U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

MSc

2.º CICLO

FCUP
2017

U.PORTO

The Genetic Basis of Plumage Colour Variation in the Gouldian finch (*Erythrura gouldiae*): an RNA-seq approach

Cristiana Isabel Jorge Marques

FC

The Genetic Basis of Plumage Colour Variation in the Gouldian finch (*Erythrura gouldiae*): an RNA-seq approach

Cristiana Isabel Jorge Marques

Dissertação de Mestrado apresentada à
Faculdade de Ciências da Universidade do Porto em
Biodiversidade, Genética e Evolução

2017

The Genetic Basis of Plumage Colour Variation in the Gouldian finch (*Erythrura gouldiae*): an RNA-seq approach

Cristiana Isabel Jorge Marques
Mestrado em Biodiversidade, Genética e Evolução
Departamento de Biologia
Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO)
2017

Orientador
Miguel Carneiro, Post-Doctoral Researcher, Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO)

Co-orientador
Ricardo Jorge Lopes, Post-Doctoral Researcher, Centro de Investigação em Biodiversidade e Recursos Genéticos (CIBIO)

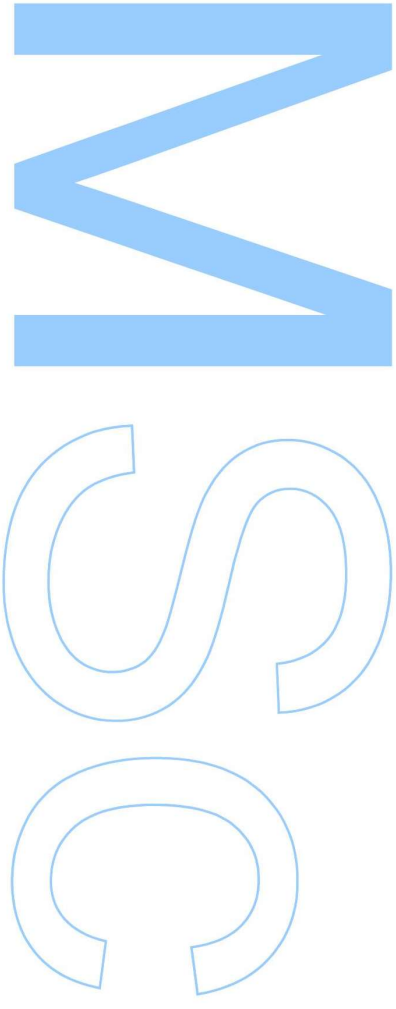
U.PORTO

FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Todas as correções determinadas pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto. / /



“Chaos was the law of nature,
Order was the dream of man”

- Henry Adams, *in “*The
Education of Henry Adams”

Acknowledgements

Throughout this year I faced many challenges that I would not be able to overcome without the help and support of the wonderful people surrounding me. I hope they know how grateful I am for having them in my life, whether they are new acquaintances or old friends. With the closing of this cycle, I cannot help to express my gratitude:

To my supervisor, Miguel Carneiro, for accepting me as his student and assembling a phenomenal project for my Master thesis. For all the guidance and understanding throughout my learning process, while keeping me on my toes. For giving me the opportunity to grow scientifically and for keeping me motivated in discouraging times. And, finally, for respecting my individuality throughout the course of writing this thesis. I honestly could not be more pleased with my choice.

To my co-supervisor, Ricardo Lopes, for helping with several basal tasks, for reviewing this document and sharing constructive ideas. Without his working skills would be difficult to achieve most of the sampling, essential to my study.

To Pedro Andrade, my colleague and friend, for the immense support, patience and help with my first steps into genomics and evolutionary genetics. For helping me put things in perspective and for the hours he spent teaching me fundamental principles and discussing new ideas. A heartfelt thank you!

To Sandra Afonso and Jolita, for handling most of the lab work involved in this and future studies. For passing on their expertise, so I can learn fast and get good results.

To my friends and family, that helped me through hard times and did not let me give up on myself. For sharing my ambitions and conquests. For all the coffee brakes, jokes, tender moments, philosophical moments and stupid playful moments that light up my smile on a daily basis.

To all the people that, in some way, contributed to my work. You are many and I will not forget your help. I hope the opportunity arises for me to repay you.

Abstract

Birds are the most diverse and successful group of vertebrates, exhibiting a wide array of colourful ornaments. Colour polymorphism is often associated with other biological traits, such as physiology and behaviour, and usually leads to rapid phenotypic evolution. Therefore, polymorphic systems provide critical insights into processes by which phenotypic diversity is generated and maintained within a species. In the Gouldian finch (*Erythrura gouldiae*), three head colour morphs (red, black and yellow) coexist in sympatry. These colour phenotypes follow simple Mendelian segregation (a Z-linked gene determines red/black head colour and an autosomal locus controls red/yellow carotenoid-based colour) and promote assortative mating. Despite valuable contributions, previous works have left the genetic basis of head colour polymorphism in this species, unresolved.

Gouldian finches have been extensively bred in captivity throughout the years. As such, several new mutations arose. The white chest mutation, in domestic gouldian finches, results from a regional loss of the structural purple colouration, therefore providing an opportunity to explore the genetic basis of melanin pigmentation and body patterning in birds. In this thesis, we combined differential expression analyses and genome-wide differentiation analyses, using RNA-seq data, to examine the molecular genetic basis of plumage colour variation in the Gouldian finch. Expression data analysis revealed a possible structural modification in *CYP2J-like* gene complex, specific to yellow-headed birds, and several upregulated genes involved in epidermal formation and follicular growth. Meanwhile, genetic differentiation analyses comparing melanic and carotenoid-based phenotypes enabled a considerable narrowing of the candidate window previously described, but were not effective in pinpointing a candidate gene involved in head melanism. We attribute these difficulties to insufficient sequencing effort and hypothesize that we may have missed the precise time of follicular growth in which our target genes were being expressed. Regarding the chest regional loss of pigmentation, we succeeded in mapping a candidate genomic region, comprising *TYR* and *RAB38*, two genes known to impair melanin biosynthesis and deposition. Exciting new *TYR* mutations were found using functional annotation. However, these mutations alone seem unlikely to induce regional loss of melanin pigmentation. Taking our results into account, follow-up studies will proceed with increased sampling and alternative genomic tools, to further confirm our hypotheses. Overall, this study highlighted the importance of experimental design when using RNA-seq approaches and provided crucial information to determine and guide future experiments. We expect more thorough

complementary analyses will finally resolve the genetic machinery underlying plumage colour variation in wild and domestic gouldian finches.

Keywords: Colour polymorphism; Gouldian finch; *Erythrura gouldiae*; RNA sequencing; Gene expression analysis; Selective sweep mapping; Association mapping

Resumo

As Aves são o grupo mais diverso e bem-sucedido de vertebrados, exibindo um leque variado de ornamentos coloridos. Os polimorfismos de cor estão frequentemente associados a outros aspetos biológicos, tal como a fisiologia ou o comportamento e, habitualmente conduzem a uma rápida evolução fenotípica. Como tal, sistemas polimórficos proporcionam uma visão aprofundada dos processos que dão origem e mantêm a diversidade de uma espécie. O Diamante de Gould (*Erythrura gouldiae*) apresenta-se em três cores de cabeça distintas (vermelho, preto e amarelo) que coexistem em simpatria. Estes fenótipos seguem uma segregação Mendeliana (um gene ligado ao cromossoma sexual determina a deposição de vermelho ou preto, enquanto um locus autossómico controla a expressão das variedades vermelho/amarelo) e promovem o acasalamento seletivo nas populações. Apesar de contribuições valiosas, estudos anteriores não foram capazes de resolver a base genética deste polimorfismo de cor. Os Diamantes de Gould têm sido criados em cativeiro extensivamente ao longo dos anos e, como tal, novas mutações emergiram. A mutação "peito-branco" em Diamantes de Gould resulta da perda localizada do tradicional roxo estrutural, portanto, abrindo uma oportunidade para explorar a base genética da pigmentação melânica, assim como dos padrões corporais nas Aves. Nesta tese, combinamos análises de expressão diferencial e estimativas de diferenciação genética ao longo do genoma, usando dados de *RNA-sequencing*, para examinar a base molecular da variação de cor da plumagem no Diamante de Gould. As análises de expressão revelaram uma possível modificação estrutural no complexo *CYP2J-like,* especificas de animais com cara amarela, e variados genes envolvidos no crescimento folicular com expressão incrementada. As análises de diferenciação comparando fenótipos melânicos e não-melânicos permitiram reduzir consideravelmente o tamanho da janela candidata, previamente mapeada. No entanto, não foi possível precisar um gene candidato para melanismo. Atribuímos estas dificuldades a um esforço de sequenciação insuficiente e a uma possível falha no timing de amostragem. Relativamente, à perda regional de pigmentação no peito, fomos capazes de mapear uma região genómica candidata que incluía dois genes reconhecidos por inviabilizar a síntese e deposição de melanina. Novas mutações na região codificante do *TYR* foram anotadas funcionalmente. Contudo, estas mutações, *per se,* não deverão induzir uma perda regional de pigmentação. Considerando os resultados, uma série de estudos complementares estão em progresso, incrementando a amostragem e usando ferramentas genómicas alternativas, para testar e confirmar novas hipóteses. No geral,

este estudo evidenciou a importância de um desenho experimental robusto em abordagens baseadas em *RNA-seq* e providenciou informação crucial que determinará estudos futuros. Finalmente, esperamos que analises complementares mais robustas possam elucidar a maquinaria genética responsável por variação de cor da plumagem, em diamantes de Gould selvagens e domésticos.

Keywords: Polimorfismo; Diamante de Gould; *Erythrura gouldiae*; *RNA sequencing*; Análise de expressão génica; *Selective sweep mapping*; *Association mapping*

Index

List of Figures

Chapter I: General Introduction

Chapter II

Chapter III

List of Tables

Chapter II

Chapter III

List of Abbreviations

Dct - Dopachrome tautomerase

TYRP1 - Tyrosinase related protein

Mc1R - Melanocortin-1 receptor

α-MSH - Melanin-stimulating hormone

Mitf - Melanogenesis Associated Transcription Factor

TYR - Tyrosinase protein

EDN3 - Endothelin 3

SOX10 - SRY-related HMG-box 10

PMEL17 - Premelanosome protein 17

MLPH - Melanophilin

ASIP - Agouti signalling protein

CYP2J19 - Cytochrome P450 2J19 member

SCARB1 - Scavenger Receptor B1

BCO2 - Beta-Carotene Oxygenase 2

DNA - Desoxiribonucleic Acid

GWAS - Genome-Wide-Association Studies

QTL - Quantitative trait loci

RNA - Ribonucleic Acid

RIN - RNA Integrity Number

cDNA - complement DNA

PCR - Polymerase Chain Reaction

qPCR - quantitative Polymerase Chain Reaction

NCBI - National Centre for Biomedical Investigation

FPKM - Fragments Per Kilobase Million

GFF - General Feature Format

GTF - Gene Transfer Format

GATK - Genome Analysis Toolkit

GVCF - Genomic Variant Call Format

CIGAR - Compact Idiosyncratic Gapped Alignment Report

VQSR - Variant Quality Score Recalibration

FS - Fisher Strand Test

QD - Quality by Depth

FST - Fixation Index

SNP - Single Nucleotide Polymorphism

SIFT - Scale-Invariant Feature Transform

MQ - Mapping Quality

IGV - Integrative Genome Viewer

FDR - False Discovery Rate

FZD5 - Frizzled Class Receptor 5

WIF1 - Wnt Inhibitory Factor 1

GLI3 - GLI Family Zinc Finger 3

Shh - Sonic hedgehog signalling

ZNF638 - Zinc Finger Protein 638

LONRF3 - LON Peptidase N-Terminal Domain Ring Finger 3

COMT - Catechol-O-Methyltransferase

ARO1 - Penta functional AROM polypeptide 1

NOP14 - Nucleolar Protein 14

NDUFS4 - NADH Ubiquinone Oxidoreductase Subunit S4

IL31RA - Interleukin 31 Receptor A

IL6ST - Interleukin 6 Signal Transducer

GAS2L3 - Growth Arrest Specific 2 Like 3

MLANA - Melan-a

SLC45A2 - Solute Carrier 45 Member A2

RAB38 - RAS Oncogene Family Member

CHST1 - Carbohydrate Sulfotransferase 1

FST - follistatin

Chapter I

General Introduction

Among all terrestrial vertebrates, birds are undoubtedly the most diverse and successful group. Such astonishing adaptability is, in great part, attributed to their permanently evolving plumage (Haavie et al. 2000; Yeh 2004; Zink et al. 2004; Stoddard & Prum 2011), which not only allowed them to take flight and insulate from the cold, but also evolved through colouration either concealing or conveying specific messages (Bradbury & Vehrencamp 2011). In fact, colouration is such an important part of birds' life-history that it often serves as a strong driver for natural and sexual selection, ultimately culminating in speciation (Greene et al. 2000; Dunn et al. 2015; Gomes et al. 2016).

1. Mechanisms of Colour Production

Plumage colouration may be attained either by deposition of pigments or through structural changes in feathers. Biochromes, or natural pigments, (e.g. Melanin, carotenoids, porphyrins and pterins, etc) are deposited in feathers during growth and vary widely in their properties, thus differing greatly in the way they are incorporated and displayed on feathers. In addition to structural colouration, here, only melanin- and carotenoid-based colouration will be explored.

1.1. Carotenoid-based pigmentation

Carotenoids generate a range of yellow to red colours in birds, being found in almost every integumentary tissue, including eggs, skin and its appendages, eyes and beaks. Carotenoids are hydrocarbons consisting of a symmetrical conjugated double

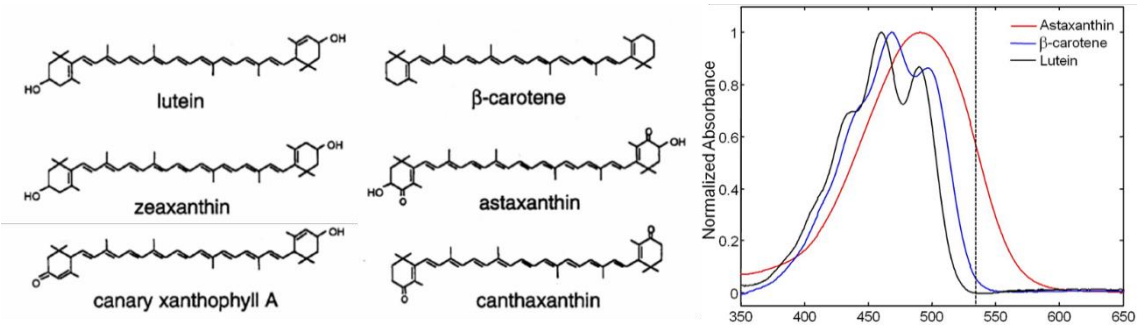


Figure 1.1: Carotenoid chemical structure and absorbance spectra

bonds backbone, cyclized at each end (Figure 1.1). Their chemical structure is highly variable, fact that inevitably affects their ability to absorb radiation (Figure 1.1). Red carotenoids, such as astaxanthin and canthaxanthin, possess ketone functional groups (ketocarotenoids) shifting their absorbance to higher frequency wavelengths, whilst yellow carotenoids like lutein or canary-xanthophyll, incorporate hydroxyl into the chromophore (hydroxy-carotenoids).

Vertebrates cannot synthesize carotenoids, instead, they must ingest it through diet by feeding on algae, plants and carotenoid-rich animal prey (Brockmann & Volker 1934; Fox 1962). Following ingestion, carotenoid absorption takes place in the intestine, where site-specific absorption mechanisms are activated for xanthophyll assimilation (Tyczkowski & Hamilton 1986; Figure 1.2). Blood circulation and delivery to internal organs is exclusively done by lipoproteins and scavenger receptors. Some birds may directly display unmodified carotenoids, whereas others profit from a complex array of enzymes to undergo biochemical modification and produce a variety of hues (Lafountain et al. 2015). For instance, ketolases often tune carotenoids to generate red colours, by adding a ketone group to the end rings and lutein can be subtly modified to produce orange by undergoing epoxidation (Cooke & Buckley 1987; Meléndez-Martínez et al. 2006). These metabolic steps can occur in the liver (Brush 1990; Wyss 2004), which harbours enzymes that breakdown lipids, or in colourful tissues directly (Kritzler 1943; Inouye et al. 2001; McGraw 2004). Lastly, processed carotenoids are incorporated into the target maturing cells, such as feather follicles, through passive lipid diffusion (Yu et al. 2004).

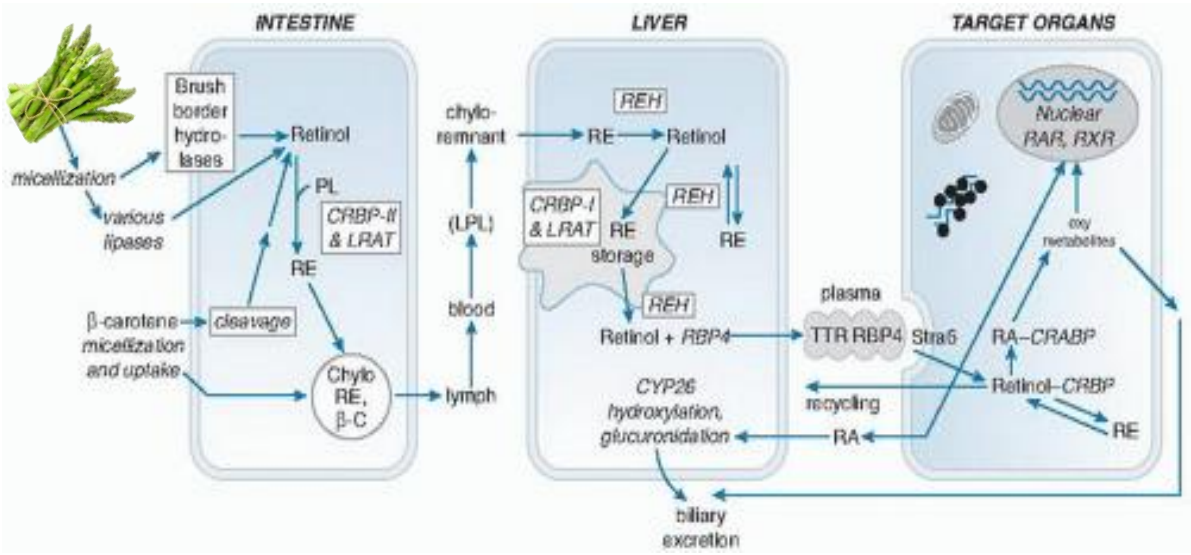


Figure 1.2: Carotenoid metabolic pathway from absorption to incorporation into target maturing cells

Because carotenoids are exclusively withdrawn from the environment, changes in dietary composition, either in quantity or quality, may influence which carotenoids are

deposited into different tissues (Hill & McGraw 2006). Furthermore, birds selectively allocate these resources according to their needs, whether it is to enhance their beauty (Lozano 2001), boost immunity and health (Fenoglio et al. 2002; McGraw & Ardia 2003) or for protection against oxidative stress (Burton 1988). Due to these nutritional and physiological benefits, as well as the inherent costs in producing it, carotenoid-based colouration evolved as an honest signal of individual fitness (Peters et al. 2007). Hence, a clear understanding of how ecological and biochemical constraints influence carotenoid metabolism, would be significantly increase the power to infer target genes (or pathways) being shaped by natural and sexual selection.

