITY_DN2491_c0_g2

kbtbd8	Kelch repeat and BTB domain-containing protein 8	0.04	TRINITY_DN2611_c0_g1
rxra	Retinoic acid receptor RXR-gamma-A	0.00	TRINITY_DN2617_c0_g1
dtnbp1	Dysbindin	0.01	TRINITY_DN3204_c0_g1
vdac1	Voltage-dependent anion-selective channel protein 1	0.00	TRINITY_DN3262_c0_g1
aldoa	Fructose-bisphosphate aldolase A	0.02	TRINITY_DN3355_c0_g1
trpm1	Transient receptor potential cation channel subfamily M member 1	0.03	TRINITY_DN4136_c0_g1
ppargc1a	Peroxisome proliferator-activated receptor gamma coactivator 1-alpha	0.00	TRINITY_DN41369_c0_g1
plxnb2	Plexin-B2	0.01	TRINITY_DN4221_c6_g2
fhl1	Four and a half LIM domains protein 1	0.00	TRINITY_DN4342_c0_g1
gnptab	N-acetylglucosamine-1-phosphotransferase subunits alpha/beta	0.01	TRINITY_DN4563_c4_g1
rad21	Double-strand-break repair protein rad21 homolog A	0.01	TRINITY_DN4883_c1_g1
rad21a	Double-strand-break repair protein rad21 homolog A	0.01	TRINITY_DN4883_c1_g1
actr3b	Actin-related protein 3B	0.00	TRINITY_DN55269_c0_g1
sytl2	Synaptotagmin-like protein 2	0.00	TRINITY_DN578_c0_g1
pah	Phenylalanine-4-hydroxylase	0.00	TRINITY_DN60150_c1_g1
lrmda	Leucine-rich melanocyte differentiation-associated protein	0.00	TRINITY_DN6211_c0_g1
pmel	Melanocyte protein PMEL	0.00	TRINITY_DN6894_c0_g1

5.7 Attribution Statement

Project conception by Emily Phelps and Martin Taylor. All wet laboratory work and analysis was performed by Emily Phelps. Supervision by Martin Taylor.

5.8 References

Alexandrou, M.A., Oliveira, C., Maillard, M., *et al.* (2011) Competition and phylogeny determine community structure in Müllerian co-mimics, *Nature*, 469(7328), pp. 84–88. https://doi.org/10.1038/nature09660.

Andrews, S. (2010) *FastQC: A Quality Control Tool for High Throughput Sequence*. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Bates, H.W. (1862) Contributions to an insect fauna of the Amazon valley. Lepidoptera: heliconinae, *Journal Of The Proceedings Of The Linnean Society Of London Zoology*, 6(22), pp. 73–77. https://doi.org/10.1111/j.1096-3642.1862.tb00932.x.

Baxter, L.L., Watkins-Chow, D.E., Pavan, W.J., *et al.* (2019) A curated gene list for expanding the horizons of pigmentation biology, *Pigment Cell & Melanoma Research*, 32(3), pp. 348–358. https://doi.org/10.1111/pcmr.12743.

Bhattacharya, I. and Modi, D. (2021) Sex Determination in Teleost Fish, *Recent Updates In Molecular Endocrinology And Reproductive Physiology Of Fish: An Imperative Step In Aquaculture*, pp. 121–138. https://doi.org/10.1007/978-981-15-8369-8_9.

Bono, A., Caserta Tencatt, L.F., Alonso, F., *et al.* (2019) Redescription of Corydoras undulatus Regan, 1912 (Siluriformes: Callichthyidae), with comments on the identity of Corydoras latus Pearson, 1924, *PloS One*, 14(1), p. e0211352. https://doi.org/10.1371/journal.pone.0211352.

Bryant, D.M., Johnson, K., DiTommaso, T., *et al.* (2017) A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors, *Cell Reports*, 18(3), pp. 762–776. https://doi.org/10.1016/j.celrep.2016.12.063.

Buscà, R. and Ballotti, R. (2000) Cyclic AMP a key messenger in the regulation of skin pigmentation, *Pigment Cell Research*, 13(2), pp. 60–69. https://doi.org/10.1034/j.1600-0749.2000.130203.x.

Cal, L., Suarez-Bregua, P., Cerdá-Reverter, J.M., *et al.* (2017) Fish pigmentation and the melanocortin system, *Comparative Biochemistry And Physiology. Part A, Molecular & Integrative Physiology*, 211, pp. 26–33. https://doi.org/10.1016/j.cbpa.2017.06.001.

Camacho, C., Coulouris, G., Avagyan, V., *et al.* (2009) BLAST+: architecture and applications, *BMC Bioinformatics*, 10. https://doi.org/10.1186/1471-2105-10-421.

Caro, T. and Ruxton, G. (2019) Aposematism: Unpacking the defences, *Trends In Ecology & Evolution*, 34(7), pp. 595–604. https://doi.org/10.1016/j.tree.2019.02.015.

Chapman, T., Arnqvist, G., Bangham, J., *et al.* (2003) Sexual conflict, *Trends In Ecology & Evolution*, 18(1), pp. 41–47. https://doi.org/10.1016/s0169-5347(02)00004-6.

Cox, R.M. and Calsbeek, R. (2009) Sexually antagonistic selection, sexual dimorphism, and the resolution of intralocus sexual conflict, *The American Naturalist*, 173(2), pp. 176–187. https://doi.org/10.1086/595841.

Danecek, P., Bonfield, J.K., Liddle, J., *et al.* (2021) Twelve years of SAMtools and BCFtools, *GigaScience*, 10(2). https://doi.org/10.1093/gigascience/giab008.

Darolti, I., Wright, A.E., Sandkam, B.A., *et al.* (2019) Extreme heterogeneity in sex chromosome differentiation and dosage compensation in livebearers, *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 116(38), pp. 19031–19036. https://doi.org/10.1073/pnas.1905298116.

Darwin, C. (1859) *On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life*. London : John Murray, 1859.

Darwin, C. (1871) *The descent of man, and selection in relation to sex*. Princeton, NJ: Princeton University Press.

Dias, A.C., Tencatt, L.F.C., Roxo, F.F., *et al.* (2024) Phylogenomic analyses in the complex Neotropical subfamily Corydoradinae (Siluriformes: Callichthyidae) with a new classification based on morphological and molecular data, *Zoological Journal Of The Linnean Society* .

Endler, J.A. (1978) A Predator's View of Animal Color Patterns, in M.K. Hecht, W.C. Steere, and B. Wallace (eds) *Evolutionary Biology*. Boston, MA: Springer US, pp. 319–364. https://doi.org/10.1007/978-1-4615-6956-5_5.