1.2. Melanin-based pigmentation

Melanin pigmentation is the most prevalent strategy of colouration in the integumentary tissues of birds, conferring the majority of the black, grey, brown and rufous colours seen in Nature. Even though birds rely on striking plumage colours to produce the most spectacular displays, they also use melanin pigments to achieve an astonishing array of hues and patterns both within and across feathers (Figure 1.3). Moreover, the strategic placement of melanic bands and spots for face-to-face interactions, or covering the whole body for concealment, speaks to its function in communication (Pérez-Rodríguez et al. 2017). Such intricacy of melanic patterns strongly suggests that melanogenesis is under local control at maturing feather follicles and keratinocytes in birds.

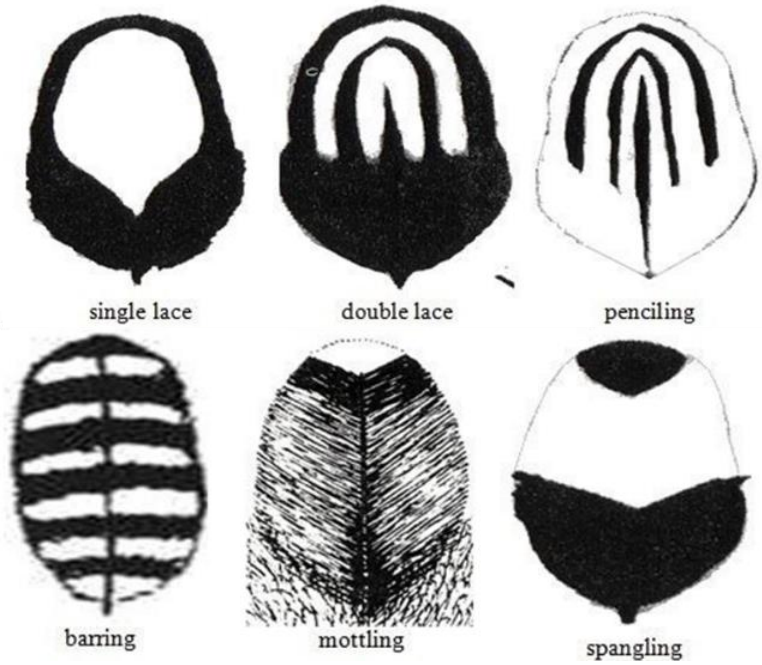


Figure 1.3: Examples of melanin-based pigmentation patterns within feathers of chickens (*Gallus gallus*). Adapted from © 2013-2017 THE 104HOMESTEAD

Eumelanin and pheomelanin are considered nitrogen-containing biochromes, although their full chemical structure has never been described due to its large size, cross-linking properties and insolubility. Synthesis of melanin starts with the conversion of tyrosine into dopaquinone, with the rate-limiting enzyme being tyrosinase. Dopaquinone is further processed to produce pheomelanin polymers. When cysteine is depleted, eumelanin synthesis begins. In the presence of the dopachrome tautomerase (Dct), dopachrome undergoes a series of biochemical processes, including oxidation by tyrosinase related protein (TYRP1), to form eumelanin granules (Ito and Wakamatu 2010; Figure 1.4). Relative amounts of eumelanin and pheomelanin are mainly regulated by the melanocortin-1 receptor (Mc1R). During feather growth, coupling of melanin-

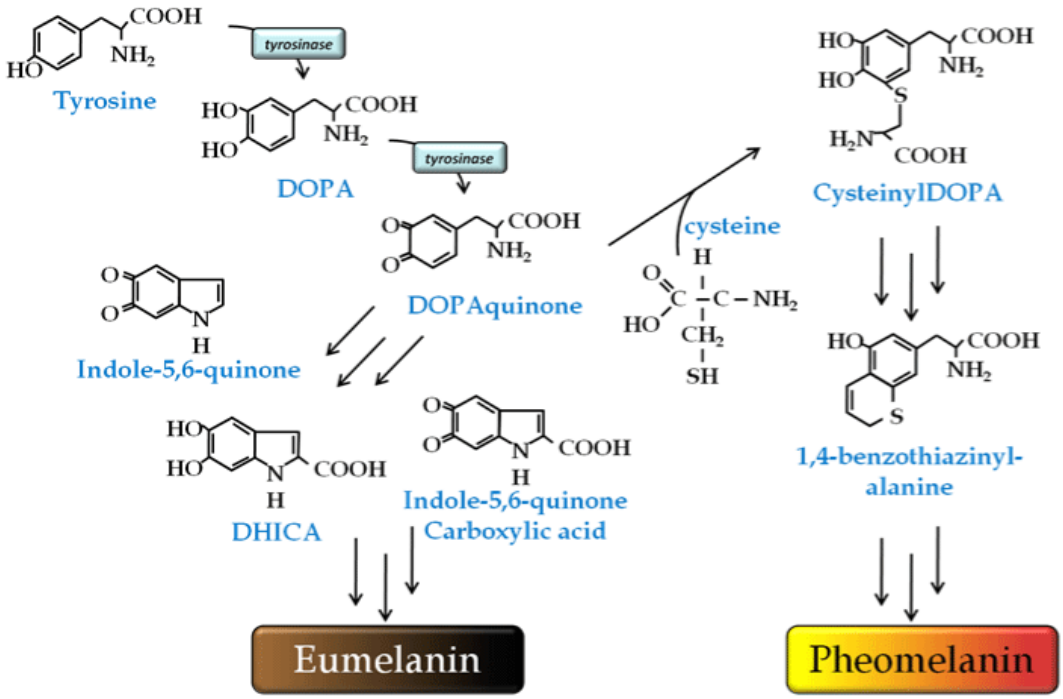


Figure 1.4: Melanin biosynthesis.
Melanins are large heteropolymers derived from the amino acids tyrosine or cysteine. Epidermal melanins are present in two major forms: **eumelanin**, an effective UV-blocking brown/black pigment and **pheomelanin**, a sulfur-containing red/blonde pigment permeable to UV photons. Image credits: Copyright © 2017 OMICS International

stimulating hormone (α-MSH) induces the transcription factor Mitf and increases melanosome biogenesis. On the other hand, binding of Agouti signalling proteins, antagonists of melanocortins, results in decreased production of eumelanin in favour of pheomelanin.

All vertebrates synthesize melanin endogenously in dedicated organelles called melanosomes. They originate as unpigmented cells, derived from the neural crest, and migrate to the epidermis, where they differentiate and begin synthesizing melanin. Eventually, melanosomes are delivered to keratinocytes by exo/endocytosis (Tarafder et al. 2014). In bird integumentary tissues, melanogenesis produces a mixture of

pheomelanin and eumelanin, and it is the relative proportion of these pigments that dictates the variability of melanin-based colours and shades (McGraw et al. 2004). Hormones such as testosterone and oestrogen, were shown to influence melanisation in birds, either by extending melanin patches (González et al. 2001) and triggering the expression of nuptial plumage (Lank et al. 1999), or by preventing any sexual dichromatism (Kimball & David Ligon 1999).

Despite the large amount of work done in this colour-producing mechanism, researchers have primarily focused on the factors controlling melanin pigmentation and its possibilities for condition-dependence, whereas less is known about melanosome transfer from follicles to feathers or what molecular mechanisms underlie melanosome site-specific control of melanin deposition to form intricate pigmentation patterns.

1.3. Structural colouration

Opposing to pigmentary colours, which depend entirely on the molecular structure and abundance of pigments, structural colours rely on the presence of nanoscale arrays built into the integument (Figure 1.5). This type of colouration is commonly seen in invertebrates giving them an iridescent look, resulting from light passing through a stack of keratin layers and creating constructive interference.

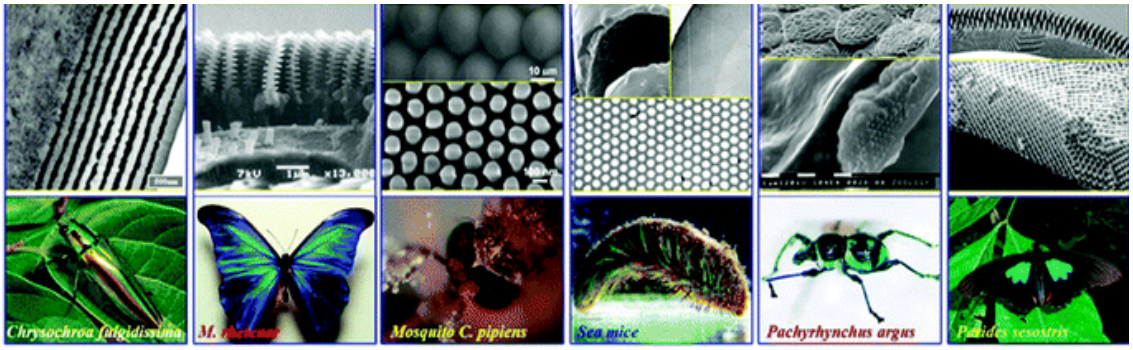


Figure 1.5: Diversity of biophysical matrices for iridescent structural colouration. Source: (Zhao et al. 2012)
Left to right: transmission microscopy of real tissues (**top**) and respective model organisms (**bottom**).

In bird feathers, these colour-producing matrices are generally composed of keratin and air organized in a quasi-ordered fashion, which selectively reflect light to generate specific hues (Figure 1.6). Furthermore, the tissue below a colour-producing nanostructure cannot itself produce colour, due to incoherent scattering of all visible wavelengths. Hence, incorporating a layer of pigment granules (usually melanin) underneath has proven to be evolutionarily profitable, since it absorbs all propagated light functioning as an optical insulator (Shawkey 2006). Such mechanism aids in the production of highly saturated brilliant colours in nature that do not change with the angle

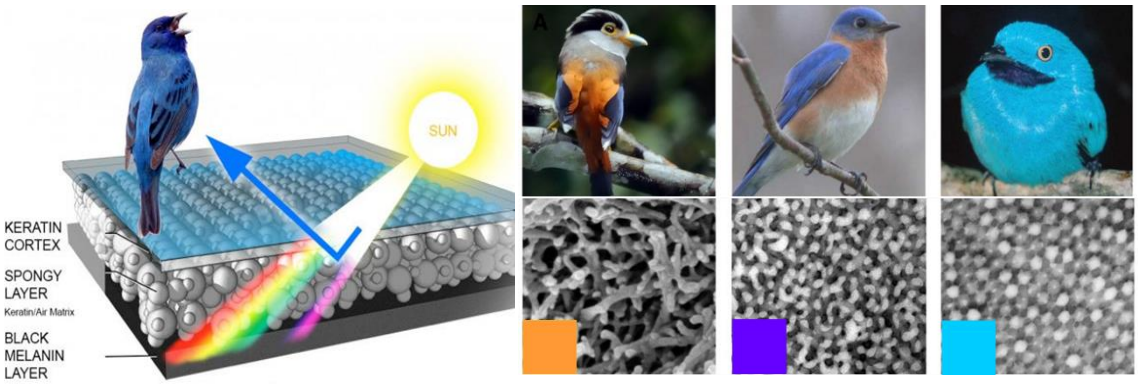


Figure 1.6: Non-iridescent structural colouration in birds.
Left: schematics of feather barb nanostructure producing non-iridescent colouration. Right: examples of structural fine tuning to produce the presented hues. Adapted from (Saranathan et al. 2012) and ©Cornell lab of Ornithology

of light incidence. Another possibility is a combined product of structural and pigmentary mechanisms, where the resulting hue is uniquely saturated and usually unavailable through any of the mechanisms on their own This is frequently seen in captive birds, such as Budgerigars (*Melopsittacus undulatus*; D'Alba et al. 2012; Marshall & Johnsen 2017) and Gouldian finches (*Erythrura gouldiae*), whose structural blue colouration was combined with yellow carotenoid pigments to produce rich green plumage. Different combinations of alternative substances such as collagen, mucopolysaccharides and cytoplasm, were also shown to adorn birds' skin and bill, and can have a determinant role in visual communication (Prum 2003).

2. Genetic basis of colour polymorphisms in birds

Quantitative variation in plumage colouration is elemental to many aspects of avian biology, including the evolution, maintenance and adaptive function of phenotypes. However, the few studies carried out have been focused almost exclusively on sexual ornaments and melanin-based plumage traits (Roulin & Ducrest 2013; Saino et al. 2013; Roulin et al. 2012; Badyaev & Young 2004; Bortolotti et al. 2006; Bourgeois et al. 2017; see Hill & McGraw 2006 and references therein)

Bird fanciers soon realized the potential in domesticated species for the study of colour inheritance, recording and elucidating these patterns in such species as the Domestic chicken (*Gallus gallus*), Zebra finches (*Taeniopygia guttata*), Budgerigars and many others. This initial body of work helped to determine the number of loci involved in a given phenotype, mode of segregation and relationships among loci. It has further allowed for demonstration of genetic interactions among alleles of different loci. For

instance, in domestic pigeons two linked autosomal loci interact with a third sex-linked locus to accomplish the full colour spectrum seen (Domyan et al. 2015).

The presence of colour polymorphisms (discrete categories of plumage colouration in an interbreeding population), provides a unique opportunity to study the genetic mechanisms of colour variation. Since many colour traits follow simple Mendelian principles and, often, colour variation in wild populations resembles that of captive animals, one can foretell candidate loci implicated in similar colour morphs in different species. The existence of dramatic colour morphs, also indicates strong genetic control of traits, usually involving few genes of major effects.

The first isolated pigmentation gene in birds was the one encoding tyrosinase (*TYR),* whose loss of function leads to the complete loss of melanin pigments (MOCHII et al. 1992). A relatively large number of studies have shown that variation in melanin-based colouration is due to polymorphisms at the *Mc1R* gene, predominantly resulting in melanism by constitutive activation of melanin deposition (Takeuchi et al. 1996; Theron et al. 2001; Mundy 2005), changes in melanin ratio (Baião & Parker 2012) or distribution possibly through regulatory features (Mundy 2005). Alternative sets of genes have also been found to modify melanin-based phenotypes in different ways, by modifying melanocyte migration and differentiation (*EDN3*, Dorshorst et al. 2011; *SOX10*, Gunnarsson et al. 2011), melanosome structure (*PMEL17*, Kerje et al. 2004), melanosome transport (*MLPH*, Vaez et al. 2008), the balance of eumelanin to pheomelanin production (*ASIP*, Hiragaki et al. 2008); *TYRP1*, Lehtonen et al. 2012) and other components of the melanin pathway (*Mitf*, Minvielle et al. 2010).

Fortunately, progress has now been made in isolating loci controlling for carotenoid deposition and metabolism on birds' feathers and bare parts, in part driven by new technological advances. Two independent studies identified a key enzyme, CYP2J19, that mediates the conversion of yellow into red ketocarotenoids (Lopes et al. 2016; Mundy et al. 2016). In a different study, the same researchers have found that mutations in a scavenger receptor (*SCARB1*) impaired carotenoid uptake into canaries' bloodstream (Toomey et al. 2017). Also, flanking sequence polymorphisms in the carotenoid-cleaving enzyme β-carotene-9′,10′-dioxygenase (BCO2) were associated to yellow/white chest morphs in wood warblers. Finally, tissue-specific regulation of *BCO2* expression has been shown to cause yellow skin in domestic chickens (Eriksson et al. 2008). To the best of our knowledge, no work has been done to elucidate *loci* affecting plumage structural colouration through changes in feather structure alone.

3. The study of colouration in the Genomic Era

Recent advances in molecular techniques and DNA sequencing have propelled us into a new era of remarkable discoveries, providing us with the tools to answer fundamental questions in evolutionary biology and opening a window for asking new ones. Indeed, current studies addressing complex topics in ecology, evolution and patterns of diversification are rapidly moving towards high-throughput approaches, which may now be more cost-effective than previous methods using a limited number of genetic markers.

Different approaches have been used to provide novel insights into the genetic basis of colour variation in birds, being whole-genome sequencing the most useful in the long run (Rubin et al. 2010; Jarvis et al. 2014; Lamichhaney, Berglund, et al. 2015; Toews et al. 2016; Lopes et al. 2016). For instance, the Domestic chicken (Hillier et al. 2004) and the Zebra finch (Warren et al. 2010) were the first avian taxa for which full genomes were produced, annotated and made available. Due to conserved genome size and high chromosomal synteny, typical of avian groups, these references keep contributing useful structural information for studies focusing in non-model systems. Genome-wide association studies (GWAS) and quantitative trait loci (QTL) mapping, have revealed a considerable amount of information regarding plumage colour diversification across taxa (Shapiro et al. 2013; Poelstra et al. 2013; Ellegren 2014; Toews et al. 2016; Lopes et al. 2016; Mundy et al. 2016; Toomey et al. 2017). A contemporary interesting case also benefitted from the use of these genomic tools to explain differences among colour morphs in breeding behaviour. Indeed, two independent studies determined that a major genomic inversion comprising a supergene promoted the coevolution of genes affecting the Ruff (*Philomachus pugnax*) reproductive traits (Lamichhaney, Fan, et al. 2015; Küpper et al. 2015).

It has been increasingly perceptible that mutations on regulatory features and patterns of gene expression between groups may be linked to phenotypic divergence. Hence, RNA sequencing data is becoming increasingly common for exploring patterns of intraspecific differentiation. RNA-seq targets coding DNA sequences being expressed in a given tissue at the time of sampling and is currently applied in genetic marker development (Ramstad et al. 2016; Kaiser et al. 2017; Höglund et al. 2017; Watson et al. 2017), transcriptome assembly and analysis (Ferreira et al. 2017; Poelstra et al. 2015) and search for candidate genes controlling important adaptive traits (Nachman et al. 2003; Poelstra et al. 2015; Ferreira et al. 2017; Pan et al. 2017). It is also useful in studies of alternative splicing and gene expression analysis of developmental timelines (Wang et al. 2010; Gouti et al. 2017). Ultimately, different strategies can be combined for

increased robustness and accuracy. One such case is that documented by Poelstra et al. 2015. Using whole-genome sequencing combined with expression data they found that phenotypic variation in carrion and hooded crows resulted from the interaction of several regulatory melanin genes. Whatever the genomic tool of choice, its application to the study of avian colouration and patterning is still in its infancy, however, with promise to revolutionize our knowledge of the evolutionary mechanisms underlying phenotypic diversity.

4. General Objectives

In wild populations, polymorphisms coordinating the regulation of melanin and carotenoids are rare, especially when occurring in the absence of geographic isolation. In Gouldian finches (*Erythrura gouldiae*) the presence of three discrete head colour morphs, involving variation in melanin and carotenoids, is suggestive of strict genetic control. We hypothesize that divergent expression patterns might be the main reason for the emergence of alternative head colour morphs in the Gouldian finch, both in carotenoid-based and melanin-based colouration. The Gouldian finch, therefore, constitutes a unique system to understand the genetic basis of colour polymorphisms.