Endler, J.A. (1983) Natural and sexual selection on color patterns in poeciliid fishes, *Environmental Biology Of Fishes*, 9(2), pp. 173–190. https://doi.org/10.1007/BF00690861.

Eyer, P.-A., Blumenfeld, A.J. and Vargo, E.L. (2019) Sexually antagonistic selection promotes genetic divergence between males and females in an ant, *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 116(48), pp. 24157–24163. https://doi.org/10.1073/pnas.1906568116.

Font, E. (2019) Mimicry, Camouflage and Perceptual Exploitation: the Evolution of Deception in Nature, *Biosemiotics*, 12(1), pp. 7–24. https://doi.org/10.1007/s12304-018-9339-6.

Fuller, I. (2012) *Breeding Corydoradinae Catfish*. 2nd edn. Kidderminster, England: Ian Fuller Enterprises.

Fuller, I.A.M. and Evers, H.E. (2005) *Identifying Corydoradinae Catfishe*. Kidderminster, England: Ian Fuller Enterprises.

Göran Arnqvist and Locke Rowe (2005) *Sexual Conflict*. Princeton, N.J.: Princeton University Press.

Grabherr, M.G., Haas, B.J., Yassour, M., *et al.* (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome, *Nature Biotechnology*, 29(7), pp. 644–652. https://doi.org/10.1038/nbt.1883.

Greven, H., Flasbecl, T. and Passia, D. (2006) Axillary glands in the armoured catfish Corydoras aeneus (Callichthyidae, Siluriformes), *Verhandlungen Der Gesellschaft Für Ichthyologie*, 5, pp. 65–69.

Jiggins, C.D. (2016) Genes on the wing: colour pattern genetics, in *The Ecology And Evolution Of Heliconius Butterflies*. Oxford University Press, pp. 112–137. https://doi.org/10.1093/acprof:oso/9780199566570.003.0008.

Joron, M. and Mallet, J.L. (1998) Diversity in mimicry: paradox or paradigm?, *Trends In Ecology & Evolution*, 13(11), pp. 461–466. https://doi.org/10.1016/s0169-5347(98)01483-9.

Jun, G., Wing, M.K., Abecasis, G.R., *et al.* (2015) An efficient and scalable analysis framework for variant extraction and refinement from population-scale DNA sequence data, *Genome Research*, 25(6), pp. 918–925. https://doi.org/10.1101/gr.176552.114.

Kaufmann, P., Howie, J.M. and Immonen, E. (2023) Sexually antagonistic selection maintains genetic variance when sexual dimorphism evolves, *The Royal Society Proceedings Biological Sciences* 290(1995). https://doi.org/10.1098/rspb.2022.2484.

Kitano, J., Ansai, S., Takehana, Y., *et al.* (2024) Diversity and convergence of sex-determination mechanisms in teleost Fish, *Annual Review Of Animal Biosciences*, 12, pp. 233–259. https://doi.org/10.1146/annurev-animal-021122-113935.

Kohda, M., Yonebayashi, K., Nakamura, M., *et al.* (2002) Male reproductive success in a promiscuous armoured catfish Corydoras Aeneus (Callichthyidae), *Environmental Biology Of Fishes*, 63(3), pp. 281–287. https://doi.org/10.1023/A:1014317009892.

Korneliussen, T.S., Albrechtsen, A. and Nielsen, R. (2014) ANGSD: Analysis of Next Generation Sequencing Data, *BMC Bioinformatics*, 15(356). https://doi.org/10.1186/s12859-014-0356-4.

Korunes, K.L. and Samuk, K. (2021) pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data, *Molecular Ecology Resources*, 21(4), pp. 1359–1368. https://doi.org/10.1111/1755-0998.13326.

Kottler, V.A. and Schartl, M. (2018) The colorful sex chromosomes of teleost fish, *Genes*, 9(5). https://doi.org/10.3390/genes9050233.

Krebs, R.A. and West, D.A. (1988) Female mate preference and the evolution of female-limited batesian mimicry, *Evolution*, 42(5), pp. 1101–1104. https://doi.org/10.2307/2408927.

Kunte, K. (2009) The diversity and evolution of batesian mimicry in Papilio swallowtail butterflies, *Evolution*, 63(10), pp. 2707–2716. https://doi.org/10.1111/j.1558-5646.2009.00752.x.

Lawrence, M., Huber, W., Pagès, H., *et al.* (2013) Software for computing and annotating genomic ranges, *PLoS Computational Biology*, 9(8). https://doi.org/10.1371/journal.pcbi.1003118.

Li, H. and Durbin, R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform, *Bioinformatics*, 25(14), pp. 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Lorin, T., Brunet, F.G., Laudet, V., *et al.* (2018) 1 Teleost fish-specific preferential retention of pigmentation gene-containing families after 2 whole genome duplications in vertebrates, *G3* , 8(5), pp. 1795–1806. https://doi.org/10.1534/g3.118.200201.

Lou, R.N., Jacobs, A., Wilder, A.P., *et al.* (2021) A beginner's guide to low-coverage whole genome sequencing for population genomics, *Molecular Ecology*, 30(23), pp. 5966–5993. https://doi.org/10.1111/mec.16077.

Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, *Genome Biology*, 15(12). https://doi.org/10.1186/s13059-014-0550-8.

Low, X.H. and Monteiro, A. (2018) Dorsal forewing white spots of male Papilio polytes(Lepidoptera: Papilionidae) not Maintained by Female Mate Choice, *Journal Of Insect Behavior*, 31(1), pp. 29–41. https://doi.org/10.1007/s10905-017-9656-7.

Lowe, A., Summers, A.P., Walter, R.P., *et al.* (2021) Scale performance and composition in a small Amazonian armored catfish, Corydoras trilineatus, *Acta Biomaterialia*, 121, pp. 359–370. https://doi.org/10.1016/j.actbio.2020.11.045.

Maisonneuve, L., Smadi, C. and Llaurens, V. (2022) Evolutionary origins of sexual dimorphism: Lessons from female-limited mimicry in butterflies, *Evolution; International Journal Of Organic Evolution*, 76(10), pp. 2404–2423. https://doi.org/10.1111/evo.14599.

Mank, J.E. (2017) Population genetics of sexual conflict in the genomic era, *Nature Reviews. Genetics*, 18(12), pp. 721–730. https://doi.org/10.1038/nrg.2017.83.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads, *EMBnet.Journal*, 17(1), pp. 10–12. https://doi.org/10.14806/ej.17.1.200.

Meisner, J. and Albrechtsen, A. (2018) Inferring population structure and admixture proportions in low-depth NGS Data, *Genetics*, 210(2), pp. 719–731. https://doi.org/10.1534/genetics.118.301336.

Mérot, C., Berdan, E.L., Cayuela, H., *et al.* (2021) Locally Adaptive inversions modulate genetic variation at different geographic scales in a seaweed fly, *Molecular Biology And Evolution*, 38(9), pp. 3953–3971. https://doi.org/10.1093/molbev/msab143.

Merrill, R.M., Dasmahapatra, K.K., Davey, J.W., *et al.* (2015) The diversification of Heliconius butterflies: what have we learned in 150 years?, *Journal Of Evolutionary Biology*, 28(8), pp. 1417–1438. https://doi.org/10.1111/jeb.12672.

Mistry, J., Chuguransky, S., Williams, L., *et al.* (2021) Pfam: The protein families database in 2021, *Nucleic Acids Research*, 49(1), pp. 412–419. https://doi.org/10.1093/nar/gkaa913.

Müller, F. (1878) Ueber Die Vortheile Der Mimicry Bei Schmetterlingen, *Zool. Anzeiger.*

Nishida, R. (2017) Chemical ecology of poisonous butterflies: Model or mimic? A paradox of sexual dimorphisms in müllerian mimicry, in *Diversity And Evolution Of Butterfly Wing Patterns*. Singapore: Springer Singapore, pp. 205–220. https://doi.org/10.1007/978-981-10-4956-9_11.

Ohsaki, N. (2005) A common mechanism explaining the evolution of female-limited and both-sex Batesian mimicry in butterflies, *The Journal Of Animal Ecology*, 74(4), pp. 728–734. https://doi.org/10.1111/j.1365-2656.2005.00972.x.

Palmer, D.H., Rogers, T.F., Dean, R., *et al.* (2019) How to identify sex chromosomes and their turnover, *Molecular Ecology*, 28(21), pp. 4709–4724. https://doi.org/10.1111/mec.15245.

Parichy, D.M. (2021) Evolution of pigment cells and patterns: recent insights from teleost fishes, *Current Opinion In Genetics & Development*, 69, pp. 88–96. https://doi.org/10.1016/j.gde.2021.02.006.

Patro, R., Duggal, G., Love, M.I., *et al.* (2017) Salmon provides fast and bias-aware quantification of transcript expression, *Nature Methods*, 14(4), pp. 417–419. https://doi.org/10.1038/nmeth.4197.

Petrie, M. (2021) Evolution by sexual selection, *Frontiers In Ecology And Evolution*, 9, p.786868. https://doi.org/10.3389/fevo.2021.786868.

Pruzsinszky, I. and Ladich, F. (1998) Sound production and reproductive behaviour of the armoured catfish Corydoras paleatus (Callichthyidae), *Environmental Biology Of Fishes*, 53(2), pp. 183–191. https://doi.org/10.1023/a:1007413108550.

Purcell, S., Neale, B., Todd-Brown, K., *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses, *American Journal Of Human Genetics*, 81(3), pp. 559–575. https://doi.org/10.1086/519795.

R Core Team (2023) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Rajendiran, P., Jaafar, F., Kar, S., *et al.* (2021) Sex determination and differentiation in teleost: Roles of genetics, environment, and brain, *Biology*, 10(10). https://doi.org/10.3390/biology10100973.

Reimchen, T.E. and Nosil, P. (2004) Variable predation regimes predict the evolution of sexual dimorphism in a population of threespine stickleback, *Evolution; International Journal Of Organic Evolution*, 58(6), pp. 1274–1281. https://doi.org/10.1111/j.0014-3820.2004.tb01706.x.

Roberts, R.B., Ser, J.R. and Kocher, T.D. (2009) Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes, *Science*, 326(5955), pp. 998–1001.https://doi.org/10.1126/science.1174705.

Rowe, L., Chenoweth, S.F. and Agrawal, A.F. (2018) The genomics of sexual conflict, *The American Naturalist*, 192(2), pp. 274–286. https://doi.org/10.1086/698198.

Rowland, H.M., Mappes, J., Ruxton, G.D., *et al.* (2010) Mimicry between unequally defended prey can be parasitic: evidence for quasi-Batesian mimicry, *Ecology Letters*, 13(12), pp. 1494–1502. https://doi.org/10.1111/j.1461-0248.2010.01539.x.

Ruxton, G.D., Allen, W.L., Sherratt, T.N., *et al.* (2018) Avoiding Attack: The evolutionary ecology of crypsis, aposematism, and mimicry, in *Avoiding Attack*. 2nd edn. Oxford University Press. https://doi.org/10.1093/oso/9780199688678.001.0001.

Ruzicka, F., Dutoit, L., Czuppon, P., *et al.* (2020) The search for sexually antagonistic genes: Practical insights from studies of local adaptation and statistical genomics, *Evolution Letters*, 4(5), pp. 398–415. https://doi.org/10.1002/evl3.192.

Sanders, K.L., Malhotra, A. and Thorpe, R.S. (2006) Evidence for a Müllerian mimetic radiation in Asian pitvipers, *The Royal Society Proceedings, Biological Sciences*, 273(1590), pp. 1135–1141. https://doi.org/10.1098/rspb.2005.3418.

Sands, D. (1994) *The behaviour and evolutionary ecology of corydoras adolfoi and corydoras imitator: Studies on two species of sympatric catfish from the upper rio negro, brazil*. PhD. Liverpool University.

Soneson, C., Love, M.I. and Robinson, M.D. (2015) Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences, *F1000Research*, 4, p. 1521. https://doi.org/10.12688/f1000research.7563.2.

Speed, M.P. (1999) Batesian, quasi-Batesian or Müllerian mimicry? Theory and data in mimicry Research, *Evolutionary Ecology*, 13(7), pp. 755–776. https://doi.org/10.1023/A:1010871106763.

Svitačová, K., Horký, P., Valchářová, T., *et al.* (2024) Pigment matters: Behavior and lateralization of albino and pigmented fish (Bronze Corydoras) in aquaculture, *Applied Animal Behaviour Science*, 272. https://doi.org/10.1016/j.applanim.2024.106205.

Todd, E.V., Liu, H., Lamm, M.S., *et al.* (2018) Female mimicry by sneaker males has a transcriptomic signature in both the brain and the gonad in a sex-changing fish, *Molecular Biology And Evolution*, 35(1), pp. 225–241. https://doi.org/10.1093/molbev/msx293.

Tosto, N.M., Beasley, E.R., Wong, B.B.M., *et al.* (2023) The roles of sexual selection and sexual conflict in shaping patterns of genome and transcriptome variation, *Nature Ecology & Evolution*, 7(7), pp. 981–993. https://doi.org/10.1038/s41559-023-02019-7.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., *et al.* (2013) From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline, *Current Protocols In Bioinformatics*, 43(1). https://doi.org/10.1002/0471250953.bi1110s43.