This species also displays a variety of colours in the remaining parts of the body in consequence of the domestication process. One interesting case is the loss of pigmentation in the chest within a very precise area of the integument, not affecting other body parts, suggesting that colour patterning in these finches is highly modular. Based on the few cases in which similar colour variation was studied from a biophysical perspective, we hypothesize that the white chest mutation in gouldian finches is a product of the loss of the melanin granules underlying the structural matrix. Thus, an impairment of genes related to melanin biosynthesis or melanosome transport and docking might explain this phenotype.

In this thesis, we profited from newly developed genomic tools, to examine the molecular genetic basis of plumage colour variation in the Gouldian finch using an RNA-sequencing approach. Chapter II addresses the genetic mechanisms related to head colour polymorphism and Chapter III focuses on chest colour variation.

Specifically, this research work aims at:

- Characterizing expression profiles and performing genetic differentiation analyses to map candidate genes and causal mutations implicated in sexually selected plumage traits, such as carotenoid and melanin-based colouration.
- Understanding the genetic basis underlying plumage colour variation in the chest area by enlightening genetic changes associated with site-specific loss of pigmentation and patterning.

Ultimately, we expect that identifying the genetic architecture of variation in carotenoid-based traits largely improves our understanding of carotenoid processing and variation, providing also novel insights into their evolution. Additionally, we expect to provide novel insights into the regional genetic regulation of body patterning in birds, by elucidating which mechanisms underlie melanin-based colouration in gouldian finches and how they affect structural colour production.

Chapter II

The genetic basis of head colour polymorphism in the Gouldian finch

1. Introduction

Assortative mating is known to be a key promoter of the speciation process and it may be of particular importance in a context of sympatric speciation, especially at the onset of reproductive isolation between populations. At these early stages of the divergence process, individuals either fail to interbreed or, if hybridization takes place, the two lineages may collapse back into a single population (Figure 2.1). This merging can be prevented if gene flow is constrained by premating barriers, since postmating barriers such as differential fertilization success or decreased offspring fitness (Price 2007) typically take longer to evolve. Though assortative mating is important in generating initial divergence, it also plays an important role in maintaining reproductive isolation after secondary contact (Figure 2.1). Nevertheless, how assortative mating evolves in diverging populations undergoing gene flow, remains obscured.

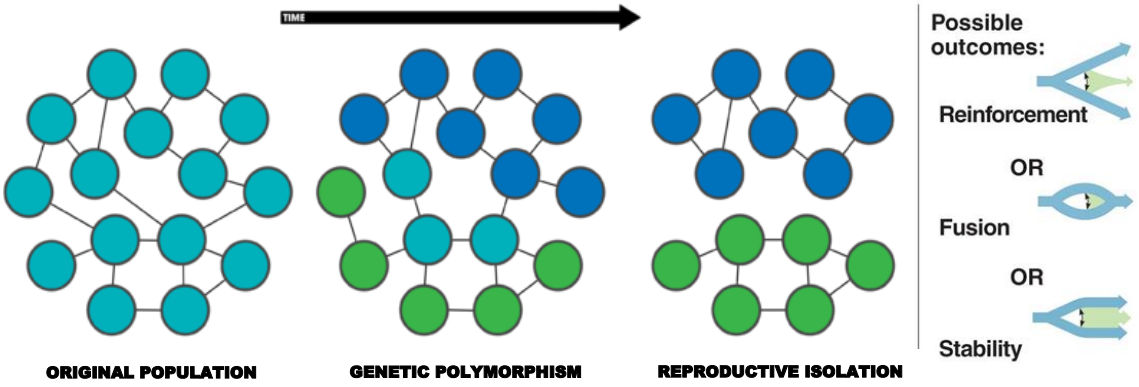


Figure 2.1: Sympatric speciation mode and possible outcomes.
Adapted from Wikipedia© and Pearson Education, Inc©

The Gouldian finch, *Erythrura gouldiae*, is a small highly polymorphic passerine, belonging to the Estrildidae family (Arnaiz-Villena et al. 2009). In this endemic Australian finch, three head colour-morphs (red, black and yellow; Figure 2.2) coexist, in sympatry, throughout its entire geographic range at stable frequencies (Franklin & Dostine 2000; Gilby et al. 2009). Even though no ecological divergence is known, alternative competitive strategies exist between different morphs. For instance, red-faced

individuals are behaviourally dominant towards black and yellow morphs (Pryke 2007), being this difference induced by physiological processes such as the hormonal control of aggressiveness and health status (Pryke & Griffith 2007).

Figure 2.2: The Gouldian finch head colour variation. Only males are shown.
1. Red morph, produced by natural red carotenoid (canthaxanthin); **2.** Black morph, melanin based pigmentation (eumelanin) and **3.** Yellow morph, also produced by natural carotenoid pigmentation (lutein). **Image credits: © 2011 gouldianfinches.eu**

The expression of red/black head colour in this species is controlled by a single Z-linked *locus* with a dominant red (Z^R) and a recessive black (Z^r) allele (Southern 1945). As in all bird species, the heterogametic sex is represented by females, whose phenotype matches the genotype directly (Z^R red head, Z^r black head), whereas males may be homozygous Z^R Z^R (red) and Z^r Z^r (black), or heterozygous Z^R Z^r (red). Heterozygous and homozygous red males are phenotypically indistinguishable. In contrast, yellow head colour is an autosomal recessive trait. An autosomal *locus* interacts with the *Red locus*, possibly through suppression, resulting in the deposition of yellow lutein epoxide instead of red canthaxanthin (Cooke & Buckley 1987). The yellow head gene is only fully expressed when the bird carries at least one copy of the red gene ($Z^{RR}A^{yy}$ or $Z^{Rr}A^{yy}$). In the absence of the red gene, birds that are homozygous for the yellow autosomal trait ($Z^{rr}A^{yy}$) will have black heads and yellow-tipped bill instead of the typical red-tipped bill. Females and males are able to discriminate between morphs and demonstrate assortative mate preference for their own morph-type (Pryke & Griffith 2007).

Due to genetic incompatibilities between different colour genotypes (black/red), which resulted in reduced offspring survival (Figure 2.3), assortative mating is likely to be adaptive. However, despite high fitness costs, different morph pairing is not uncommon (Sarah R. Pryke & Griffith 2009). This fact corroborates Haldane's rule in which hybrid sterility or unviability affects the heterogametic sex preferentially (female offspring originating from mixed-morph crosses, have about 40% less chance of survival; Haldane 1922). It is unclear whether yellow headed females show the same preference and if mating with a genetically opposite morph yields similar results.

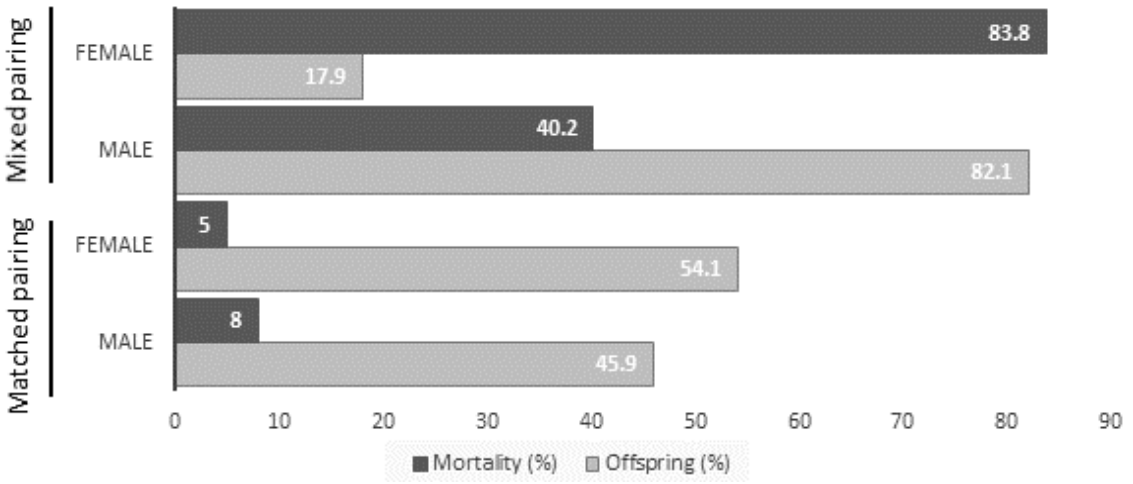


Figure 2.3: Offspring and Mortality associated to morph pairing. Adapted from (Pryke & Griffith 2009)**.** Offspring sex ratios become strongly male-biased when females are paired with genotypically different males (**Mixed pairing**) and mortality rates increase considerably, when compared to same-morph pairings (**Matched pairing**). In addition, daughters originating from mixed-morph pairings experience 43.6% higher mortality than sons. Mortality rates shown correspond to chicks' cumulative mortality (egg to adulthood). Offspring resulting from partial-mixed crosses (with a red heterozygous male) suffered similar mortality rates to offspring produced in mixed pairing (not shown).

In the Gouldian finch, all three components of both pre- and post-zygotic reproductive barriers seem to be Z-linked traits: (i) morph genetic expression (Kim et al. 2016), (ii) female assortative morph recognition (Pryke 2010) and (iii) inter-morph genetic incompatibility (Pryke & Griffith 2009; Pryke & Griffith 2009). Altogether, these factors indicate that sympatric morphs may represent a snap-shot of the early stages of the speciation-hybridization process. Additionally, strong evidence for disruptive selection on colour morphs has been put forward, revealing that predators selectively targeted black morphs on sandy backgrounds and red birds on rocky grey backgrounds. This was proposed as a mechanism maintaining populations' polymorphism (Hatton 2013)

Despite making important contributions, recent works using linkage mapping were only able to map a candidate region responsible for red/black head colour, in the Gouldian finch (Kim et al. 2016). This genomic region is quite broad, containing dozens of genes, leaving the genetic basis of head colour polymorphism in this species unresolved. We hypothesize that differences in gene expression levels may be in the origin of alternative colour morphs and that variation between red/yellow colouration might be explained by the same autosomal gene implicated in zebra finches and canaries' carotenoid colour variation (Lopes et al. 2016; Mundy et al. 2016). Since it has been suggested that the *Red locus* might be associated with pre- and postzygotic incompatibilities in Gouldian finches and that genes involved in colouration often regulate other metabolic processes, identifying a causal gene responsible for colour polymorphisms would help to investigate the causes behind intermorph incompatibility. Furthermore, it would also enable a more thorough study of the genetic architecture and pleiotropic effects of alternative variants, involved in assortative mating. Ultimately, we

hope our work contributes to a better understanding of how complete reproductive isolation is prevented between interbreeding morphs; and that it increments our knowledge respecting the genetics of carotenoid colour variation.

2. Materials and Methods

2.1. Sampling

Gouldian finches were acquired from Portuguese licensed breeders and kept in experimental facilities under the same conditions, in compliance with national and international regulations for the maintenance of live birds in captivity (*FELASA*, Federation of European Laboratory Animal Science Associations). A total of 8 finches from 3 different colour morphs (red, black and yellow) were sampled, as these colours are representative of wild-type extant morphs (Southern 1945). Only male subjects were obtained to avoid sex-dependent expression profiles (Ranz et al. 2003).

Following acquisition, birds were held for approximately 10-15 days. During this time, feather regeneration was induced by plucking small feather patches from the subjects' faces. New feathers were allowed to grow for ~10 days prior to tissue excision. In agreement with the *IACUC* (The Institutional Animal Care and Use Committee: American Association for Laboratory Animal Science, Memphis) and *AVMSA* (American Veterinary Medical Association) subjects were euthanized in a two-step procedure: each bird was rendered unconscious followed by manual cervical dislocation. After sacrificed, scalp tissue samples were collected and immediately frozen in liquid nitrogen for later RNA extraction.

2.2. RNA extraction, library preparation and sequencing

Tissue samples were grinded and homogenized by impact and friction in a *Ball Mills*, to recover RNA from disrupted cells. Total RNA from skin samples was isolated and purified using the *RNeasy*® Mini Kit (*QIAGEN)*. An additional *RNase-Free DNase*® digestion step was performed for a more complete removal of contaminant DNA. Following isolation, initial estimates of RNA concentration and purity were accessed through the usual 260/280 and 260/230 absorbance ratios using *Qubit*® RNA BR assay

kit. By disregarding anything other than RNA, this method provides an accurate and highly selective way for the quantitation of high abundance RNA samples. To ensure that differential degradation of samples would not mislead our differential expression analysis further ahead, the RNA Integrity Number (RIN) was determined in an *Agilent 2200 TapeStation* system with a RNA *ScreenTape.*

All samples had a fair amount of RNA, but RIN scores below 8, which is not the ideal for library preparation, as it might show some extent of RNA degradation. In our case, due to the nature of the tissue (skin) it would be expected some sample degradation. Skin tissue inevitably carries dead cells from natural abrasion and keratinized appendices, which biases sample quality estimates. Likewise, the presence of molecular pigments often shifts the absorbance peaks to indicate sample contamination and/or degradation. Takin this fact into account, we chose to proceed to library preparation.

Illumina libraries were prepared, using 1µg of total RNA per sample, following the *TruSeq*® RNA Sample Preparation Kit v2 standard protocol. In short, polyA containing mRNA molecules were selectively enriched, by binding oligo dT attached magnetic beads. Then, enzymatically cleaved and primed RNA fragments were reverse transcribed into the first cDNA strand. A second strand of cDNA was further synthesized to generate double stranded cDNA, end repair was performed and 3' ends of the cDNA were adenylated. The latter to two steps allow adapter ligation while preventing fragments to stick to one another. Each double stranded cDNA was coupled to single indexed adapters, such that each library represented one individual subject. To amplify the amount of starting material, cDNA fragments were selectively enriched through a small number of PCR cycles. Final libraries were, then, quantified by qPCR according to *KAPA* library quantification kit (KAPA BIOSYSTEMS) and normalized to achieve optimal cluster densities across the flow cell. Finally, samples were pooled together in equal volumes and sequenced on a *HiSeq 1500* Illumina sequencer, producing 125bp paired-ends reads.

2.3. Sequence processing and reference mapping

Prior to genome mapping, samples were submitted to pre-processing steps to filter out poor quality data. *TruSeq* Illumina adapters and barcodes were removed from raw reads using *Cutadapt v1.7.119* (Martin 2011) and low base quality ends were trimmed with *Trimmomatic v0.3220* (Bolger et al. 2014). For optimal trimming, reads were scanned through a 4bp sliding window, establishing a *Phred* quality score of 15

(SLIDINGWINDOW=4:15). All reads smaller than 30bp were discarded. All samples were inspected with *FastQC v0.10.1* (Simon Andrews 2010) before and after trimming to evaluate the need for adjusted parameters.

Since the Gouldian finch reference genome is not yet available, reads were aligned against the Zebra finch genome, as this is the closest and most complete reference to our focal species. On that account, both the Zebra finch (*Taeniopygia guttata*) genome assembly version 3.2.4 and the corresponding transcriptome were downloaded from the NCBI Reference Sequence Database (*RefSeq*, https://www.ncbi.nlm.nih.gov/refseq/) and used as reference for mapping. The need for two reference types is explained below.

We resorted to *HISAT2* (Kim et al. 2015) for a highly efficient alignment to our genomic reference. Indeed, this implemented algorithm is the first of its kind to employ a hierarchical indexing strategy for spliced alignment. In addition to the usual whole-genome indexing, it constructs numerous local indexes for regions that collectively cover the genome. Also, by using a scheme based on the Burrows-Wheeler transformation, it significantly decreases memory requirements. Finally, the combination with different alignment algorithms designed to target specific RNA-seq problems, allows for high speed and quality, often exceeding other aligners in speed and accuracy.

Mapping to zebra finch transcriptome reference was handled by *Bowtie2* (Langmead & Salzberg 2012), also integrated into *HISAT*'s engine. One of the advantages standing out on this algorithm is its ability to cope with large genomes and supporting gapped, local, and paired-end alignment modes, enabling a flexible yet accurate alignment method.

2.4. Differential Gene Expression analysis

RNA sequencing enables the detection and quantification of expressed transcripts in biological samples, being applicable to novel transcript discovery, differential expression analyses and alternative splicing studies. When working with transcriptome data, read depth is used as a proxy for quantifying expression levels and compare it across samples. Due to considerable variation in coverage across transcripts, abundance is often expressed as FPKM (fragments per kilobase of exon per million reads mapped) which provides length and depth normalization to allow for comparisons within and between groups. We made use of two software packages *DESeq2* (Love et al. 2014) and *Cuffdiff2* (Trapnell et al. 2012) to detect possible divergent expression patterns between head colour phenotypes. By implementing a two-fold approach we

aimed to increase detection power as well as to capture as many differentially expressed transcripts as possible. Both methods follow a negative binomial distribution model, however, they differ in the type of data used to estimate expression levels. *DESeq2* is a count-based method, whereas *Cuffdiff2* utilizes a FPKM-based strategy. Therefore, it is expected that dissimilarities in differentially expressed genes emerge, as *Cuffdiff2* algorithm performs additional steps to tease apart changes in isoform- and gene-level abundances. In addition, *Cuffdiff2* requires a reference genome to annotate transcripts and *DESeq2* works directly with count-matrices independent from annotation, thus the need for two separate types of reference.

To produce relative abundance estimations, a *pileup script* incorporated in *BBmap*'s toolbox (http://sourceforge.net/projects/bbmap/) was applied to the BAM files resulting from *Bowtie2*'s mapping to zebra finch transcriptome. Retrieved read counts, were shaped into a counts-matrix that was fed into *DESeq2* package. For *Cuffdiff2*, we took advantage of our already existing genome annotation files, GFF and GTF, and quantified both isoform- and gene-level expression directly from *HISAT2*'s output BAM files (comprising genome mapped reads). This implied using *Cufflinks* (Trapnell et al. 2012) to assemble the reads from each sample into a sample-specific transcriptome and, then merge them into a master transcriptome with *Cuffmerge* (Trapnell et al. 2012). Finally, expression levels based on FPKM counts were compared between conditions.

All possible pairwise comparisons were tested in both methodologies (i.e. "Red vs. Black"; "Red vs. Yellow" and "Black vs. Yellow"), including two additional contrasts ("Z-linked vs. Autosomal" and "Melanin vs. Carotenoid") that group samples according to inheritance mode and shared pathway, respectively. Grouping samples in this fashion allows phenotype-based comparisons, but also to explore putative genes associated to specific pathways, hence gaining a better perception for biological interpretation.

After applying a Benjamin-Hochberg correction for multiple-testing, genes for which FDR < 0.05 (False Discovery Rate), were considered differentially expressed. To increase the power of detection, the two resulting lists of differentially expressed genes were combined and scanned for obvious candidates involved either in pigment biosynthesis - for the melanic morph - or pigment metabolism and deposition – for carotenoid related morphs. Lastly, using the web-based *Ensembl BioMart* (Kinsella et al. 2011) data mining tool, Gene Ontology terms were enlisted whenever possible, for each of the differentially expressed components to investigate the presence of non-obvious candidates.

2.5. Variant calling in High-Throughput Sequencing data

To identify locus-specific sequence polymorphisms (SNPs), in each sample and colour phenotype, BAM files were put through a series of *Picard* (http://broadinstitute.github.io/picard) pre-processing steps to make it suitable for variant calling using *GATK* tools (Mckenna et al. 2010). Such steps included adding read group information, sorting, marking duplicates and indexing.