Weitzman, S.H. and Nijssen, H. (1970) Four new species and one new subspecies of the catfish genus Corydoras from Ecuador, Colombia and Brazil (Pisces, Siluriformes, Callichthyidae), *Beaufortia*, 18(233), pp. 119–132. https://repository.naturalis.nl/pub/504794/BEAU1970018233001.pdf.

Wen, M., Pan, Q., Larson, W., *et al.* (2023) Characterization of the sex determining region of channel catfish (Ictalurus punctatus) and development of a sex-genotyping test, *Gene*, 850, p. 146933. https://doi.org/10.1016/j.gene.2022.146933.

6 General Discussion

6.1 Synthesis

Protective mimicries have captured the interest of evolutionary biologists from the very conception of the field (Darwin, 1859; Endler, 1978; Caro, 2017). Despite the longevity of study, many questions regarding the evolution and maintenance of these systems remain (Merrill *et al.*, 2015). The use of molecular evidence in the Lepidopteran *Heliconius* and *Papilio* have demonstrated the importance of genetic conservation, introgression and complex genetic architectures in the evolution of protective mimicries (Pardo-Diaz *et al.*, 2012; Merrill *et al.*, 2015; Zhang *et al.*, 2016; Iijima, Yoda and Fujiwara, 2019; Komata, Lin and Fujiwara, 2022). However, it is not known how universal these inferences are, particularly in vertebrate systems. A major roadblock in studying the mimetic genotypes of vertebrates has been the ability to synthesise genetic resources due to the large genome sizes of mimetic vertebrates (Stuckert *et al.*, 2024). Improvements in sequencing technologies and bioinformatic techniques are making the study of the genetics underpinning protective mimicry more accessible than ever before (Pardo-Diaz, Salazar and Jiggins, 2015; Hjelmen, 2024). Recently, the poison arrow frog *Ranitomeya imitator* genome was sequenced, marking an important milestone in the study of mimicry (Stuckert *et al.*, 2024). The overarching aim of this thesis was to generate novel genetic resources for the Corydoradinae and utilise these resources to elucidate the evolution of protective mimicry in this system, adding to the current literature. As such, this thesis marks not only an important contribution to the study of mimetic systems, but a very timely addition.

6.1.1 Genome Evolution in the *Corydoradinae*

Participation in Müllerian mimicry rings is not ubiquitous across all Corydoradinae genera, with those from *Hoplisoma*, *Corydoras* and *Brochis* being overrepresented within mimetic communities (Alexandrou *et al.*, 2011). Additionally, *Corydoras* and *Brochis* are only co-mimics when *Hoplisoma sp* are present in a community (Alexandrou *et al.*, 2011). This suggests members of *Hoplisoma* possess biological characteristics that increase the likelihood of mimetic communities evolving when they are present. One such characteristic is the increased genome size in the group (Marburger *et al.*, 2018). This has been hypothesised to be driven by the presence of a whole genome duplication (WGD) event and transposable element (TE) proliferation (Marburger *et al.*, 2018). In this first data chapter, we focused on confirming a role for both WGD and TE expansion in the evolutionary history of *Hoplisoma*. We achieved this by generating novel genomic resources, including a scaffold level assembly for *Corydoras fulleri,* as well as, generating genome annotations for both *C. fulleri* and *H. metae.* Evidence for a WGD in the *Hoplisoma* was demonstrated via various whole genome techniques. These revealed

that not only was a WGD present in *Hoplisoma* but provided evidence to suggest it is a recent autotetraploid, originating only 4.86 mya. This is the youngest autopolyploid event to have been identified in vertebrates. Known teleost autopolyploidy events include the Salmonid specific duplication, as well as the duplication in the paddlefish/sturgeon ancestor (Robertson *et al.*, 2017; Gundappa *et al.*, 2022; Redmond *et al.*, 2023). Thus, the event in *Hoplisoma* could elucidate the evolutionary mechanisms shaping recent autopolyploid genomes in a way not previously available. Additionally, we found evidence to suggest TE expansion is also driving genome size shifts in *H. metae,* relative to *C. fulleri*. We demonstrated that TEs had a higher probability of insertion within introns in *H. metae*, which was driving the increase in gene length despite relatively similar exon lengths. Insertions were also more likely to occur within the first intron of *H. metae* than subsequent introns. The role of the first intron in gene expression, as well as evidence for these insertions resulting in phenotypic changes in pigmentation indicates that perhaps the TE expansion could also be playing a role in diversity of mimetic phenotypes (Van't Hof *et al.*, 2016; Kratochwil *et al.*, 2022).

6.1.1.1 Genome annotation quality

In this work, we found limited evidence for polyploidy within specific gene families. A potential reason for this was the fragmented and collapsed nature of the *H. metae* genome assembly (Chapter 2). Being a recent autopolyploid and having a high repeat content, makes *H. metae* a challenging genome to assemble. This likely has knock on effects on the genome annotation, which is also highly fragmented (Yandell and Ence, 2012; Salzberg, 2019). Additionally, the long intron lengths of the *H. metae* genome may have impacted alignment of expression evidence as well as *ab initio* prediction (Campbell *et al.*, 2014). If only genes with shorter introns are initially identified, these will represent the *ab initio* training set, which will then have a downstream impact on the next iteration of the genome annotation (Card *et al.*, 2019). Subsequently, inherent genomic characteristics of *H. metae* may be responsible for the low BUSCO score, indicating incomplete annotation (Manni *et al.*, 2021). Now having a greater understanding of these features will aid improvement of the genome annotation, even in the absence of a more contiguous assembly.

6.1.2 The genetics of polymorphic mimicry in the Neotropical Catfish, *Corydoras fulleri*.

Despite its prevalence in nature, the diversity of warning colouration in Müllerian mimicry is contradictory, remaining one of the unanswered questions in the field (Merrill *et al.*, 2015). Novel phenotypes should be preferentially predated, due to positive frequency dependent selection on rare morphs (Joron and Mallet, 1998). In this chapter, we introduced the polymorphic mimetic species *C. fulleri* as a novel model polymorphic mimicry system. Utilising windowed F_{ST} across the genome, we

were able to identify genes related to the presence or absence of a blotch pattern on an individual's flank. Ultimately, we largely identified genes that played a known role in melanin density and melanosome distribution. Indicating that mimicry in *C. fulleri* is utilising known pigment pathways to participate in mimetic communities. Additionally, these genes provided an initial insight into the genomic architecture of the trait as well as the nature of the polymorphism in this system, including whether the trait is likely to be a local polymorphism, a hybrid zone between geographically monomorphic populations or deviation from the ancestral state.