Next, *GATK's SplitNCigarReads* tool identifies all N elements in the reference CIGAR, which correspond to splicing events, and compares it to the same elements in each sample, splitting reads into exons and hard-clipping overhanging sequences. This way, the number of false variants called due to inaccurate splicing events is reduced. There was no need for mapping quality reassignment, as *HISAT* already reports meaningful quality scores compatible with the *GATK* pipeline. Once the data had been pre-processed, the variant discovery process began, i.e. the process by which variable sites relative to the reference genome are identified and genotyped. Because some of the variation observed is caused by mapping and sequencing errors, the greatest challenge here is balancing the need for sensitivity (failing to identify real variants) vs. specificity (failing to reject artifacts). For this reason, GATK has decomposed the analysis into separate steps: variant call and variant filtering. As recommended, we performed each step per-sample.

GATK's HaplotypeCaller tool was used in the GVCF mode, to perform per-sample variant calling, via local de-novo assembly of haplotypes. Once a list of plausible haplotypes is produced, a Smith-Waterman alignment (SWA) of each haplotype to the original reference is performed to reconstruct a CIGAR string for the haplotype. This yields the putative variation sites. For each potentially variant site, the program applies Bayes' rule to calculate the likelihoods of each genotype, given the read data observed for each sample. The most likely genotype is then assigned to the sample. In a second step, all sample files in the cohort are submitted to a joint genotyping analysis using the package's *GenotypeGVCFs* tool to create a raw SNP and Indel record. By performing variant discovery in a way that includes joint genotyping analysis, it is possible to leverage population-wide information from a cohort of multiple samples empowering the detection of variants with great sensitivity and genotyping with accuracy. Moreover, using GVCF mode we profit from the ability to re-run the population-level genotyping analysis at any time as the cohort grows avoiding computational costs.

To end, as *Variant Recalibration* (VQSR) is not yet supported for RNA-seq experiments, we hard-filtered variants manually applying specific thresholds suggested by *GATK* developers. Clusters of minimum 3 SNPs within a window of 35bp were filtered

based on Fisher Strand values (FS>30), which measures the *Phred*-scaled probability for strand bias at the site, and Quality by Depth (QD<2) intended to normalize variant quality to prevent inflation from deep coverage.

2.6. Genetic differentiation estimates

At this point, we were mainly interested in genetic variation caused by bi-allelic SNP data and as such we discarded Indels and multi-allelic SNPs from further analyses.

To identify regions of the genome that were highly differentiated between head colour morphs, allele frequency differentiation was estimated across scaffolds using the fixation index (F_{ST}), implemented by *VCFtools* (Danecek et al. 2011). In short, F_{ST} values were averaged across each scaffold using a sliding-window approach, with three window sizes: 50kb, 100kb and 200kb. Step size between windows was chosen as 25% of the window size. Preference was given to larger window-sizes to ensure averaged values would not be biased by sampling variance, rather than phenotype-associated variation. This approach aims to target regions of high genetic differentiation, irrespective of any assumption about demographic history. Also, because *VCFtools* applies Weir and Cockerham's estimator (Weir and Cockerham 1984), no missing genotype data was allowed, since in combination with small sample size it can potentially bias F_{ST} calculations (Bhatia et al. 2013).

Based on previous knowledge of breeding experiments, we admitted the traits under scope followed an independent mode of segregation, thus being Mendelian traits (Lewis et al. 2005). Hence, we grouped the subjects accordingly:

i) Yellow (homozygous recessive) **vs. Others** (either homozygous dominant or heterozygous) with respect to the autosomal locus

ii) Black (homozygous recessive) **vs. Others** (either homozygous dominant or heterozygous) concerning the Z-linked locus

We defined candidate genomic regions under selection for each pairwise-comparison as regions with window-based F_{ST} estimates above the 99.99th percentile of the empirical distribution (F_{ST} >0.30). SNPs found within consecutive windows were also considered, as long as these windows fell above the 99.9th percentile and allele frequency differences were higher than 0.60. All candidate regions were visually inspected on *IGV v2.3.81* to identify a genotypic pattern associated with phenotypic differences.

2.7. SNP Validation

For each pairwise-contrast, information from the resulting SNP catalogue and the list of differentially expressed genes was cross-checked to identify strong candidate *loci* influencing gene expression. Those candidates were required to i) be listed as differentially expressed; ii) be highly differentiated (F_{ST}>0.60) and iii) matching genotype/phenotype segregation mode. Due to the nature of our dataset, regions surrounding each SNP were manually inspected on the zebra finch genome to avoid designing primers overlapping intronic regions, which could affect PCR efficiency. Overall, 16 individuals (including the initial 8 referred to in *2.1*) from three head colour morphs (Red, n=6; Black, n=5 and Yellow, n=5) were genotyped. SNP validation and association with phenotypes was carried out by Sanger sequencing on a Biosystems™ 3130XL Sanger Sequencer, available at CIBIO's facilities (Porto, Portugal). Genotype calls' correction was done manually by means of *BioEdit v7.2.5* (Hall 1999).

2.8. SNP Functional Annotation and variant impact

To functionally annotate variants, we used the genetic variant annotation and effect prediction toolbox *SnpEff* (Cingolani et al. 2012). We focused strictly on SNP variation that was found to be highly differentiated between cohorts (allele frequency difference ≥0.60), and each variant was manually inspected. Specifically, we looked for variants of potential functional significance, such as nonsynonymous, frame-shift, STOP, and splice site mutations. The impact of amino acid substitutions on the function of proteins was inferred using the Case/Control tool implemented in *SIFT v1.03* (Ng & Henikoff 2003). Here, one can define case and control groups based on segregation mode and posteriorly filter out variants not following a pattern of interest.

3. Results

3.1. Dataset overview and Quality control

Our RNA sequencing protocol yielded a total of ~400 million 125bp reads evenly distributed across samples, with the exception of one for which a larger number of reads was obtained (Table 2.1). Following pre-processing steps, at which point adapters and

barcodes are removed, the number of reads dropped to ~376 million, with sizes ranging between 30 and 112bp (Table 2.1).

Table 2.1: Number of RNA-seq reads per sample and head phenotype, before and after filtering

Phenotype	Sample Code	No. Reads raw data	No. Reads after filtering[a]
Yellow	GO01AR1	52,319,958	49,083,970
	GO02AR3	49,796,176	46,763,048
	GO03AR8	46,456,990	43,464,078
Black	GB01AR9	50,461,530	47,189,574
	GB02AR10	45,130,642	42,362,154
	GB03AR11	47,450,142	44,622,604
Red	GR01AR20	45,177,360	42,318,174
	GR02AR21	64,308,606	60,348,366
Total		401,101,404	376,151,968

[a] After removing adapters with Cutadapt and trimming with Trimmomatic

3.2. Mapping quality statistics

Reads passing quality control were, then, aligned to Zebra finch reference genome and transcriptome. Overall, a high percentage (~95%) of good quality reads were successfully mapped to zebra finch's transcriptome, with 77% being properly mapped (Table 2.2). Mapping reads to the reference genome resulted in ~40% of the initial reads being properly mapped (Table 2.2). As the reference transcriptome contains exclusively protein-coding genes, which typically show higher sequence conservation, this difference in the amount of reads mapping to different references was expected. Mapping quality scores (MQ) quantify the probability of a read being misplaced and it is measured in a scale of 0-60. For paired-end alignments, distance and strand orientation is also considered when quantifying mapping quality (Li et al. 2008). So, the higher this score is, the more unique is the alignment of a read and its mate.

Table 2.2: Summarized mapping statistics by phenotype and reference used

Head Colour	No. Samples	No. Reads[a]	Percentage of reads mapping[b]		% of reads *MQ ≥ 40[c]	% of reads *MQ ≥ 60[d]
			To Zebra Finch Transcriptome	To Zebra Finch Genome		
Yellow	3	139,311,096	94.94(77.59)	56.44(41.82)	55.52	49.03
Black	3	134,174,332	94.89(77.57)	56.04(41.49)	55.44	48.36
Red	2	102,666,540	94.84(77.11)	52.36(38.72)	56.32	44.81

[b]After removing adapters with cutadapt and trimming with Trimmomatic; [b]Properly mapped reads are shown in parentheses; [c]From Bowtie2 alignment to reference transcriptome; [d]From HISAT2 alignment to reference genome; *Different MQ thresholds were applied due to distinct software outputs

3.3. *Taking a first look at CYP2J homologs*

Compelled by previous studies, showing a cytochrome P450 oxygenase family member - *CYP2J19* - is implicated in carotenoid ketolation (Lopes et al. 2016; Mundy et al. 2016), we investigated read depth within the *CYP2J*-family genomic region, homologous to zebra finch chromosome 8 (24,500,000-24,600,000 bp).

Figure 2.4: RNA-seq read depth at the CYP2J-like genomic region homologous to zebra finch.
Panel **a** showing coverage within chromosome 8: 24,572,000-24,582,000 genomic interval (retrieved from reference annotation) and **b** displaying NW_002198990.1 unplaced scaffold, homologous to CYP2J2 members. Gene architecture is depicted in between panels. Coverage was group auto-scaled for proper comparison.

By visualizing read depth on IGV, we could perceive a perfect overlap of produced reads with *CYP2J* exons. However, no differences were apparent in *CYP2J* expression levels between Yellow and Red birds, within the genomic region homologous to zebra finch chromosome 8 (Figure 2.4, panel a). This drove us away from our hypothesis that, as in the canary and zebra finch, carotenoid conversion is catalysed by CYP2J ketolases. Notwithstanding, a second homology was found to zebra finch's NW_002198990.1 unplaced scaffold comprising the same *CYP2J* complex (Figure 2.4, panel b). Interestingly, read depth in this region seems to tell a different story. For red birds, we found reads mapping to the 5' end of the transcript and none were found on yellow birds. This might reflect a deletion between red and yellow haplotypes, consistent with the one described for zebra finches (Mundy et al. 2016), or alternative usage of gene isoforms in each morph. Overall, and due to the relatively low gene expression levels for this region, these results must be seen as tentative.

3.4. Differential Expression Analysis

Next, we performed genome-wide differential expression analysis to seek for candidate genes underlying head colour polymorphism. Multiple pairwise-comparisons were carried out by grouping data in alternative ways, either in simple phenotype-based contrasts or combining groups for increased sampling number and explore putative pathway switching genes. For a matter of simplicity, here we report findings for what we considered to be relevant contrasts in answering the questions raised by this thesis (Red vs Yellow and Melanin vs Carotenoids). Additional information on gene ontology for each of the following sections will be provided in supplementary materials (Appendix 1).

3.4.1. Gene expression profiling of Red vs Yellow birds

From a total of known 17,488 genes (encoded by 19,334 transcripts) present in zebra finch's reference genome and transcriptome, 439 were considered differentially expressed by DESeq2 and 136 were reported by Cuffdiff2 (FDR < 0.05). Of these, 69 genes were shared amongst algorithms (Table 2.3).

Table 2.3: Number of DEGs retrieved from each method and shared DEGs by relevant pairwise-contrast

	DESeq2	Cuffdiff2	Shared DEGs
Red vs Yellow	**439**	**136**	**69**
Melanin vs Carotenoids	**7**	**44**	**3**

A substantial proportion of shared differentially expressed genes, 54 in total, were upregulated in the Red colour morph, whilst less than 25% were downregulated. Even though the number of genes exclusive to each algorithm was largely different (Figure 2.5), the exact same shared differentially expressed genes were up and downregulated between them, maintaining an average Log_2 (FC) difference inferior to 0.1.

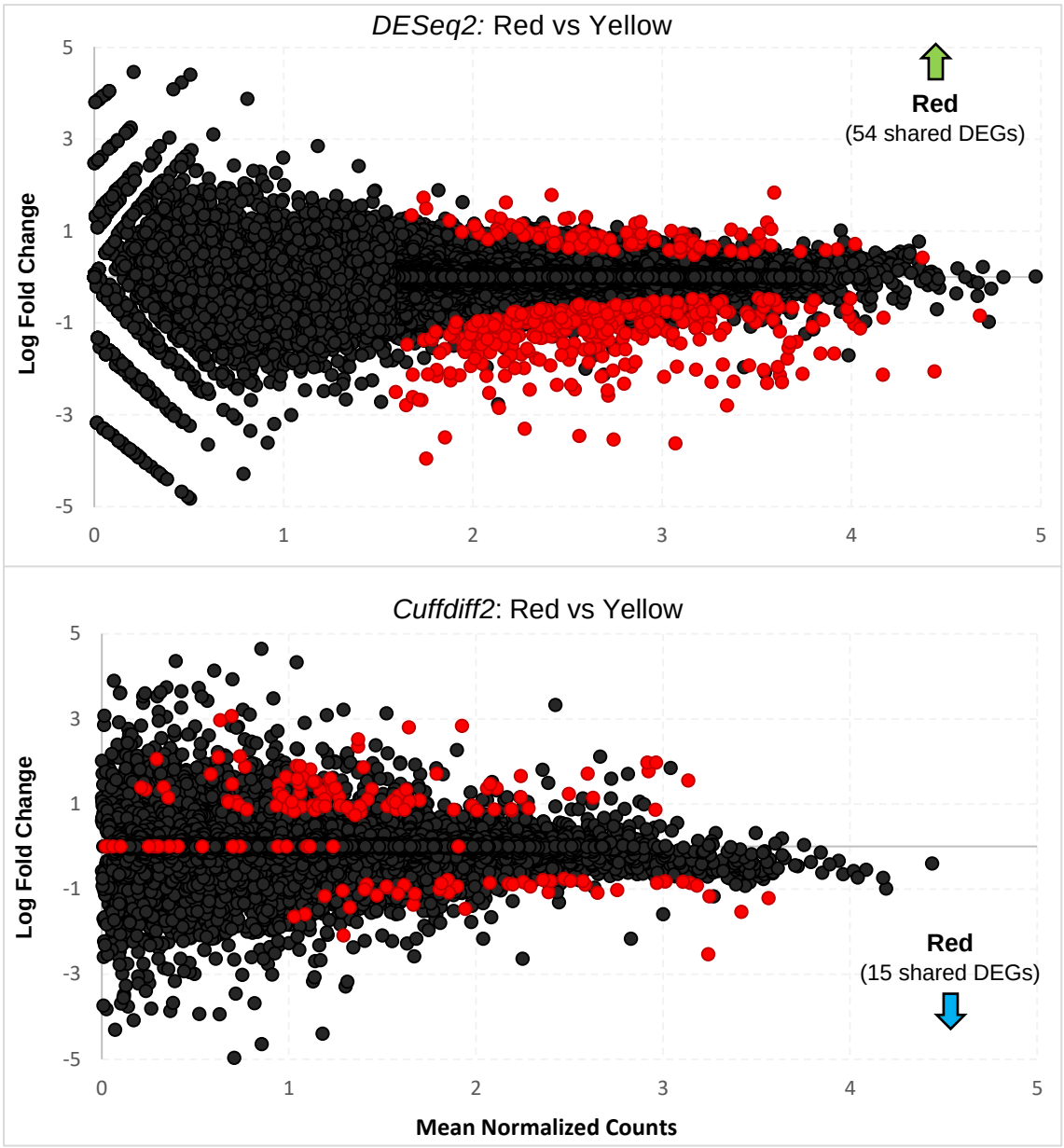


Figure 2.5: Smear plots resulting from differential expression analyses with DESeq2 (above) and Cuffdiff2 (below) for Red vs Yellow pairwise-contrast.
For each analytical method, mean normalized counts were plotted against $\log_2$Fold Change, depicting differentially expressed genes (FDR<0.05) **in red** over a background of non-differentially expressed genes. Arrows represent regulation of differentially expressed genes shared amongst algorithms – total number in brackets – in the specified morph.

No genes directly involved in carotenoid metabolism were found among differentially expressed genes, instead, 22% of the upregulated genes in the Red colour morph were involved in keratinization and/or skin development and approximately 14% were membrane transporters and/or solute carriers. Among the upregulated genes, Wnt-signalling pathway components such as *FZD5 (Frizzled Class Receptor 5)*, *WIF1* (Wnt Inhibitory Factor 1) and *GLI3 (GLI Family Zinc Finger 3)*, being the latter also a mediator of Sonic hedgehog (*Shh*) signalling (Borycki et al. 2000). All three belong to pathways

that are crucial for skin and melanocyte development (Logan & Nusse 2004; Park et al. 2014; Alvarez-Medina et al. 2007). Most downregulated genes between *Red* and *Yellow* morphs were transcription factors and regulators of cell proliferation and differentiation through several pathways. For instance, one downregulated gene -*ZNF638 (Zinc Finger Protein 638)* - is worthy of mention, since it is known to regulate cytochrome P450 enzymes through the action of peroxisome proliferators (Aninat et al. 2006). Importantly, *CYP2J*-like genes did not appear as differentially expressed in any of the contrasts.

3.4.2. Differential expression in Melanin vs Carotenoid phenotypes

When merging carotenoid-based phenotypes and contrasting it with black-headed birds, considerably fewer genes stood out as differentially expressed in either algorithm, when compared to Red vs Yellow. Still, DESeq2 retrieved only 7 genes characterizing this pathway switch, while Cuffdiff2 was shown less restrictive (Table 2.3 and Figure 2.6). Only three genes were common to both algorithms - *LONRF3 (*LON Peptidase N-Terminal Domain Ring Finger 3*)*, *COMT* (Catechol-O-Methyltransferase) and *ARO1 (*Penta functional AROM polypeptide 1) also known as *CYP19A1* and all of them were downregulated in melanic colour-morphs. The first, encodes a ubiquitin-dependent peptidase known to be part of protein catabolism (provided by Refseq. 2008), while the latter encodes an aromatase that catalyses the formation of sex steroid hormones (Simpson et al. 1994). Together with COMT, this aromatase is important for regulating oestrogen/androgen balance in the organism (Tworoger et al. 2004). In addition to this, *COMT* is also known to play a role in memory and learning due to inactivation of neurotransmitters and neuroactive drugs (Chen et al. 2011). We failed to find any obvious candidate involved in melanin biosynthesis.

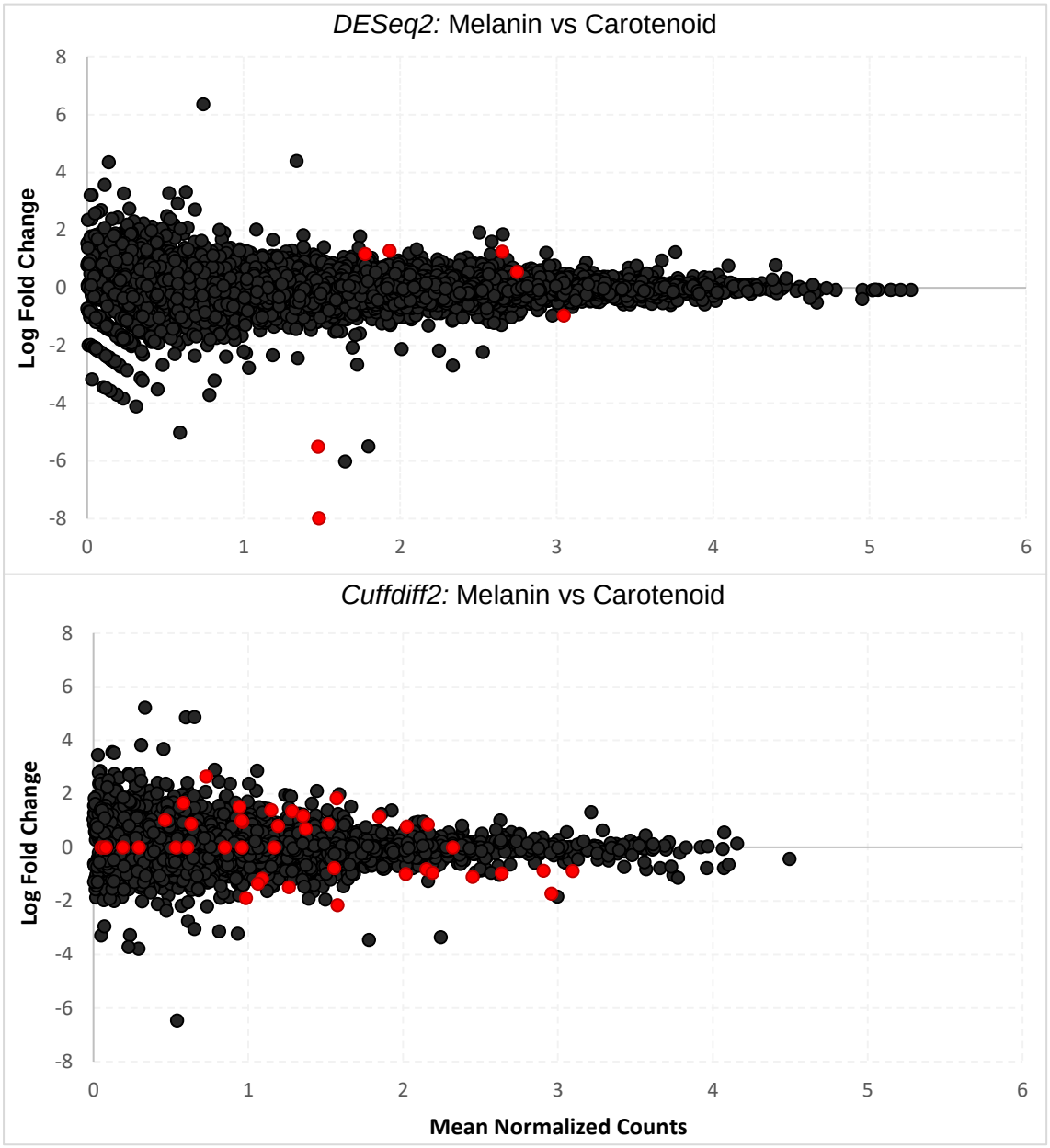


Figure 2.7: Smear plots resulting from differential expression analyses with DESeq2 (above) and Cuffdiff2 (below) for Melanin vs Carotenoid phenotypes.
For each analytical method, mean normalized counts were plotted against $\log_2$Fold Change, depicting differentially expressed genes (FDR<0.05) **in red** over a background of non-differentially expressed genes. Arrows represent regulation of differentially expressed genes shared amongst algorithms – total number in brackets – in the specified morph.

3.5. Association mapping of head colour polymorphisms

To further investigate chromosomal regions of high genetic differentiation that may underlie these unmistakable plumage traits, allele frequency differences were averaged, on a sliding-window approach, between morphs of interest (see *Materials and Methods* for a more detailed description).

3.5.1. Selective sweep mapping of the Yellow locus

In total, 69,581 polymorphic sites (SNPs) were extracted when comparing yellow headed and the remaining birds. Average measures of differentiation across all scaffolds were low (Mean F_{ST} = 0.038) when considering other domestic bird species for which similar data is available (Lopes et al. 2016; Long et al. 2017). This low average differentiation is conducive for identifying regions of high differentiation that might underlie the head colour polymorphism.

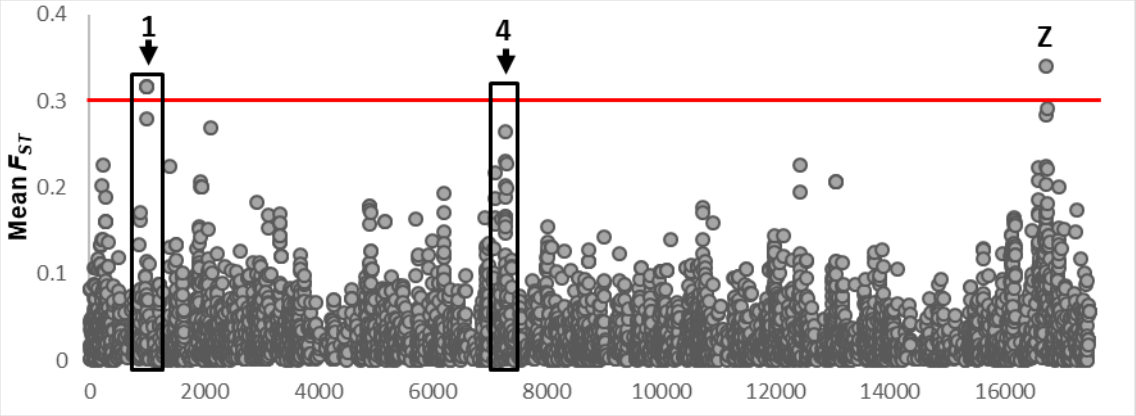


Figure 2.7: **Transcriptome-wide F_{ST} scan, between Yellow and Non-Yellow Gouldian finches, using RNA sequencing data.**
Dots represent F_{ST} estimates averaged over 100kb windows iterated in 25kb steps. Also depicted is the 99.9th percentile of the empirical distribution (**red line**). **Black rectangles** signal the position of top windows and respective chromosomes.

The clear majority of windows above the 99.9th empirical percentile (F_{ST} > 0.30) mapped to two genomic regions (Figure 2.7): one homologous to zebra finch's chromosome 1 (~80,335,000 – 80,718,000 bp) and the other located on zebra finch's chromosome 4 (~62,000,000 – 62,088,000 bp). Based on our analyses, we still had no indication that the most plausible candidate genes, *CYP2J*-like were involved in alternative colour morphs in the Gouldian finch. Indeed, none of the top windows mapped to scaffolds homologous to zebra finch's chromosome 8. Furthermore, all highly differentiated SNPs (N = 16; F_{ST} > 0.60) within chromosome 8 fell outside *CYP2J*-like known coordinates by more than 1Mb.

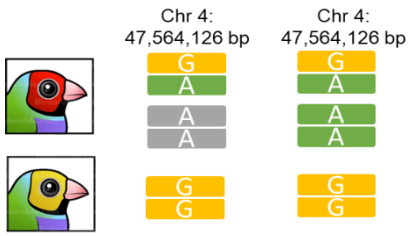


Figure 2.8: **The *Yellow* associated genotype on *NOP14* coding sequence.**
Perfect association located on zebra finch's chromosome 4: 62,074,784 - 62,074,828. Colours were chosen according to base-specific fluorophores. Non-polymorphic bases are indicated in grey. Distance not to scale. Black headed birds (not shown) share genotype with red headed birds.

Most of the highly divergent polymorphic sites clustered within the above-mentioned chromosome 4. Of these, only 2 SNPs were congruent with the expected genotypic pattern: yellow birds being homozygous for the alternative allele and non-yellow birds heterozygous or homozygous for the reference allele (Figure 2.8). These polymorphisms fell within the coding sequence of the *NOP14* gene, encoding a protein with no known function related to vitamins, lipids, or any other process associated with carotenoid metabolism.

3.5.2. Selective sweep mapping of the Red locus

With regards to the *Red locus* association mapping, we found that the highest level of differentiation was accumulated in the avian sexual chromosome (Mean F_{ST} = 0.072) by contrasting melanic and non-melanic birds. Also, approx. 30% of the whole 76,718 SNPs found were Z-linked. The previously mapped Z-linked interval (Kim et al. 2016) spanning ~10Mb, was taken as a candidate window for strong positive selection on *Black*/*Red* phenotypes.

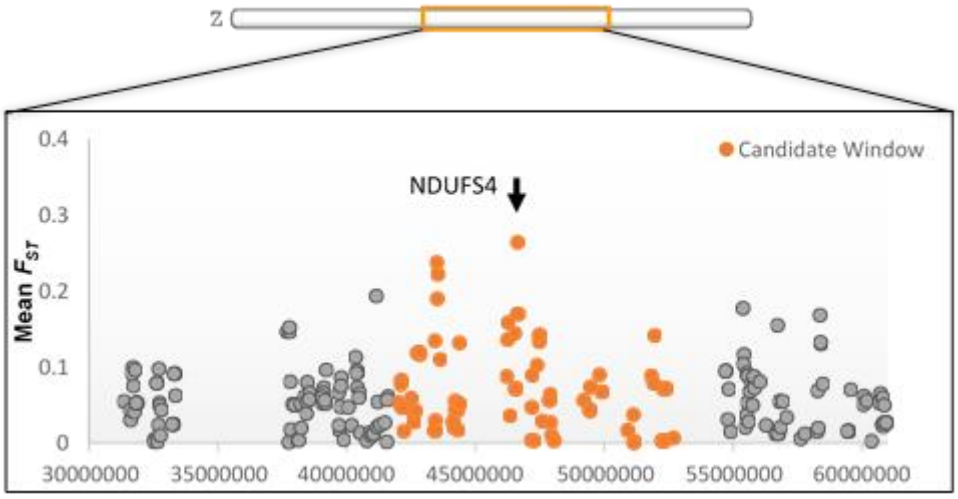


Figure 2.9: Transcriptome-wide F_{ST} scan, between Melanic and Non-melanic Gouldian finches, using RNA sequencing data.
Points represent F_{ST} estimates averaged over 100kb windows iterated in 25kb steps, along the Z chromosome. Relative position of the candidate window is given **in orange**. The **Black arrow** signals the position of our top window and corresponding gene.

Our sliding-window analyses revealed a series of genomic regions above the 99.9th percentile of the empirical distribution (F_{ST} > 0.30), including one within the ~10Mb candidate window (Figure 2.9). Since it had been shown that no other genomic region was linked to the *Red* locus, we focused our efforts exclusively on the Z chromosome.

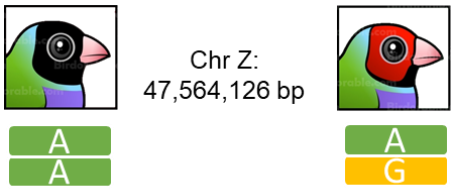


Figure 2.10: The *Red locus* associated genotype on IL6ST coding sequence.
Perfect association located within our candidate window, homologous to zebra finch's chromosome Z: 47,564,100 - 47,564,170. Colours were chosen according to base-specific fluorophores. Non-polymorphic bases are indicated in grey. The two closest polymorphisms are also shown. Distance not to scale. Yellow headed birds (not shown) share the genotypic pattern with red headed birds.

Despite its large range, only 2 highly differentiated SNPs (F_{ST} > 0.60) stood out from the candidate region, falling either on *NDUFS4* (NADH Ubiquinone Oxidoreductase Subunit S4; ~46,636,200 – 46,683,735) or *IL31RA* (Interleukin 31 Receptor A; ~47,504,000 – 47,534,408) homologous genes. Perfect genotype/phenotype association was found for variants on different gene, *IL6ST* (Interleukin 6 Signal Transducer; ~47,554,477 – 47,582,109; Figure 2.10), but not its paralog *IL31RA* or any other gene within the defined candidate window. However, estimates related to this *locus* did not reach our threshold to be considered a strong candidate in explaining the emergence of a phenotypic alternative (F_{ST} < 0.45), nor was the window containing it included in the 99.9th percentile. In summary, we failed to find any strong candidate gene or mutation within the candidate region that might be responsible for the *Red/Black* head phenotype.

3.6. SNP genotyping and Variant annotation

Considering the requirements for SNP confirmation, three variants were genotyped in a larger cohort of samples: two of them located on chromosome 1A: within *GAS2L3* (Growth Arrest Specific 2 Like 3, SNP: 46,733,812) and *WIF1* (SNP: 33,224,127) coding sequences and another one located on chromosome 4 (SNP: 62,074,784) within *NOP14* coding sequence. All amplicons synthesized were ~300bp long. We successfully amplified SNPs on chromosome 4 for all 16 samples of 3 colour morphs and found that genotypes extracted with our bioinformatics pipeline were correctly assigned to all 8 initial samples, in which yellow headed birds carried the alternative form in homozygosity -G/G- and the remaining phenotypes carried two copies of the reference allele -A/A- or were heterozygous -A/G (Figure 2.11).

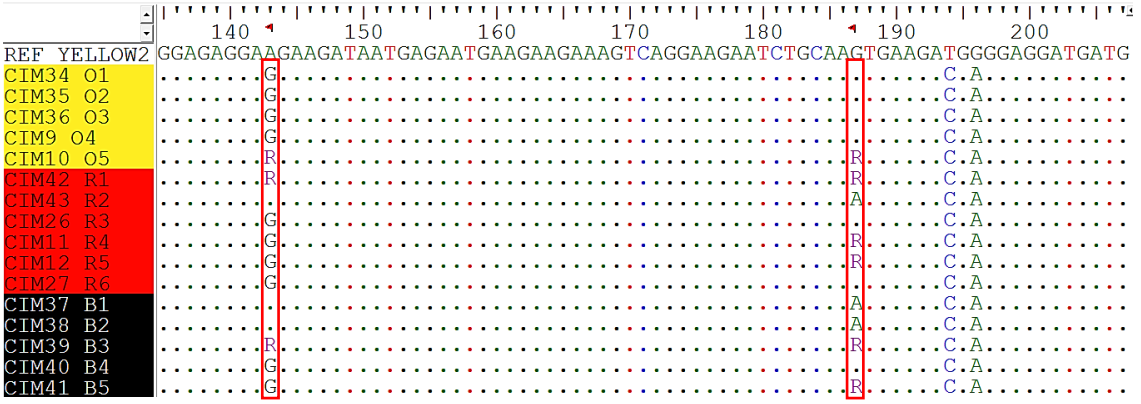


Figure 2.8: NOP14 allelic pattern is not related to head plumage colouration.
Comparison of amplified fragments across samples (N=16) in two polymorphic sites, within NOP14 coding sequence. Samples were highlighted according to head plumage colouration. Samples included in bioinformatic analyses were CIM34 to 36 for yellow morphs, CIM42,43 for red morphs and CIM37-39 for black morphs. Ambiguous bases symbolize heterozygous sites (R = A or G).

Regardless, when we extended this to other samples, we found no significant association between genotypes and plumage colour (*p > 0.210, Fisher's exact test*). An additional polymorphic site was found within the same ~300bp fragment, but its association to phenotype remained non-significant (*p > 0.05, Fisher's exact test).* Even though all considered SNPs either failed to be sequenced for all samples or could not unambiguously distinguish between head phenotypes, variant impact was evaluated on *SIFT*. None of the variants had a functional impact on the respective translated proteins, being classified as synonymous, or intergenic, mutations (Table 2.4).

Table 2.4: Variant annotation at candidate polymorphic sites

Gene	Position (bp)	Reference	Alternative	Annotation
WIF1	33,224,127	C	G	Synonymous
GAS2L3	46,733812	C	T	Intergenic
NOP14	62,074,784	A	G	Intergenic
NOP14	62,074,828	G	A	Intergenic

Chapter III

Novel insights into structural colour variation in the Gouldian finch

1. Introduction

The study of different colour-producing mechanisms has always been an appealing subject to avian biologists, one that yielded a considerable record of research towards understanding the physics and evolution of colouration (Hill & McGraw 2006). Still, our knowledge on the genetic architecture and especially the mechanisms controlling for variation of these traits is lacking, which is quite surprising given the millennial association of humans with bird domestication (Blechman 2007; Birkhead 2014; Zhang et al. 2017). Indeed, the work recently developed in a variety of captive breeding birds has had a substantial impact on melanin-related phenotypic variation (Price 2002; Badyaev 2006; Kerje et al. 2004; Gunnarsson et al. 2011; Dorshorst et al. 2011; Domyan et al. 2015). Notwithstanding, evidence on the genetic mechanisms controlling for variation in carotenoids has just now been put forward (Lopes et al. 2016; Mundy et al. 2016; Toomey et al. 2017) and no information is available on structural colouration, in either wild or captive birds.

Gouldian finches (*Erythrura gouldiae*) are exquisite Australian birds, exhibiting numerous vibrant colours organized in perfectly delineated patches (Figure 3.1). Their colour mosaic look has earned them a place among the most beautiful birds in the world and caught the interest of bird fanciers worldwide (Evans & Fidler 1986). Apart from iridescent colouration, gouldian finches seem to gather all types of colour-producing mechanisms, either pigmentary (head), non-iridescent structural (chest, tail coverts and thin blue line around the head) or a combination of both (on the back) making this species the ideal model to study colour evolution both in captive and wild populations (Figure 3.1). Aviculturists have successfully developed an extensive number of colour mutations throughout the decades, fully exploring gouldian finches' colour gamut and documenting inheritance patterns for each of the colour variants (Evans & Fidler 1986; Lewis et al.

2005). As a result, their practices paved the way to identify candidate *loci* underlying in phenotypic variation.

Figure 3.1: Chest colour variation and sexing in the Gouldian finch.
Gouldian finches exhibit 3 phenotypes controlled by 3 different alleles (A^P; A^p and A^w) within the same autosomal locus. Purple chest (**a**) is the original wildtype form in this species whereas white chest (**c)** is a mutation emergent from domestication. Lilac chest (**b**) is recessive to purple but dominant to white. **On the right**, females always appear duller than males, despite having the same genotype. Chest colour may be combined with any head and body colour.

In this chapter, we explore the genetic basis of chest colour variation in captive bred gouldian finches, hoping to get new insights on the molecular mechanisms controlling for variation in structural colouration.

Currently, three chest colours are recognized in domestic populations of the Gouldian finch (Figure 3.1) - purple, lilac and white - resulting from variation in a single autosomal *locus* with a dominant purple (A^P) and recessive lilac (A^p) alleles. A third allele controls for the expression of white chest (A^w), such that lilac is dominant to white and purple is dominant to both lilac and white alleles ($A^P > A^p > A^w$; Evans & Fidler 1986; Lewis et al. 2005). Interestingly, even though wildtype (purple) individuals may carry identical known genotypes, females are unable to fully express a purple chest, instead, they exhibit a duller tone (Figure 3.1, right). This may be an effect of alternative hormonal backgrounds, as shown in previous studies on hormone-dependent colouration (Haase & Schmedemann 1992; Lank et al. 1999; but see Kimball 2006) or sex-limited expressivity. A peculiar feature of chest colouration in this species, in addition to its independent genetic control, is the sharp transition between adjacent colour patches, suggesting a finely tuned regulation of protein expression in specific cells or cell lineages. In the gouldian finch, for instance, one may easily question why, within the same individual, melanin is expressed on its back and throat but not in the chest (Figure 3.1 see white chest mutation). Furthermore, perfectly delimited chest patches are widespread across Aves in varying colours (Figure 3.2), opening the possibility for its

use as a signal of high quality if the abrupt transition is maintained and the colour does not fade since birds in poor condition lose the transition and colour brilliance.