6.1.2.1 Background noise and sample size

The regions identified in this study were distributed across the genome. Additionally, many other regions demonstrated elevated genetic diversity, but on inspection did not separate by the phenotypes of interest. Additionally, we observed high background levels of F_{ST}. The variation between morphs (requiring only extremes to be used) and population structuring exacerbated issues with an already small population size. In simulations, when comparing populations connected by geneflow background noise can be high, even at coverage up to 8x, when sample sizes are low (Lou *et al.*, 2021). This may be responsible for the elevated background noise across the genome in the *C. fulleri* study. To account for false positives, we only retained regions where PCA of SNPs in the region separated by phenotype. Additionally, there is evidence to suggest that small sample size does not limit the inferences made from F_{ST} when the number of markers is high (Willing, Dreyer and van Oosterhout, 2012). Therefore, whilst we have confidence in the results presented here, ultimately this work would greatly benefit from additional sampling, as many studies on wild populations often do.

6.1.3 Quantifying venom strength and venom related gene expression to better understand the evolution of mimicry in the Corydoradinae.

Unpalatability and the advertisement of unpalatability via conspicuous warning colouration are the key tenants of protective mimicries. Relative to work on the evolution of warning colour, unpalatability is often overlooked but has fundamental implications on the interactions between taxa in mimetic communities. In this chapter, we investigated venom potency of *Corydoras sp* and *Hoplisoma sp*, to better understand these interactions. Envenomation of researchers and hobbyists suggested that members of *Hoplisom*a are more painful than other Corydoradinae genera (Alexandrou *et al.*, 2011). Using a brine shrimp cytotoxicity assay, we demonstrated that whilst venom gland extracts were more toxic to brine shrimp than muscle tissue, *Hoplisoma* did not have more potent venom than *Corydoras*. We also explored the genetic basis for venom within the group, identifying candidate genes with domains similar to those of known venoms, as well as venom housekeeping genes, using Hidden Markov Models, differential expression and co-expression

networks. The genes identified in this study contain C-type lectin domains, similar to those found in the Scorpion fish (Nakagawa *et al.*, 2015; Campos *et al.*, 2016; Ziegman *et al.*, 2019). This may indicate there are some biological qualities of these proteins that make them amenable to piscine venoms. For example, perhaps they are more stable in aquatic environments or perhaps it is parallel evolution with selection occurring on similar ichthyocrinotoxins (Harris and Jenner, 2019).

6.1.3.1 Polyploid transcriptomes and cross-species expression

This study utilised six species, across two genera, which are known to represent mimetic pairs. The use of cross-species comparisons is technically challenging. Standard practice utilises inferences made from orthologs, as we did in this chapter (Chung *et al.*, 2021). However, difficulty in identifying orthologs or many-to-one orthologous groups may prevent some of the more nuanced gene-specific differences from being detected (Chung *et al.*, 2021). This is a more pertinent issue when using a recent polyploid, as suspected in *Hoplisoma*, due to mechanisms such as fractionation (gene loss or retention) and dosage which can impact interpretation of evolutionary relationships between orthologs (Chen *et al.*, 2017; Cheng *et al.*, 2018). Similarly, the assembly of *de novo* polyploid transcriptomes is subject to many of the same limitations as polyploid genome assembly (Kyriakidou *et al.*, 2018). Gene duplication and the associated increase in heterozygosity can complicate the building of de Bruijn graphs used in transcriptome assembly, resulting in increases in collapse of transcripts, chimerism and additional redundancy (Madritsch, Burg and Sehr, 2021; Oh, Lee and Park, 2021). Similarly to genome assembly, long read sequencing may reduce these artefacts due to its ability to sequence whole transcripts (Kyriakidou *et al.*, 2018). However, this is not always accessible to those studying non-model organisms.

6.1.3.2 A framework for quasi-Batesian mimicry

We did not find evidence for differences in venom potency in the Corydoradinae, suggesting that they are traditional Müllerian mimics rather than quasi-Batesian mimics. However, other studies have found evidence for differences in unpalatability between co-mimics (Stuckert *et al.*, 2014; Winters *et al.*, 2018; Mattila *et al.*, 2021; Mattila, Jiggins and Saastamoinen, 2022). The poison dart frog mimic *Ranitomeya imitator* was found to have a higher alkaloid diversity than its co-mimics at different sites (Stuckert *et al.*, 2014). In this system, quasi-Batesian mimicry is unlikely as *R. imitator* has been shown to adverge on local models (Symula, Schulte and Summers, 2001; Muell *et al.*, 2022). However, mimetic nudibranchs found along the east coast of Australia were shown to be differentially chemically defended (Winters *et al.*, 2018). This was not considered to be quasi-Batesian mimicry as those which were less chemically defended (putative mimics) clustered in appearance with those that were the most chemically defended (putative models), indicating a lack of imperfect mimicry (Winters

et al., 2018). In Batesian mimicry, an imperfect resemblance is thought to reflect the arms race between the model and mimic (Ruxton *et al.*, 2018). As such the accuracy of mimicry in the nudibranchs was interpreted as not reflective of Batesian mimicry (Joron and Mallet, 1998; Winters *et al.*, 2018). Whether this is true for quasi-Batesian mimics, who are partially defended, remains to be seen, but highlights the need for a framework in which differences between co-mimics can be interpreted to test for quasi-Batesian mimicry. The first step in the framework would be to explore differences in toxicity between co-mimics, as we have done in *Hoplisoma* and *Corydoras*. On identifying differences, a next step to confirming or disregarding quasi-Batesian mimicry would include measures of predator driven selection (Speed, 1999; Rowland *et al.*, 2010; Chouteau and Angers, 2011).

6.1.4 The molecular underpinnings of sex-limited Müllerian mimicry in *Corydoras simulatus.*

Intraspecific variation between the sexes can be caused by sexually antagonistic selection, resulting in a different phenotypic optimum between the sexes, leading to the evolution of sexual dimorphism (Chapman *et al.*, 2003; Rowe, Chenoweth and Agrawal, 2018; Ruzicka *et al.*, 2020). In *Corydoras simulatus*, different anti-predation strategies have been adopted resulting in male-limited Müllerian mimicry. In this chapter, we identified candidate genes associated with intersex genetic differentiation. Additionally, these genes demonstrated differences in gene expression both between sexes and within females, between regions with different melanocyte density. This evidence suggests these genes play a role in pigmentation in *C. simulatus* and subsequently the adoption of contrasting predation strategies. Whilst pigment genes were found to be differentially expressed or represented in the group, they were not present in both genomic and expression data. Whilst we identified genes with known roles in sex differentiation within the same region of elevated genetic differentiation as our candidate pigment genes, the distance between the pigment and sex differentiation candidates was large (> 1mb). Whilst we are unsure of the rate of recombination in *C. simulatus* within this region of the genome, work in *C. fulleri* showed linkage decay around 40 kb. However, this could be larger for regions with reduced recombination such as sex determining regions. Ultimately, we cannot determine a role for sex-linkage from the results in this study. However, we do see some sex-specific expression, indicating that this does play a role in the resolution of sexual conflict *C. simulatus* (Rice, 1984; Ruzicka *et al.*, 2020). Additionally, whilst indicated by phenotypic diversification and elevated genetic differentiation, sexually antagonistic selection in the group has not been specifically explored, representing an avenue for future work (Ruzicka *et al.*, 2020).