Figure 3.2: Chest-delimited colouration is a widespread phenotype in Aves and it can occur in a variety of forms. From top to bottom, left to right: White unpigmented feathers (Dipper *Cinclus cinclus*; Collared Inca *Coeligena torquata*; Toucan *Ramphastos toco*); **Eumelanin-based pigmentation** (Variegated Fairy Wren *Malurus lamberti*; Pied puffbird *Notharchus tectus*; Diamond firetail *Stagonopleura guttate*); **Carotenoid-based pigmentation** (Gray's malimbe *Malimbus nitens*; Northern parula *Setophaga Americana*; Olive-bellied sunbird *Cinnyris chloropygius*); **Structural colouration** (Azure-breasted pitta *Pitta steerii*; Gouldian finch *Erythrura gouldiae*; Lilac-breasted roller *Coracias caudatus*); **Pheomelanin-based pigmentation** (Green kingfisher *Chloroceryle Americana*; Lazuli bunting *Passerina amoena*; Bluethroat *Luscinia svecica*).

In this study, we hypothesize that regional loss of pigmentation, as in the white chest mutation, might be the result of multiple regulatory elements acting on the same coding sequence, so that pigment production is impaired in a restricted plumage patch. This is the case of the well-known *agouti* gene (*ASIP*), where alternative isoforms are independently regulated by either dorsal or ventral promoters, in mice (Vrieling et al. 1994), galliformes (Nadeau et al. 2008; Yoshihara et al. 2012) and fish (Cerdá-Reverter et al. 2005). Differential dorso-ventral gene expression of the same isoform has also been reported in mice (Manceau et al. 2011) and fish (Cal et al. 2017). On the other hand, *ASIP* is not the only possible candidate for colour variation. Transcription factors like *MITF* and its wide array of target genes (*TYR, MC1R, MLANA, SLC45A2*, etc.) are known to regulate lineage-specific pathways essential to melanin synthesis and allocation (Hoek et al. 2008; Cheli et al. 2010) which easily implies any mutation could lead to absence of pigmentation. We do not rule out the possibility for a fully structural alteration leading to loss of colour through incoherent light scattering (Dyck 1979), despite being documented that white feathers have evolved from structural blue and purple by derived loss of the underlying melanin-granules (Igic et al. 2016).

Here, we will tackle a fundamental question in bird colouration and colour evolution by revealing the genetic foundations for cell-specific loss of pigmentation in a non-model species. Finally, understanding which genes regulate the several existing pathways implicated in colour production and how they potentiate or constrain its variation is as alluring as it is crucial, especially in overlooked topics being structural colouration.

2. Materials and Methods

2.1. Sampling

As mentioned previously (see *Introduction*), the Gouldian finch bears numerous colours that segregate independently, as pieces to a larger puzzle. We benefited from this unique feature by planning our sampling so that we could use it to answer several different questions. As such, we saw fit to keep the same individuals from CHAPTER I and pursue the genetics of chest colouration in the current chapter, since the sampled individuals also segregated for different chest colouration.

From the initially secured 8 male finches, four exhibited purple breasts, as in the wild type form, and the remaining four had white breasts, a mutation emerged from the domestication process. The intent of this chapter was to explore the underlying processes associated with the phenotypic variation, rather than to understand less likely divergent expression patterns, ergo no tissue was collected from the chest area.

2.2. Genetic differentiation estimates or Selective sweep mapping

Since our sampling was balanced, we allowed for 25% missing genotype data in each cohort, meaning that 1 out of 4 individuals (maximum) was allowed not to have genetic information at that genomic position. In doing so, we felt we reached a good compromise between background noise (introduced by low precision when averaging values) and number of samples used. Population groups were established based on phenotype, i.e samples from purple chested subjects were grouped together, as were samples from white chested birds.

The magnitude of genetic differentiation was inferred across scaffolds via the fixation index (F_{ST}), by means of the *VCFtools* package (Danecek et al. 2011). Similarly to the previous chapter, a sliding-window approach was implemented, whereby F_{ST}

values were averaged over 100kb windows with a 25kb step across scaffolds. An additional 200kb window size was attempted in combination with a 50kb step. As previously, preference was given to larger window-sizes to ensure averaged values would not be biased by low sample sizes for each position. Candidate genomic regions under selection were considered as regions with averaged F_{ST} estimates above the 99.99th percentile of the empirical distribution (F_{ST} >0.30). We maintained the same selection criteria used in CHAPTER II to find SNPs within consecutive windows (included in the 99.99th percentile and F_{ST} >0.60). All candidate regions were visually inspected on *IGV v2.3.81* (Robinson et al. 2011) to identify valid genotype-phenotype associations.

2.3. SNP Validation

One of three strong candidate loci, located in two independent genomic regions, was genotyped. All three candidates fulfilled a two-fold selection criterion: i) high genetic differentiation (F_{ST}>0.60) between purple and white chest and ii) matching genotype/phenotype segregation mode. Overall, 16 individuals (the same from CHAPTER II) were genotyped (Purple, n=7 and White, n=9). SNP validation and association was carried out by Sanger sequencing on a Biosystems™ 3130XL Sanger Sequencer, available at CIBIO's facilities (Porto, Portugal). Haplotype confirmation and genotype calls' correction were done manually by means of *BioEdit v7.2.5* (Hall 1999).

2.4. SNP Functional Annotation and variant impact

Variants found were functionally annotated using the genetic variant annotation and effect prediction toolbox *SnpEff* (Cingolani et al. 2012). We focused strictly on SNP variation that was found to be highly differentiated between groups (allele frequency difference≥0.60), and each variant was manually inspected for confirmation. Again, we looked for variants of potential functional significance, such as nonsynonymous, frame-shift, STOP, and splice site mutations. The impact of amino acid substitutions was inferred using the Case/Control tool implemented in *SIFT v1.03* (Ng & Henikoff 2003), by defining case and control groups based on inheritance pattern and, posteriorly, filtering out variants not following a pattern of interest. To manually validate *SIFT* categorization, we retrieved the corresponding zebra finch sequence from *Ensembl*, artificially introduced the mutation and translated the full sequence.

3. Results

3.1. Dataset overview and Mapping quality

As explained before (see *Materials and Methods*), the same dataset was used in both chapters to explore rather different questions. Thus, the number of sequenced reads is common to both workflows, only differing in the way we grouped populations (Table 3.1). A total of ~400 million 125bp reads were initially generated and pre-processed, however approximately 376 million reads passed quality filters and were posteriorly mapped to zebra finch's references.

Table 3.1: Number of RNA-seq reads per sample and head phenotype, before and after filtering

Phenotype	Sample Code	No. Reads raw data	No. Reads after filtering[a]
Purple	GO01AR1	52,319,958	49,083,970
	GO03AR8	46,456,990	43,464,078
	GB03AR11	47,450,142	44,622,604
	GR01AR20	45,177,360	42,318,174
White	GO02AR3	49,796,176	46,763,048
	GB01AR9	50,461,530	47,189,574
	GB02AR10	45,130,642	42,362,154
	GR02AR21	64,308,606	60,348,366
Total		401,101,404	376,151,968

[c] After removing adapters with *cutadapt* and trimming with Trimmomatic

3.2. Selective sweep mapping of chest colour variation

To unravel genomic regions of heightened differentiation between purple and white chested gouldian finches, F_{ST} estimates were averaged across scaffolds using a sliding window approach. Altogether, 109,284 polymorphic sites (SNPs) were obtained by comparison of these two varieties. Average measures of differentiation across scaffolds were indicative of low population subdivision (F_{ST} = 0.048; Cormack et al. 1990).

Considering our sliding window analyses, all windows above the 99.9th percentile of the empirical distribution (F_{ST} > 0.30) were restricted to two main genomic locations (Figure 3.3): one homologous to zebra finch's chromosome 1 (~82250000 – 82550000 bp) and the other homologous to chromosome 5 (~22700000 – 22950000 bp).

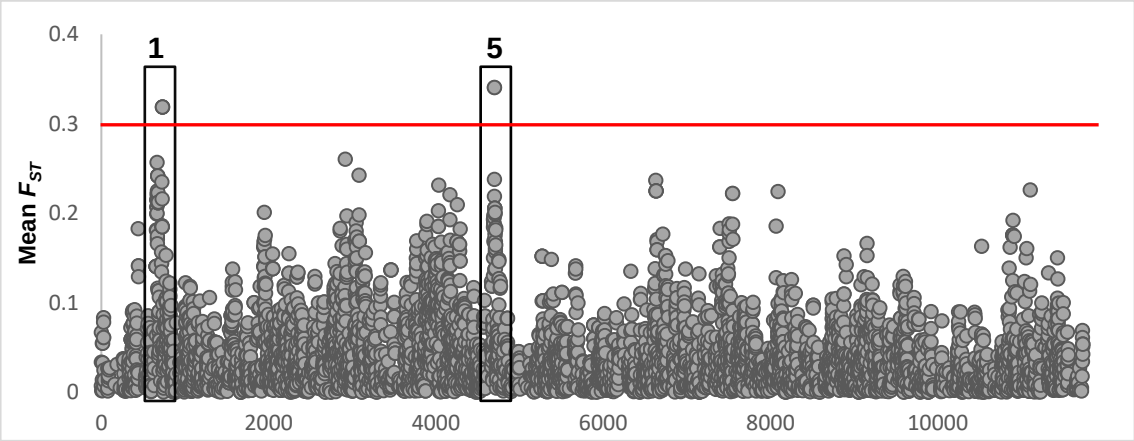


Figure 3.3: Transcriptome-wide F_{ST} scan, between Purple and White chested Gouldian finches, using RNA sequencing data.
Dots represent F_{ST} estimates averaged over 200kb windows iterated in 50kb steps. Also depicted is the 99.9th percentile of the empirical distribution (**red line**). **Black rectangles** signal the position of top windows and respective chromosomes.

Within the referred intervals, we identified two genes, *RAB38 (RAS* Oncogene Family Member*)* and *CHST1 (*Carbohydrate Sulfotransferase 1*)*, using the published annotation for zebra finch's genome and homology search tools. The first, on chromosome 1 was recently recognized as a tissue specific protein involved in melanosome biogenesis and trafficking, thus causing pigmentation defects when functionally impaired (Ohbayashi & Fukuda 2012). The latter, is not comprehensively described, yet, it encodes a sulfotransferase essential to keratin metabolism (provided by Genecards© v4.5.0, Weizmann Institute of Science 1996-2017). Interestingly, a third gene came to our attention when investigating the content of our candidate regions, one encoding the enzyme Tyrosinase (TYR), a widely recognized protein for its control of melanin synthesis. This gene was located on chromosome 1 approx. 400kb upstream from *RAB38* (~82001752 - 82044800 bp). Since variation in this gene is linked to several pigmentation disorders (Oetting 2000), mainly alternative forms of albinism, it was included as a strong candidate worthy of assessment.

3.3. SNP genotyping and Variant annotation

We genotyped a set of highly differentiated variants (F_{ST} > 0.60) within *RAB38* coding sequence, to confirm its association to chest phenotype. Amplicons produced were ~500bp long and were tested for a larger group of samples (N = 16). For all these variants, we found significant genotype/phenotype association (*p < 0.0015, Fisher's exact test*), with purple chested birds being consistently heterozygous and all white chested birds carrying two copies of the same mutant allele (Figure 3.4).

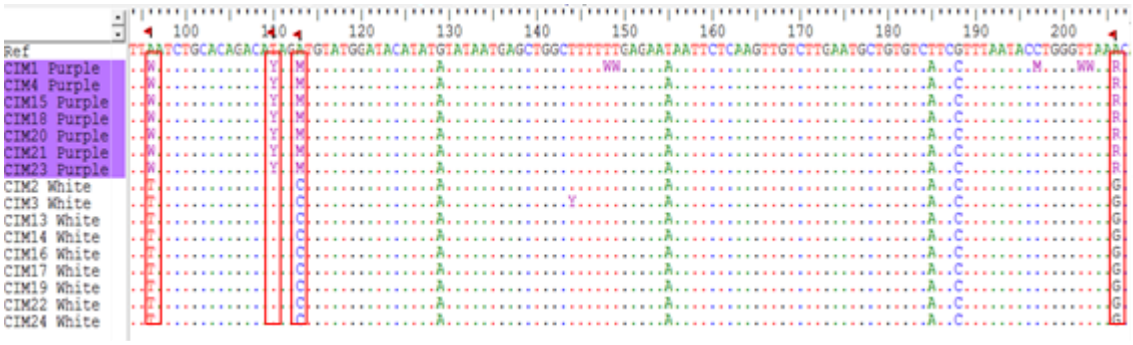


Figure 3.4: *RAB38* allele frequency pattern corroborates its association to chest colour variation.
Comparison of amplified fragments across samples (N=16) in four polymorphic sites, within *RAB38* coding sequence. Samples were highlighted according to chest colouration. Samples included in bioinformatic analyses were CIM15, 18, 20 and 21 for purple morphs and CIM13, 14, 19 and 22 for white morphs. Ambiguous bases symbolize heterozygous sites (W=A or T; Y=C or T; M=A or C; R = A or G).

All variants from the 3 candidates (*TYR*, *RAB38* and *CHST1*) were submitted to impact annotation, regardless of whether they were genotyped or not. Most of these polymorphisms were considered intergenic mutations, including those genotyped (Table 3.2). Moreover, detected two missense mutations on *TYR* gene, more specifically placed on exon 1, which is the most important exon to this enzyme's well-functioning (Table 3.2).

Table 3.2: Variant annotation at candidate polymorphic sites for chest colour variation

Gene	Position (bp)	Reference	Alternative	Protein changes	Annotation
TYR	**82044692**	**G**	**C**	**c. 109C>G | p. Pro37Ala**	**Missense**
TYR	**82044787**	**G**	**C**	**c. 14C>G | p. Ala5Gly**	**Missense**
RAB38	82411396	C	A	N/A	Intergenic
RAB38	82414411	A	G	N/A	Intergenic
RAB38	82664892	G	A	N/A	Intergenic
RAB38	82665918	G	C	N/A	Intergenic
CHST1	22770817	T	C	N/A	Intergenic
CHST1	22770982	C	T	N/A	Intergenic
CHST1	22771240	A	C	N/A	Intergenic

Chapter IV

General Discussion & Future Prospects

1. The genetic basis of head colour polymorphism

Initially, we believed divergent expression patterns could be the main reason for the emergence of alternative head colour morphs in the Gouldian finch, whether they were carotenoid-based or melanic. We posed the question of whether variation between yellow/red head colouration in gouldian finches would be explained by the same genetic architecture underlying differences in yellow/red canary breeds (Lopes et al. 2016) and yellow/red billed zebra finches (Mundy et al. 2016). Since in these two independent studies it was demonstrated the role of CYP2J19 monooxygenase in carotenoid metabolism, namely converting yellow into red ketocarotenoids, we took this gene, homologous to the *CYP2J2-like* cluster in zebra finch, as our strongest candidate gene for carotenoid ketolases involved in this phenotypic trait.

With respect to the *Red* locus, our main objective was to narrow down, to few putative genes, the chromosomal region known to accommodate loci controlling for selective deposition of carotenoid or melanin pigments. Based on knowledge from previous studies we hypothesized that any good candidate would induce melanin expression while supressing carotenoid deposition and as such we looked for melanogenesis inducing candidate genes, falling within the abovementioned window.

CYP2J-like complex might explain the conversion of yellow dietary pigments into red pigments in Gouldian finches

We were unable to get deep sequencing depth in this specific region, homologous to zebra finch's *CYP2J-like* complex. This lack of proper read depth translated into our inability to statistically distinguish if differential expression patterns were clearly driving phenotypic variation or not. The most interesting result, however, lies on the apparent deletion in the 5' end of *CYP2J-like* transcripts only seen in yellow-headed birds. Despite the lack of statistical significance, this result was clear when this region was inspected visually. The lower depth in this region might be the result of insufficient sequencing effort or be explained by the time of sample collection.

Assuming the red/yellow phenotype maps, indeed, to the *CYP2J-like* locus and that the lack of expression specific to a few exons in yellow-headed birds results from a deletion, our results are in line with the findings of a previous work performed by Mundy et al. in zebra finches. In this work, two *CYP2J19* homologs were found adjacent to a

third distantly related locus, homologous to chicken's *CYP2J40* and consistent with the exact same genetic organization we observed in our raw expression data. Additionally, Mundy et al. 2016 demonstrated an association of the "*yellowbeak*" zebra finch phenotype with a full deletion in one portion of the complex, corresponding to the entire *CYP2J19A* locus. Since our preliminary analysis of the expression data for *CYP2J-like* complex yielded a similar result, this tentatively suggests that deletions in the *CYP2J* locus might underlie red/yellow phenotypes in gouldian finches. If this is the case, we might be facing a case of convergent evolution, whereby an identical mutation evolved independently in two closely related species – gouldian and zebra finches - leading to a similar phenotypic result.

In an identical independent study, Lopes et al. 2016 further demonstrated how *CYP2J19* was introgressed from red siskins into the canary genome to breed the red factor canary. The addition of this genomic region, clearly prompted the ability of birds to produce ketocarotenoids as evidenced by the extremely elevated expression of *CYP2J19* in red birds when comparing to yellow birds. Even if the basal architecture differs from that of the zebra finch, the absence of *CYP2J19* obviously hampers carotenoid ketolation in either species. Interestingly, both studies found *CYP2J19* to be only differentially expressed in the skin but not in other important tissues, such as the retina where it contributes to enhancement of colour discrimination. They concluded yellow/red differences in feather pigmentation are most likely due to cis-regulatory expression changes at *CYP2J19* locus. In this thesis, only skin tissue was used to perform differential expression analyses, however, we may also hypothesize that differences between yellow and red headed gouldian finches are *cis*-regulatory in origin, based on their selective allocation of red ketocarotenoids to the head. In fact, yellow carotenoids are present throughout the rest of the body both in red- and yellow-headed birds. Since carotenoids are known for their role in communication as an honest signal of individual quality, mate choice and resource acquisition in gouldian finches, it is possible this is an evolutionary response to sexual selection.

Why did we fail to pinpoint a strong candidate gene being differentially expressed?

Next, we performed a genome-wide differential gene expression analyses to identify candidate genes or altered pathways underlying head colour polymorphism. However, we failed to find any obvious candidates either implicated in carotenoid metabolism or melanin biosynthesis. Nonetheless, by contrasting red and yellow birds we found several genes known to be part of major developmental pathways regulating follicular growth in vertebrates (*FZD5, WIF1, GLI3* all known to take part in Wnt-signalling cascade either directly or mediating accessory pathways, such as Sonic Hedgehog

signalling; (Borycki et al. 2000; Alvarez-Medina et al. 2007). Interestingly, recent work shows that dietary supplementation of astaxanthin, a ketocarotenoid, blocks nuclear translocation of the transcription factor beta-catenin, therefore inhibiting the Wnt canonical signalling pathway (Kavitha et al. 2013). However, in our analysis of red vs yellow birds Wnt signalling components are upregulated, suggesting either that astaxanthin is no longer being expressed in gouldian finches skin or that in gouldian finches, astaxanthin does not inhibit this major pathway. Wnt components regulate innumerable biological processes, including cell proliferation and migration, follicular development, pigment cell mobilization and differentiation from embryonic neural crest (Logan & Nusse 2004; Liu et al. 2014).

None of the differentially expressed genes retrieved for melanin versus carotenoid-based phenotypes, were contained within the defined candidate interval previously identified using linkage mapping in controlled crosses (chromosome Z: 43,000,000 – 53,000,000 bp). This might be explained by several non-mutually exclusive factors. One potential explanation is that this phenotype is not explained by gene expression differences but, instead, by protein coding changes. Another alternative is that we may have missed, again, the precise time of sampling when the causal gene(s) were exerting their functions. Third, because we sampled adult individuals only, any phenotypical determination taking place during developmental stages would not be perceptible at time of the sampling, impairing our ability to pinpoint a candidate gene. Fourth, exchanging between largely different metabolic pathways, such as carotenoid metabolism and melanin biosynthesis, might involve the recruitment of ubiquitous pathways. Consequently, a wide variety of genes being only slightly differentially expressed could go undetected with our sample sizes.

Genome-wide differentiation analysis between interbreeding colour morphs

Head colour morphs in the Gouldian finch are independently segregated and go hand in hand with assortative mating. Unlike in many domesticated species, many Gouldian breeders allow birds to interbreed freely (personal communication) and there are no established breeds. This means that the population is suitable for association mapping. In this context, we performed a set of differentiation analyses, by calling SNPs using the RNA-seq data, to identify chromosomal regions highly differentiated between morphs that might underlie phenotypic variation. We note that if the causative loci are small and map to regulatory features, the probability of mapping causal genes is null, since RNA-seq datasets are composed of translated coding sequences. Still, if the signature is strong enough one can easily find associated variation in nearby genomic regions.

Low average measures of differentiation computed across the genome suggest that the described genetic incompatibility between different colour morphs, that leads to inviable offspring (Pryke & Griffith 2009; Pryke & Griffith 2009), is not preventing the homogenization of allele frequencies due to gene flow. In fact, in a more recent study this apparent incompatibility was portrayed as a domestication artefact (Bolton et al. 2017) and many breeders of the domesticated variant also do not detect evidence for inviability in their stocks (personal communication). It has also been hypothesized that phenotypic differences between morphs, other than colouration, could be explained by an inversion affecting a large genomic segment, containing dozens to hundreds of genes. Such a mechanism has been implicated in several species with colour polymorphic systems that differ in a variety of morphological, behavioural, and biochemical traits (Lamichhaney, Fan, et al. 2015; Küpper et al. 2015; Zinzow-Kramer et al. 2015). However, we failed to find evidence for an extended genomic region of high genetic differentiation. Although we cannot discard the existence of inversions explaining colouration differences between morphs, if they exist they will certainly be of small size.

Genome-wide comparisons between yellow and non-yellow headed birds immediately showed two elements to take into consideration: one is the augmented level of genetic differentiation in the Z chromosome, due to its reduced effective population size (N_e), and second is the lack of F_{ST} peaks within the chromosome 8. Since our best candidate region responsible for the emergence of yellow head phenotypes - *CYP2J-like* complex - is harboured by chromosome 8, it was expectable to see increased differentiation levels in this genomic region. However, we do not see any signature of elevated differentiation in this region. Upon visual inspection and, as mentioned before, genes in this region have low sequencing depth, which prevented SNP calling in large segments that may contain a potential causative region. In addition, any structural modification, as suggested by the expression pattern in yellow birds, would immediately exclude the causative region from the analysis, since all yellow birds have missing data. This was probably the case as within the *CYP2J-like* region variants were called for the 3' end of the transcript but no variants were found in the 5' end of the same transcript.

We found that most of the highly differentiated SNPs to cluster in chromosome 4 with two of them exhibiting a genotypic pattern compatible to our expectations for a gene implicated in yellow phenotypes. However, by genotyping these variants on a large cohort of samples, we realized that they were false positives and yellow and red birds' genotypes did not confirm our expectations. These elevated peaks might be explained by the low sampling size in this contrast (2 versus 3 birds). Thus, these results do not exclude the *CYP2J-like* locus as the potential causative region for the red/yellow phenotype.

Despite working with a previously mapped candidate interval, which was helpful in guiding our expectations regarding differences between melanic and carotenoid-based head phenotypes, our sliding-window approach was not effective in finding plausible candidate genes. Nonetheless, it was valuable in the sense that significantly narrowed down the candidate genomic region from 10Mb to 1Mb, where we found two highly differentiated SNP within *NDUFS4* and *IL31RA* coding sequences. Because these variants did not follow a genotypic pattern compatible to the one predicted nor were directly involved in melanin biosynthesis, we did not investigate further associations by sequencing. Curiously, follistatin (*FST*) a gene involved in feather follicular growth and identified in yellow/black phenotypic variation in hybridizing warblers (Toews et al. 2016), is located immediately upstream from *NDUFS4,* leading us to believe that this might be a good candidate gene for further inspection. Its absence in differentiation analysis and lack of differential expression at the time of sampling may indicate, again, a time-dependency in phenotypic determination that we failed to grasp.

2. The genetic basis of site-specific loss of pigmentation

Structural mechanisms of colouration generate some of the most exquisite hues in birds' feathers. However, the genetic mechanisms that control these traits are largely unexplored. Structural colouration relies on a matrix of keratin and air that selectively reflects light to produce highly saturated hues. Since structural colouration depends on two main elements to achieve finely tuned colour, nanoscale organization and the presence of a melanin layer, one can easily cogitate that modifications in any of these components might result in loss of colouration. For instance, in his study of moulting stages, Dyck 1979 related the snow-white aspect of the Rock ptarmigan winter plumage to an alteration of feather nanostructure. Moulted winter feathers would contain larger and more variable air spaces, scattering all visible wavelengths and resulting in brilliant white colouration. In contrast, loss of the melanin layer in Manakins (*Lepidothrix sp.*) was associated with bright white feather patches (Igic et al. 2016).

In gouldian finches, the most perceptible structural colour patches are restricted to the chest and back areas, although we cannot exclude the possibility that other striking colour patches also have a structural component (e.g. the brightness of carotenoid-based colouration can be increased by structural modifications to the feather; Shawkey & Hill 2005). The white chest mutation in gouldian finches emerged from the domestication process and results in the loss of the structural purple colouration.

Based on the few cases in which similar colour variation was studied from a biophysical perspective and on previous notions on strict, localized genetic control of melanin-based phenotypes, we hypothesized that the white chest mutation in gouldian finches was a product of the loss of the melanin granules underlying the structural matrix. Using genome-wide F_{ST} scans comparing purple and white chested birds, we found two regions of elevated differentiation, each containing one putative gene involved in biological processes that might lead to loss of pigmentation or structural modifications. CHST1 is described as an intervenient protein in keratin metabolism, however, we are not aware of any scientific study linking this gene to keratin structural modifications or metabolism. RAB38 was recently recognized as a tissue specific protein involved in melanosome biogenesis and trafficking, thus causing pigmentation defects when its function is impaired (Loftus et al. 2002; Ohbayashi & Fukuda 2012). Furthermore, immunological studies have demonstrated a cell-specific expression pattern related to melanosome docking (Jäger et al. 2000). Finally, it was implicated in the chocolate mice dilution phenotype by decreasing efficiency of targeting TYRP1 to end-stage melanosomes, exhibiting an expression pattern similar to other melanocyte genes (*PMEL17, DCT and TYRP1*; Loftus et al. 2002). Interestingly, *RAB38* is located downstream from *TYR*, a gene commonly known for its control of melanin synthesis and directly linked to several forms of albinism. Despite perfect genotypic association for all the variants in *RAB38* coding sequence, our functional analysis only predicted variants with an impact on protein function in TYR, specifically two missense mutations. These mutations fall on tyrosinase's exon 1 overlapping the signalling peptide, which is known to harbour most of the mutations causing oculocutaneous albinism (Oetting & King 1999). Most mutations falling in this region are translated on tyrosinase knock outs (Oetting 2000). One of the protein coding changes falls in a position that is conserved across all bird genomes available and shows conservation all the way to medaka (data not shown), thus suggesting functional relevance.

Since the white chest patch appears as a precise delimited phenotype, we were expecting that it would be a product of *cis*-regulatory mutations acting locally to hamper melanin deposition and exceptionally this might not be the case in the gouldian finch. A Tyrosinase knock out would affect pigment deposition in the whole body. Instead, our results show indication of a conditional knock out, whereby a specific gene is inactivated in a given target tissue. It is also possible that tyrosinase impairment does not have a direct effect on pigment deposition but, rather, influences downstream genes involved in melanocyte mobilization to keratinocytes. These findings open up exciting new possibilities for melanin synthesis regulation, deposition and pattern formation.

3. Towards effective validation: two-folded analysis and dataset integration

Drawing conclusions from each analysis, *per se*, might be limited in the sense that independently they may not capture the entire variation in phenotypic traits or neglect to find subtle changes in genetic mechanisms. As such, it is now common to integrate multiple methods to increase the power to detect accurate and comprehensive associations. For that reason, we sought candidate *loci* common to both expression profiling and variant calling, unfortunately, with no great success. Problems with insufficient data or low sampling might have been factors affecting our efficiency to find reliable signatures of selection in interbreeding morphs, notwithstanding, pitfalls arising from the use of RNA sequencing data and the choice of a two-folded bioinformatic pipeline to characterize expression profiles may have significantly obstructed our findings, especially since many of our phenotypic traits under study are likely to arise from the interplay between regulatory features. Indeed, when looking at smear plots depicting expression patterns, it is clear that each algorithm considered genes to be differentially expressed in alternative ways, a product of distinct estimation and normalization methods (DESeq2 is based in raw read counts and logarithmic normalization whereas cuffdiff2 uses already normalized read estimates, FPKMs). It is hard to say which algorithm surpasses the other as their retrieved patterns seem almost complementary (*see Results*). As mentioned several times throughout this discussion, if the genetic machinery in question is time-dependent and the available time frame in which the gene is being expressed is narrow, a simple misguidance in stipulating the time of skin tissue harvest would largely hamper positive findings.

Future Prospects

To collect skin tissue samples, we performed a similar protocol to what has been done before, by our group, in canaries (sampling 10 days after feather plucking, Lopes et al. 2016). However, head feathers only accumulate pigments at the very tip and we might have caught the regeneration of feather follicles on a later stage, where pigmentation genes are already lowly expressed. Further experiments using qPCR and earlier feather follicle regeneration stages are now underway, that will help to determine the approximate time in which pigmentation genes are being expressed in feather follicles.

Establishing explicit genotype/phenotype associations using RNA-seq data proved to be an arduous task, one that will certainly be solved with increased sampling and extending the sequencing range to whole-genome approaches. Complementing this dataset with whole-genome resequencing (WGS) data and performing genome-wide association mapping (GWAS), would allow a more thorough investigation of the machinery behind head colour polymorphisms in the Gouldian finch and an in-depth understanding of important physiological processes, such as carotenoid metabolism and deposition in feathers and melanin metabolic triggers inducing selective melanin deposition. In our analyses, many unknown transcripts were captured that were differentially expressed in the skin. Because they were not annotated in Zebra finch's reference genome, they may have not been considered as candidates, which suggests the possibility for overlooking relevant pigmentation genes. Further studies would, therefore, profit from a *de novo* genome assembly of the Gouldian finch and careful annotation, currently ongoing in our lab.

Having found an unexpected novel mutation, affecting TYR protein function in white chested gouldian finches, stresses the need for a more thorough study of tyrosinase expression in the chest area, to confirm its involvement in pigment loss. Ongoing work using *in situ* hybridization, aims at characterizing the expression patterns of tyrosinase and other melanin-related genes, comparing distinct plumage colour patches in gouldian finches. A better characterization of phenotypes, namely through melanin quantification and distribution across the integument, is pressing to reveal which steps in the molecular cascade are being modified by the causative mutations. Also, supplementary enzymatic activity assays, on mutant TYR, would largely enhance our perception with respect to the specificity of substrate binding and the real impact of protein modification on kinetic behaviour of tyrosinase, as well as helping to untangle the interplay between its related proteins.

Finally, extending whole-genome analysis to other gouldian finch domestic breeds, such as the yellow-body and blue body mutations, will increase our notion regarding the genetic interplay producing all the astonishing colours displayed by domestic gouldian finches. Specifically, since yellow- and blue-body mutations represent melanin and carotenoid knock-outs, respectively, they may serve as good control groups for lack of pigmentation pathways and clarify the genetic machinery underlying the remaining colour mutations.

References

Alvarez-Medina, R. et al. 2007. Wnt canonical pathway restricts graded *Shh/Gli* patterning activity through the regulation of *Gli3* expression. *Development*, 135, pp.237–47, doi:10.1242/dev.012054.

Andrews, S. 2010. FastQC: a quality control tool for high throughput sequence data. Http://Www.Bioinformatics.Babraham.Ac.Uk/Projects/ Fastqc.

Aninat, C. et al. 2006. Expression of Cytochrome P450, Conjugating Enzymes and Nuclear Receptors in Human Hepatoma HepaRG Cells. *Drug Metabolism and Disposition*, 34, pp.75–83, doi:10.1124/dmd.105.006759.

Arnaiz-Villena, A. et al. 2009. Estrildinae Finches (Aves, Passeriformes) from Africa, South Asia and Australia: a Molecular Phylogeographic Study. *The Open Ornithology Journal*, 2, pp.29–36, doi:10.2174/1874453200902010029.

Badyaev, A. 2006. Colorful Phenotypes of Colorless Genotypes : Toward a New Evolutionary Synthesis of Color Displays. In *Bird Coloration, vol. 2: Function and evolution*. Harvard University Press, Cambridge, pp. 349–79.

Badyaev, A. & Young, R. 2004. Complexity and integration in sexual ornamentation: an example with carotenoid and melanin plumage pigmentation. *Journal of Evolutionary Biology*, 17, pp.1317–27, doi:10.1111/j.1420-9101.2004.00781.x.

Baião, P. & Parker, P. 2012. Evolution of the melanocortin-1 receptor (MC1R) in boobies and gannets (Aves, Suliformes). *Journal of Heredity*, 103, pp.322–29, doi:10.1093/jhered/esr151.

Bhatia, G. et al. 2013. Estimating and interpreting F_{ST}: the impact of rare variants. *Genome Research*, 23, pp.1514–21, doi:10.1101/gr.154831.113.

Birkhead, T. 2014. *The Red Canary: The Story of the First Genetically Engineered Animal*, Bloomsbury Publishing, New York.

Blechman, A. 2007. *Pigeons: The Fascinating Saga of the World's Most Revered and Reviled Bird*, University of Queensland Press, Queensland.

Bolger, A., Lohse, M. & Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30, pp.2114–20, doi:10.1093/bioinformatics/btu170.

Bolton, P. et al. 2017. The colour of paternity: extra-pair paternity in the wild Gouldian finch does not appear to be driven by genetic incompatibility between morphs. *Journal of Evolutionary Biology*, 30, pp.174–90, doi:10.1111/jeb.12997.

Bortolotti, G. et al. 2006. A complex plumage pattern as an honest social signal. *Animal Behaviour*, 72, pp.423–30, doi:10.1016/j.anbehav.2006.01.016.

Borycki, A., Brown, A. & Emerson, C. 2000. Shh and Wnt signaling pathways converge to control Gli gene activation in avian somites. *Development*, 127, pp.2075–87.

Bourgeois, Y. et al. 2017. A novel locus on chromosome 1 underlies the evolution of a melanic plumage polymorphism in a wild songbird. *Royal Society Open Science*, 4, pp.1608-05, doi:10.1098/rsos.160805.

Bradbury, J. & Vehrencamp, S. 2011. *Principles of Animal Communication,* 2nd ed., Sinauer Associates Inc., Massachusetts.

Brockmann, H. & Volker, O. 1934. The yellow colour of the canary's feathers and the occurrence of carotenoids in birds. *Hoppe-Seyler's Zeitschrift fur physiologische Chemie*, 224, pp.193–215.

Brush, A.H. 1990. Metabolism of carotenoid pigments in birds. *The FASEB Journal*, 4, pp.2969–77.

Burton, G.W. 1988. Antioxidant action of carotenoids. *Journal of Nutrition*, 119, pp.109–11.

Cal, L. et al. 2017. BAC Recombineering of the Agouti Loci from Spotted Gar and Zebrafish Reveals the Evolutionary Ancestry of Dorsal-Ventral Pigment Asymmetry in Fish. *Journal of Experimental Zoology. Part B, Molecular and Developmental Evolution*, 328, pp.697-708, doi:10.1002/jez.b.22748.

Cerdá-Reverter, J. et al. 2005. Gene structure of the goldfish agouti-signaling protein: A putative role in the dorsal-ventral pigment pattern of fish. *Endocrinology*, 146, pp.1597–610, doi:10.1210/en.2004-1346.

Cheli, Y. et al. 2010. Fifteen-year quest for microphthalmia-associated transcription factor target genes. *Pigment Cell and Melanoma Research*, 23, pp.27–40, doi:10.1111/j.1755-148X.2009.00653.x.

Chen, J. et al. 2011. Orientation and cellular distribution of membrane-bound catechol-O-methyltransferase in cortical neurons: Implications for drug

development. *Journal of Biological Chemistry*, 286, pp.34752–60, doi:10.1074/jbc.M111.262790.

Cingolani, P. et al. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w^{1118}; iso-2; iso-3. *Fly*, 6, pp.80–92, doi:10.4161/fly.19695.

Cooke, F. & Buckley, P. 1987. *Avian Genetics: A Population and Ecological Approach*. Academic Press Inc., London.

Cormack, R., Hartl, D. & Clark, A. 1990. *Principles of Population Genetics, 4th ed.* Sinauer Associates Inc., Massachusetts.

D'Alba, L., Kieffer, L. & Shawkey, M. 2012. Relative contributions of pigments and biophotonic nanostructures to natural color production: a case study in budgerigar (*Melopsittacus undulatus*) feathers. *Journal of Experimental Biology*, 215, pp.1272–77, doi:10.1242/jeb.064907.

Danecek, P. et al. 2011. The variant call format and VCFtools. *Bioinformatics*, 27, pp.2156–58, doi:10.1093/bioinformatics/btr330.

Domyan, E. et al. 2014. Epistatic and Combinatorial Effects of Pigmentary Gene Mutations in the Domestic Pigeon. *Current Biology*, 24, pp.459–64, doi:10.1016/j.cub.2014.01.020.

Dorshorst, B. et al. 2011. A complex genomic rearrangement involving the Endothelin 3 locus causes dermal hyperpigmentation in the chicken. *PLoS Genetics*, 7:e1002412, doi:10.1371/journal.pgen.1002412.

Dunn, P., Armenta, J. & Whittingham, L. 2015. Natural and sexual selection act on different axes of variation in avian plumage color. *Science Advances*, 1, pp 140-55, doi:10.1126/sciadv.1400155.

Dyck, J., 1979. Winter plumage of the Rock Ptarmigan: Structure of air-filled barbules and function of the white colour. *Dansk Ornithologisk Forenings Tidsskrift*, 73, pp.41–58.

Ellegren, H. 2014. Genome sequencing and population genomics in non-model organisms. *Trends in Ecology and Evolution*, 29, pp.51–63, doi:10.1016/j.tree.2013.09.008.