In the *C. fulleri* study, we demonstrated pigment genes to play a role in blotch formation. The female *C. simulatus* also has a blotch on the flank, although this is lower contrast to the rest of the body and is more widely spread (Fuller and Evers, 2005). We identified likely candidates controlling this trait in *C. simulatus* that are not currently known to be pigment genes, such as *arpp21*. Whilst this gene is not a known pigment gene, it does play a role in cAMP mediated signalling, which is related to the melanin pathway (Parichy, 2021). Additionally, a large region on Scaffold 9 contained most of our candidate genes in the *C. simulatus* study. This is in contrast to *C. fulleri*, where genes were distributed across the genome. The differences between the architecture of the two traits indicates the genetics of mimicry is complex in the Corydoradinae, with different regions being utilised between even closely related species.

6.2 Directions and Final Remarks

While this thesis has made significant contributions to the understanding of mimicry in the Corydoradinae there remain many promising avenues for further research. In this final section, two such avenues are briefly highlighted, chosen as they build upon this existing work to further elucidate the evolution of mimicry in this subfamily.

6.2.1 Contribution of gene duplication and TE insertion to aposematic phenotypic diversification.

Whilst *Hoplisoma* members are overrepresented in mimetic communities, the underlying reasons remain unknown. Advergence of one defended taxon to the colour pattern of the other is the primary mode through which Müllerian mimetic resemblances occur (Symula, Schulte and Summers, 2001; Marek and Bond, 2009; Wilson *et al.*, 2012; Ruxton *et al.*, 2018; Muell *et al.*, 2022; Cui *et al.*, 2024). The overrepresentation of *Hoplisoma* indicates other Corydoradinae taxa are adverging on the *Hoplisoma* colour pattern. However, this could be formally tested using a time calibrated phylogeny to indicate when colour divergence and speciation is occurring in each co-mimic, as in Muell *et al.* (2022). Both WGD and TE insertion have been linked to colour diversification in teleosts, with pigment genes being preferentially retained after whole genome duplication events (Braasch, Volff and Schartl, 2008; Santos *et al.*, 2014; Lorin *et al.*, 2018; Kratochwil *et al.*, 2022). Subsequently, a comparative analysis investigating gene loss and retention, selection and positional TE insertion between members of *Hoplisoma* and other Corydoradinae genera would be an important contribution to the evolution of colour pattern across the subfamily.

6.2.2 Adaptive introgression as a mechanism to facilitate mimicry ring participation.

The introgression of adaptive alleles across species boundaries seems to a widespread in *Heliconius*, allowing a close resemblance to evolve between two species. For example, *H. cyndo, H. timareta* and *H. heurippa* have gained the regulatory alleles upstream of the *optix* gene from *H. melpomene*, with whom they form a mimetic community (Pardo-Diaz *et al.*, 2012). Additionally, there is evidence for introgression in the evolution of the south variable pitohui *Pitohui uropygialis* from its co-mimic *Pitohui dichrous* (Garg *et al.*, 2019). In this system, SNPs within introgressed regions are linked to genes related to melanin synthesis (Garg *et al.*, 2019). A study of a role for introgression and sharing of adaptive alleles across Corydoradinae co-mimics would be a valuable addition to the literature, indicating whether introgression is common across taxonomically diverse mimetic systems. Initial studies could focus on the pigment candidate genes presented in this thesis, creating gene trees and utilising ABBA-BABA statistics (Zhang *et al.*, 2016; Garg *et al.*, 2019). Alternatively, in additional mimetic Corydoradinae systems, a genome wide phylogenetic tree approach could highlight introgressed regions of the genome, as achieved in the *Oreochromis niloticus* (Etherington *et al.*, 2022). However, introgression between sympatric species from different genera is unlikely due to different genomic structure, including chromosome number. In aquaria, hybridisation among *Hoplisoma* species appear to be relatively common, but there are no confirmed reports of cross-genera hybrids.

6.3 References

Alexandrou, M. A., Oliveira, C., Maillard, M., *et al.* (2011) Competition and phylogeny determine community structure in Mullerian co-mimics, *Nature*, 469(7328), pp. 84–88. https://doi.org/10.1038/nature09660.

Braasch, I., Volff, J.N. and Schartl, M. (2008) The evolution of teleost pigmentation and the fish-specific genome duplication, *Journal Of Fish Biology*, 73(8), pp. 1891–1918. https://doi.org/10.1111/j.1095-8649.2008.02011.x.

Campbell, M.S., Holt, C., Moore, B., *et al.* (2014) Genome annotation and curation using MAKER and MAKER-P, *Current Protocols In Bioinformatics*, 48(1), p. 4.11.1-39. https://doi.org/10.1002/0471250953.bi0411s48.

Campos, F.V., Menezes, T.N., Malacarne, P.F., *et al.* (2016) A review on the Scorpaena plumieri fish venom and its bioactive compounds, *The Journal Of Venomous Animals And Toxins Including Tropical Diseases*, 22(1). https://doi.org/10.1186/s40409-016-0090-7.

Card, D.C., Adams, R.H., Schield, D.R., *et al.* (2019) Genomic Basis of Convergent Island Phenotypes in Boa Constrictors, *Genome Biology And Evolution*, 11(11), pp. 3123–3143. https://doi.org/10.1093/gbe/evz226.

Caro, T. (2017) Wallace on coloration: Contemporary perspective and unresolved insights, *Trends In Ecology & Evolution*, 32(1), pp. 23–30. https://doi.org/10.1016/j.tree.2016.10.003.

Chapman, T., Arnqvist, G., Bangham, J., *et al.* (2003) Sexual conflict, *Trends In Ecology & Evolution*, 18(1), pp. 41–47. https://doi.org/10.1016/s0169-5347(02)00004-6.

Chen, E.C.H., Morin, A., Chauchat, J.-H., *et al.* (2017) Statistical analysis of fractionation resistance by functional category and expression, *BMC Genomics*, 18(366). https://doi.org/10.1186/s12864-017-3736-0.

Cheng, F., Wu, J., Cai, X., *et al.* (2018) Gene retention, fractionation and subgenome differences in polyploid plants, *Nature Plants*, 4(5), pp. 258–268. https://doi.org/10.1038/s41477-018-0136-7.

Chouteau, M. and Angers, B. (2011) The role of predators in maintaining the geographic organization of aposematic signals, *The American Naturalist*, 178(6), pp. 810–817. https://doi.org/10.1086/662667.

Chung, M., Bruno, V.M., Rasko, D.A., *et al.* (2021) Best practices on the differential expression analysis of multi-species RNA-seq, *Genome Biology*, 22(1). https://doi.org/10.1186/s13059-021-02337-8.

Cui, J., Chen, Y., Hines, H.M., *et al.* (2024) Does coevolution in refugia drive mimicry in bumble bees? Insights from a South Asian mimicry group, *Science Advances*, 10(24). https://doi.org/10.1126/sciadv.adl2286.

Darwin, C. (1859) *On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life*. London : John Murray, 1859.

Endler, J.A. (1978) A Predator's View of Animal Color Patterns, in M.K. Hecht, W.C. Steere, and B. Wallace (eds.) *Evolutionary Biology*. Boston, MA: Springer US, pp. 319–364.

Etherington, G.J., Nash, W., Ciezarek, A., *et al.* (2022) Chromosome-level genome sequence of the Genetically Improved Farmed Tilapia (GIFT, Oreochromis niloticus) highlights regions of introgression with O. mossambicus, *BMC Genomics*, 23(1), p. 832. https://doi.org/10.1186/s12864-022-09065-8.

Fuller, I.A.M. and Evers, H.E. (2005) *Identifying Corydoradinae Catfishe*. Kidderminster, England: Ian Fuller Enterprises.

Garg, K.M., Sam, K., Chattopadhyay, B., *et al.* (2019) Gene flow in the müllerian mimicry ring of a poisonous Papuan songbird clade (Pitohui; Aves), *Genome Biology And Evolution*, 11(8), pp. 2332–2343. https://doi.org/10.1093/gbe/evz168.

Gundappa, M.K., To, T.-H., Grønvold, L., *et al.* (2022) Genome-wide reconstruction of rediploidization following autopolyploidization across one hundred million years of Salmonid evolution, *Molecular Biology And Evolution*, 39(1). https://doi.org/10.1093/molbev/msab310.

Harris, R.J. and Jenner, R.A. (2019) Evolutionary ecology of fish venom: Adaptations and consequences of evolving a venom system, *Toxins*, 11(2). https://doi.org/10.3390/toxins11020060.

Hjelmen, C.E. (2024) Genome size and chromosome number are critical metrics for accurate genome assembly assessment in Eukaryota, *Genetics*, 227(4). https://doi.org/10.1093/genetics/iyae099.

Iijima, T., Yoda, S. and Fujiwara, H. (2019) The mimetic wing pattern of Papilio polytes butterflies is regulated by a doublesex-orchestrated gene network, *Communications Biology*, 2(1), p. 257. https://doi.org/10.1038/s42003-019-0510-7.

Joron, M. and Mallet, J.L. (1998) Diversity in mimicry: paradox or paradigm?, *Trends In Ecology & Evolution*, 13(11), pp. 461–466. https://doi.org/10.1016/s0169-5347(98)01483-9.

Komata, S., Lin, C.-P. and Fujiwara, H. (2022) doublesex Controls Both Hindwing and Abdominal Mimicry Traits in the Female-Limited Batesian Mimicry of Papilio memnon, *Frontiers In Insect Science*, 2. https://doi.org/10.3389/finsc.2022.929518.

Kratochwil, C.F., Kautt, A.F., Nater, A., *et al.* (2022) An intronic transposon insertion associates with a trans-species color polymorphism in Midas cichlid fishes, *Nature Communications*, 13(1). https://doi.org/10.1038/s41467-021-27685-8.

Kyriakidou, M., Tai, H.H., Anglin, N.L., *et al.* (2018) Current srategies of polyploid plant genome sequence assembly, *Frontiers In Plant Science*, 9, p. 1660. https://doi.org/10.3389/fpls.2018.01660.

Lorin, T., Brunet, F.G., Laudet, V., *et al.* (2018) 1 Teleost fish-specific preferential retention of pigmentation gene-containing families after 2 whole genome duplications in vertebrates, *G3* , 8(5), pp. 1795–1806. https://doi.org/10.1534/g3.118.200201.

Lou, R.N., Jacobs, A., Wilder, A.P., *et al.* (2021) A beginner's guide to low-coverage whole genome sequencing for population genomics, *Molecular Ecology*, 30(23), pp. 5966–5993. https://doi.org/10.1111/mec.16077.

Madritsch, S., Burg, A. and Sehr, E.M. (2021) Comparing de novo transcriptome assembly tools in di- and autotetraploid non-model plant species, *BMC Bioinformatics*, 22(1), p. 146. https://doi.org/10.1186/s12859-021-04078-8.

Manni, M., Berkeley, M.R., Seppey, M., *et al.* (2021) BUSCO update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes, *Molecular Biology And Evolution*, 38(10), pp. 4647–4654. https://doi.org/10.1093/molbev/msab199.

Marburger, S., Alexandrou, M.A., Taggart, J.B., *et al.* (2018) Whole genome duplication and transposable element proliferation drive genome expansion in Corydoradinae catfishes, *Proceedings Of The Royal Society B*, 285(1872). https://doi.org/10.1098/rspb.2017.2732.

Marek, P.E. and Bond, J.E. (2009) A Müllerian mimicry ring in Appalachian millipedes, *Proceedings Of The National Academy Of Sciences Of The United States Of America*, 106(24), pp. 9755–9760. https://doi.org/10.1073/pnas.0810408106.

Mattila, A.L.K., Jiggins, C.D., Opedal, Ø.H., *et al.* (2021) Evolutionary and ecological processes influencing chemical defense variation in an aposematic and mimetic Heliconius butterfly, *PeerJ*, 9. https://doi.org/10.7717/peerj.11523.

Mattila, A.L.K., Jiggins, C.D. and Saastamoinen, M. (2022) Condition dependence in biosynthesized chemical defenses of an aposematic and mimetic Heliconius butterfly, *Ecology And Evolution*, 12(6). https://doi.org/10.1002/ece3.9041.

Merrill, R.M., Dasmahapatra, K.K., Davey, J.W., *et al.* (2015) The diversification of Heliconius butterflies: what have we learned in 150 years?, *Journal Of Evolutionary Biology*, 28(8), pp. 1417–1438. https://doi.org/10.1111/jeb.12672.

Muell, M.R., Chávez, G., Prates, I., *et al.* (2022) Phylogenomic analysis of evolutionary relationships in Ranitomeya poison frogs (Family Dendrobatidae) using ultraconserved elements, *Molecular Phylogenetics And Evolution*, 168(107389). https://doi.org/10.1016/j.ympev.2022.107389.

Nakagawa, H., Nagasaka, K., Sakai, H., *et al.* (2015) Isolation of a novel lectin from the dorsal spines of the devil stinger, Inimicus japonicus, *International Aquatic Research*, 7(2), pp. 143–150. https://doi.org/10.1007/s40071-015-0101-2.

Oh, J., Lee, S.-G. and Park, C. (2021) PIC-Me: paralogs and isoforms classifier based on machine-learning approaches, *BMC Bioinformatics*, 22(Suppl 11). https://doi.org/10.1186/s12859-021-04229-x.

Pardo-Diaz, C., Salazar, C., Baxter, S.W., *et al.* (2012) Adaptive introgression across species boundaries in Heliconius butterflies, *PLoS Genetics*, 8(6), p. e1002752. https://doi.org/10.1371/journal.pgen.1002752.

Pardo-Diaz, C., Salazar, C. and Jiggins, C.D. (2015) Towards the identification of the loci of adaptive evolution, *Methods In Ecology And Evolution*, 6(4), pp. 445–464. https://doi.org/10.1111/2041-210X.12324.

Parichy, D.M. (2021) Evolution of pigment cells and patterns: recent insights from teleost fishes, *Current Opinion In Genetics & Development*, 69, pp. 88–96. https://doi.org/10.1016/j.gde.2021.02.006.

Redmond, A.K., Casey, D., Gundappa, M.K., *et al.* (2023) Independent rediploidization masks shared whole genome duplication in the sturgeon-paddlefish ancestor, *Nature Communications*, 14(1). https://doi.org/10.1038/s41467-023-38714-z.

Rice, W.R. (1984) Sex Chromosomes and the Evolution of Sexual Dimorphism, *Evolution*, 38(4), pp. 735–742. https://doi.org/10.2307/2408385.

Robertson, F.M., Gundappa, M.K., Grammes, F., *et al.* (2017) Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification, *Genome Biology*, 18(1). https://doi.org/10.1186/s13059-017-1241-z.

Rowe, L., Chenoweth, S.F. and Agrawal, A.F. (2018) The genomics of sexual conflict, *The American Naturalist*, 192(2), pp. 274–286. https://doi.org/10.1086/698198.

Rowland, H.M., Mappes, J., Ruxton, G.D., *et al.* (2010) Mimicry between unequally defended prey can be parasitic: evidence for quasi-Batesian mimicry, *Ecology Letters*, 13(12), pp. 1494–1502. https://doi.org/10.1111/j.1461-0248.2010.01539.x.

Ruxton, G.D., Allen, W.L., Sherratt, T.N., *et al.* (2018) Avoiding attack: The evolutionary ecology of crypsis, aposematism, and mimicry, in *Avoiding Attack*. 2nd ed. Oxford University Press. https://doi.org/10.1093/oso/9780199688678.001.0001.

Ruzicka, F., Dutoit, L., Czuppon, P., *et al.* (2020) The search for sexually antagonistic genes: Practical insights from studies of local adaptation and statistical genomics, *Evolution Letters*, 4(5), pp. 398–415. https://doi.org/10.1002/evl3.192.

Salzberg, S.L. (2019) Next-generation genome annotation: we still struggle to get it right, *Genome Biology*, 20(1), pp. 1–3. https://doi.org/10.1186/s13059-019-1715-2.

Santos, M.E., Braasch, I., Boileau, N., *et al.* (2014) The evolution of cichlid fish egg-spots is linked with a cis-regulatory change, *Nature Communications*, 5(1), pp. 1–11. https://doi.org/10.1038/ncomms6149.

Speed, M.P. (1999) Batesian, quasi-Batesian or Müllerian mimicry? Theory and data in mimicry Research, *Evolutionary Ecology*, 13(7), pp. 755–776. https://doi.org/10.1023/A:1010871106763.

Stuckert, A.M.M., Saporito, R.A., Venegas, P.J., *et al.* (2014) Alkaloid defenses of co-mimics in a putative Müllerian mimetic radiation, *BMC Evolutionary Biology*, 14(1), pp. 1–8. https://doi.org/10.1186/1471-2148-14-76.

Stuckert, A.M.M., Chouteau, M., McClure, M., *et al.* (2024) The genomics of mimicry: Gene expression throughout development provides insights into convergent and divergent phenotypes in a Müllerian mimicry system, *Molecular Ecology*, 33(14). https://doi.org/10.1111/mec.17438.

Symula, R., Schulte, R. and Summers, K. (2001) Molecular phylogenetic evidence for a mimetic radiation in Peruvian poison frogs supports a Müllerian mimicry hypothesis, *Proceedings Of The Royal Society Of London. Series B*, 268(1484), pp. 2415–2421. https://doi.org/10.1098/rspb.2001.1812.

Van't Hof, A.E., Campagne, P., Rigden, D.J., *et al.* (2016) The industrial melanism mutation in British peppered moths is a transposable element, *Nature*, 534(7605), pp. 102–105. https://doi.org/10.1038/nature17951.

Willing, E.-M., Dreyer, C. and van Oosterhout, C. (2012) Estimates of genetic differentiation measured by FST do not necessarily require large sample sizes when using many SNP markers, *PloS One*, 7(8). https://doi.org/10.1371/journal.pone.0042649.

Wilson, J.S., Williams, K.A., Forister, M.L., *et al.* (2012) Repeated evolution in overlapping mimicry rings among North American velvet ants, *Nature Communications*, 3(1), pp. 1–7. https://doi.org/10.1038/ncomms2275.

Winters, A.E., Wilson, N.G., van den Berg, C.P., *et al.* (2018) Toxicity and taste: unequal chemical defences in a mimicry ring, *The Royal Society Proceedings. Biological Sciences*, 285(1880). https://doi.org/10.1098/rspb.2018.0457.

Yandell, M. and Ence, D. (2012) A beginner's guide to eukaryotic genome annotation, *Nature Reviews. Genetics*, 13(5), pp. 329–342. https://doi.org/10.1038/nrg3174.

Zhang, W., Dasmahapatra, K.K., Mallet, J., *et al.* (2016) Genome-wide introgression among distantly related Heliconius butterfly species, *Genome Biology*, 17(1). https://doi.org/10.1186/s13059-016-0889-0.

Ziegman, R., Undheim, E.A.B., Baillie, G., *et al.* (2019) Investigation of the estuarine stonefish (Synanceia horrida) venom composition, *Journal Of Proteomics*, 201, pp. 12–26. https://doi.org/10.1016/j.jprot.2019.04.002.