Eriksson, J. et al. 2008. Identification of the Yellow skin gene reveals a hybrid origin of the domestic chicken. *PLoS Genetics*, 4:e1000010.

Evans, S. & Fidler, M. 1986. *The Gouldian Finch, 1st ed.*, Blandford Press, London.

Fenoglio, S., Cucco, M. & Malacarne, G. 2002. The effect of a carotenoid-rich diet on immunocompetence and behavioural performances in Moorhen chicks. *Ethology Ecology & Evolution*, 14, pp.149–56, doi:10.1080/08927014.2002.9522753.

Ferreira, M. et al. 2017. The transcriptional landscape of seasonal coat colour moult in the snowshoe hare. *Molecular Ecology*, 26, pp.4173–85, doi:10.1111/mec.14177.

Fox, D. 1962. Carotenoids of the roseate spoonbill. *Comparative Biochemistry and Physiology*, 6, pp.305–10, doi:10.1016/0010-406X(62)90134-2.

Franklin, D. & Dostine, P. 2000. A note on the frequency and genetics of head colour morphs in the Gouldian Finch. *Emu*, 100, pp.236–39, doi:10.1071/MU00911.

Gilby, A., Pryke, S. & Griffith, S. 2009. The historical frequency of head-colour morphs in the Gouldian Finch (*Erythrura gouldiae*). *Emu*, 109, pp.222–29, doi:10.1071/MU09013.

Gomes, A., Sorenson, M. & Cardoso, G. 2016. Speciation is associated with changing ornamentation rather than stronger sexual selection. *Evolution*, 70, pp.2823–38, doi:10.1111/evo.13088.

González, S. et al. 2001. Real-Time Evidence of In Vivo Leukocyte Trafficking in Human Skin by Reflectance Confocal Microscopy. *Journal of Investigative Dermatology*, 117, pp.384–86, doi:10.1046/j.0022-202x.2001.01420.x.

Gouti, M. et al. 2017. A Gene Regulatory Network Balances Neural and Mesoderm Specification during Vertebrate Trunk Development. *Developmental Cell*, 41, p.243–61, doi:10.1016/j.devcel.2017.04.002.

Greene, E. et al. 2000. Disruptive sexual selection for plumage coloration in a passerine bird. *Nature*, 407, pp.1000–03, doi:10.1038/35039500.

Gunnarsson, U. et al. 2011. The Dark brown plumage color in chickens is caused by an 8.3-kb deletion upstream of *SOX10. Pigment Cell and Melanoma Research*, 24, pp.268–74, doi: 10.1111/j.1755-148X.2011.00825.x.

Haase, E. & Schmedemann, R. 1992. Dose-dependent effect of testosterone on the

induction of eclipse coloration in castrated wild mallard drakes (*Anas platyrhynchos L.*). *Canadian Journal of Zoology*, 70, pp.428–31, doi:10.1139/z92-065.

Haavie, J., Saetre, G. & Moum, T. 2000. Discrepancies in population differentiation at microsatellites, mitochondrial DNA and plumage colour in the pied flycatcher: inferring evolutionary processes. *Molecular Ecology*, 9, pp.1137–48, doi: 10.1046/j.1365-294x.2000.00988.x.

Haldane, J. 1922. Sex-ratio and unisexual sterility in hybrid animal. *Journal of Genetics*, 12, pp.101–09, doi:10.1007/BF02983075.

Hall, T. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 41, pp.95–98.

Hatton, L. 2013. *Predation as a mechanism maintaining polymorphism: Evidence for disruptive selection in the Gouldian finch*. PhD thesis. Australian National University.

Hill, G. & McGraw, K. 2006. *Bird Coloration, vol.1: Mechanisms and Measurements*, Harvard University Press, Cambridge.

Hillier, L. et al. 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, 432, pp.695–716, doi:10.1038/nature03154.

Hiragaki, T. et al. 2008. Recessive black Is Allelic to the yellow Plumage Locus in Japanese Quail and Associated With a Frameshift Deletion in the *ASIP* Gene. *Genetics*, 178, pp.771–75, doi:10.1534/genetics.107.077040.

Hoek, K. et al. 2008. Novel *MITF* targets identified using a two-step DNA microarray strategy. *Pigment Cell and Melanoma Research*, 21, pp.665–76, doi:10.1111/j.1755-148X.2008.00505.x.

Höglund, J. et al. 2017. Blood transcriptomes and *de novo* identification of candidate loci for mating success in lekking great snipe (*Gallinago media*). *Molecular Ecology*, 26, pp.3458–71, doi:10.1111/mec.14118.

Igic, B., D'Alba, L. & Shawkey, M. 2016. Manakins can produce iridescent and bright feather colours without melanosomes. *The Journal of Experimental Biology*, 219, pp.1851–59, doi:10.1242/jeb.137182.

Inouye, C. et al. 2001. Carotenoid pigments in male House Finch plumage in relation to age, subspecies, and ornamental coloration. *Auk*, 118, pp.900–15, doi:10.2307/4089841.

Jäger, D. et al. 2000. Serological cloning of a melanocyte rab guanosine 5'-triphosphate- binding protein and a chromosome condensation protein from a melanoma complementary DNA library. *Cancer Research*, 60, pp.3584–91.

Jarvis, E. et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, 346, pp.1320–31, doi:10.1126/science.1253451.

Kaiser, S. et al. 2017. A comparative assessment of SNP and microsatellite markers for assigning parentage in a socially monogamous bird. *Molecular Ecology Resources*, 17, pp.183–93, doi:10.1111/1755-0998.12589.

Kavitha, K. et al. 2013. Astaxanthin inhibits NF-κB and Wnt/β-catenin signaling pathways via inactivation of *Erk/MAPK* and *PI3K/Akt* to induce intrinsic apoptosis in a hamster model of oral cancer. *Biochimica et Biophysica Acta - General Subjects*, 1830, pp.4433–44, doi:10.1016/j.bbagen.2013.05.032.

Kerje, S. et al. 2004. The Dominant white, Dun and Smoky color variants in chicken are associated with insertion/deletion polymorphisms in the *PMEL17* gene. *Genetics*, 168, pp.1507–18, doi:10.1534/genetics.104.027995.

Kim, D., Langmead, B. & Salzberg, S. 2015. HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, 12, pp.357–60, doi:10.1038/nmeth.3317.

Kim, K., Griffith, S. & Burke, T. 2016. Linkage mapping of a polymorphic plumage locus associated with intermorph incompatibility in the Gouldian finch (*Erythrura gouldiae*). *Heredity*, 116, pp.409–16, doi:10.1038/hdy.2015.114.

Kimball, R. & David Ligon, J. 1999. Evolution of avian plumage dichromatism from a proximate perspective. *The American Naturalist*, 154, pp.182–93, doi:10.1086/303228.

Kimball, R. 2006. Hormonal control of coloration. In *Bird Coloration, vol. 1: Mechanisms and Measurements*. Harvard University Press, Cambridge, pp.431–68.

Kinsella, R. et al. 2011. Ensembl BioMarts: A hub for data retrieval across taxonomic

space. *Database*, vol. 2011, doi:10.1093/database/bar030.

Kritzler, H. 1943. Carotenoids in the Display and Eclipse Plumages of Bishop Birds. *Physiological Zoology*, 16, pp.241–55, doi:10.1086/physzool.16.3.30151697.

Küpper, C. et al. 2015. A supergene determines highly divergent male reproductive morphs in the Ruff. *Nature Genetics*, 48, pp.79–83, doi:10.1038/ng.3443.

Lafountain, A., Prum, R. & Frank, H. 2015. Diversity, physiology, and evolution of avian plumage carotenoids and the role of carotenoid-protein interactions in plumage color appearance. *Archives of Biochemistry and Biophysics*, 572, pp.201–12, doi:10.1016/j.abb.2015.01.016.

Lamichhaney, S., Berglund, J., et al. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, 518, pp.371–75, doi:10.1038/nature14181.

Lamichhaney, S., Fan, G., et al. 2015. Structural genomic changes underlie alternative reproductive strategies in the Ruff (*Philomachus pugnax*). *Nature Genetics*, 48, pp.84–8, doi:10.1038/ng.3430..

Langmead, B. & Salzberg, S. 2012. Fast gapped-read alignment with Bowtie2. *Nature Methods*, 9, pp.357–59, doi:10.1038/nmeth.1923.

Lank, D., Coupe, M. & Wynne-Edwards, K. 1999. Testosterone-induced male traits in female Ruffs (*Philomachus pugnax*): autosomal inheritance and gender differentiation. *Proceedings of the Royal Society B: Biological Sciences*, 266, pp.2323–30, doi:10.1098/rspb.1999.0926.

Lehtonen, P. et al. 2012. Candidate genes for colour and vision exhibit signals of selection across the pied flycatcher (*Ficedula hypoleuca*) breeding range. *Heredity*, 108, pp.431–40, doi:10.1038/hdy.2011.93.

Lewis, M. et al. 2005. *A Guide to Gouldian Finches and Their Mutations,* revised ed., ABK Publications, Australia.

Li, H., Ruan, J. & Durbin, R. 2008. Mapping short DNA sequencing reads and calling variants using mapping. *Genome Research*, 18, pp.1851–58, doi:10.1101/gr.078212.108.

Liu, J. et al. 2014. Developmental pathways activated in melanocytes and melanoma. *Archives of Biochemistry and Biophysics*, 563, pp.13–21, doi:10.1016/j.abb.2014.07.023.

Loftus, S. et al. 2002. Mutation of melanosome protein RAB38 in chocolate mice. *Proceedings of the National Academy of Sciences*, 99, pp.4471–76, doi:10.1073/pnas.072087599.

Logan, C. & Nusse, R. 2004. Wnt Signaling in Development and Disease. *Annual Review of Cell and Developmental Biology*, 20, pp.781–810, doi:10.1146/annurev.cellbio.20.010403.113126.

Long, A. et al. 2017. Genetic Variation of Nine Chicken Breeds Collected from Different Altitudes Revealed by Microsatellites. *Poultry Science*, 54, pp.18-25, doi:10.2141/jpsa.0160033.

Lopes, R. et al. 2016. Genetic Basis for Red Coloration in Birds. *Current Biology*, 26, pp.1427–34, doi:10.1016/j.cub.2016.03.076.

Love, M., Huber, W. & Anders, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 15, pp.550, doi:10.1186/s13059-014-0550-8.

Lozano, G. 2001. Carotenoids, immunity, and sexual selection: comparing apples and oranges? *The American naturalist*, 158, pp.200–03, doi:10.1086/321313.

Manceau, M. et al. 2011. The Developmental Role of Agouti in Color Pattern Evolution. *Science*, 331, pp.1062–65, doi:10.1126/science.1200684.

Marshall, J. & Johnsen, S. 2017. Fluorescence as a means of colour signal enhancement. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372:20160335, doi:10.1098/rstb.2016.0335.

Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 17, pp.10, doi:10.14806/ej.17.1.200.

McGraw, K. 2004. Colorful songbirds metabolize carotenoids at the integument. *Journal of Avian Biology*, 35, pp.471–76, doi:10.1111/j.0908-8857.2004.03405.x.

McGraw, K. et al. 2004. Red-winged blackbirds, *Agelaius phoeniceus,* use carotenoid and melanin pigments to color their epaulets. *Journal of Avian Biology*, 35, pp.543–50, doi:10.1111/j.0908-8857.2004.03345.x.

McGraw, K. & Ardia, D. 2003. Carotenoids, immunocompetence, and the information content of sexual colors: an experimental test. *The American naturalist*, 162, pp.704–12, doi:10.1086/378904.

Mckenna, A. et al. 2010. The Genome Analysis Toolkit : A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20, pp.1297–1303, doi:10.1101/gr.107524.110.

Meléndez-Martínez, A. et al. 2006. Relationship between the colour and the chemical structure of carotenoid pigments. *Food Chemistry*, 101, pp.1145–50, doi:10.1016/j.foodchem.2006.03.015.

Minvielle, F. et al. 2010. The "*silver*" Japanese quail and the *MITF* gene: causal mutation, associated traits and homology with the *"blue"* chicken plumage. *BMC Genetics*, 11, pp.1471-2156, doi:10.1186/1471-2156-11-15.

Mochii, M. et al. 1992. Isolation and Characterization of a Chicken Tyrosinase cDNA. *Pigment Cell Research*, 5, pp.162–67, doi:10.1111/j.1600-0749.1992.tb00454.x.

Mundy, N. 2005. A window on the genetics of evolution: MC1R and plumage colouration in birds. *Proceedings of the Royal Society B: Biological Sciences*, 272, pp.1633–40, doi:10.1098/rspb.2005.3107.

Mundy, N. et al. 2016. Red Carotenoid Coloration in the Zebra Finch Is Controlled by a Cytochrome P450 Gene Cluster. *Current Biology*, 26, pp.1435–40, doi:10.1016/j.cub.2016.04.047.

Nachman, M., Hoekstra, H. & D'Agostino, S. 2003. The genetic basis of adaptive melanism in pocket mice. *Proceedings of the National Academy of Sciences,* 100, pp.5268–73, doi:10.1073/pnas.0431157100.

Nadeau, N. et al. 2008. Characterization of Japanese quail yellow as a genomic deletion upstream of the avian homolog of the mammalian *ASIP* gene. *Genetics*, 178, pp.777–86, doi:10.1534/genetics.107.077073.

Ng, P. & Henikoff, S. 2003. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research*, 31, pp.3812–14, doi:10.1093/nar/gkg509.

Oetting, W. 2000. The Tyrosinase Gene and Oculocutaneous Albinism Type 1 (OCA1): A Model for Understanding the Molecular Biology of Melanin Formation. *Pigment Cell Research*, 13, pp.320–25, doi:10.1034/j.1600-0749.2000.130503.x.

Oetting, W. & King, R. 1999. Molecular basis of albinism: Mutations and polymorphisms of pigmentation genes associated with albinism. *Human*

Mutation, 13, pp.99–115, doi:10.1002/(SICI)1098-1004(1999)13:2<99::AID-HUMU2>3.0.CO;2-C.

Ohbayashi, N. & Fukuda, M. 2012. Role of Rab family GTPases and their effectors in melanosomal logistics. *Journal of Biochemistry*, 151, pp.343–51, doi:10.1093/jb/mvs009.

Pan, S. et al. 2017. Population transcriptomes reveal synergistic responses of DNA polymorphism and RNA expression to extreme environments on the Qinghai–Tibetan Plateau in a predatory bird. *Molecular Ecology*, 26, pp.2993–3010, doi:10.1111/mec.14090.

Park, T. et al. 2014. Wnt inhibitory factor (*WIF1)* promotes melanogenesis in normal human melanocytes. *Pigment Cell and Melanoma Research*, 27, pp.72–81, doi:10.1111/pcmr.12168.

Pérez-Rodríguez, L., Jovani, R. & Stevens, M. 2017. Shape matters: animal colour patterns as signals of individual quality. *Proceedings of the Royal Society B: Biological Sciences*, 284:20162446, doi:10.1098/rspb.2016.2446.

Peters, A. et al. 2007. The Condition-Dependent Development of Carotenoid-Based and Structural Plumage in Nestling Blue Tits: Males and Females Differ. *The American Naturalist*, 169, pp.22–36, doi:10.1086/510139.

Poelstra, J. et al. 2015. Transcriptomics of colour patterning and coloration shifts in crows. *Molecular Ecology*, 24, pp.4617–28, doi:10.1111/mec.13353.

Poelstra, J., Ellegren, H. & Wolf, J. 2013. An extensive candidate gene approach to speciation: diversity, divergence and linkage disequilibrium in candidate pigmentation genes across the European crow hybrid zone. *Heredity*, 111, pp.467–73, doi:10.1038/hdy.2013.68.

Price, T. 2007. *Speciation in Birds*. Roberts & Company Publishers, Colorado.

Price, T. 2002. Domesticated birds as a model for the genetics of speciation by sexual selection. *Genetica*, 116, pp.311–27, doi:10.1023/A:1021248913179.

Prum, R. 2003. Structural colouration of avian skin: convergent evolution of coherently scattering dermal collagen arrays. *Journal of Experimental Biology*, 206, pp.2409–29, doi:10.1242/jeb.00431.

Pryke, S. 2007. Fiery red heads: Female dominance among head color morphs in the Gouldian finch. *Behavioral Ecology*, 18, pp.621–27,

doi:10.1093/beheco/arm020.

Pryke, S. 2010. Sex chromosome linkage of mate preference and color signal maintains assortative mating between interbreeding finch morphs. *Evolution*, 64, pp.1301–10, doi:10.1111/j.1558-5646.2009.00897.x.

Pryke, S. & Griffith, S. 2009. Genetic incompatibility drives sex allocation and maternal investment in a polymorphic finch. *Science*, 323, pp.1605–07, doi:10.1126/science.1168928.

Pryke, S. & Griffith, S. 2009. Postzygotic genetic incompatibility between sympatric color morphs. *Evolution*, 63, pp.793–98, doi:10.1111/j.1558-5646.2008.00584.x.

Pryke, S. & Griffith, S. 2007. The relative role of male vs. female mate choice in maintaining assortative pairing among discrete colour morphs. *Journal of Evolutionary Biology*, 20, pp.1512–21, doi:10.1111/j.1420-9101.2007.01332.x.

Ramstad, K., Miller, H. & Kolle, G. 2016. Sixteen kiwi (*Apteryx spp*) transcriptomes provide a wealth of genetic markers and insight into sex chromosome evolution in birds. *BMC Genomics*, 17, [no pagination], doi:10.1186/s12864-016-2714-2.

Ranz, M. et al. 2003. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. *Science*, 300, pp.1742–45, doi:10.1126/science.1085881.

Robinson, J. et al. 2011. Integrative Genomics Viewer. *Nature Biotechnology*, 29, pp.24–26, doi:10.1038/nbt.1754.

Roulin, A. & Ducrest, A.-L. 2013. Genetics of colouration in birds. *Seminars in Cell & Developmental Biology*, 24, pp.594–608, doi:10.1016/j.semcdb.2013.05.005.

Roulin, A., Da Silva, A. & Ruppli, C. 2012. Dominant nestlings displaying female-like melanin coloration behave altruistically in the barn owl. *Animal Behaviour*, 84, pp.1229–36, doi:10.1016/j.anbehav.2012.08.033.

Rubin, C.-J. et al. 2010. Whole-genome resequencing reveals loci under selection during chicken domestication. *Nature*, 464, pp.587–91, doi:10.1038/nature08832.

Saino, N. et al. 2013. Sexual Dimorphism in Melanin Pigmentation, Feather Coloration and Its Heritability in the Barn Swallow (*Hirundo rustica*). *PLoS*

ONE, 8:e58024, doi:10.1371/journal.pone.0058024.

Saranathan, V. et al. 2012. Structure and optical function of amorphous photonic nanostructures from avian feather barbs: a comparative small angle X-ray scattering (SAXS) analysis of 230 bird species. *Journal of The Royal Society Interface*, 9, pp.2563–80, doi:10.1098/rsif.2012.0191.

Shapiro, M. et al. 2013. Genomic diversity and evolution of the head crest in the Rock pigeon. *Science*, 339, pp.1063–67, doi:10.1126/science.1230422.

Shawkey, M. 2006. Significance of a basal melanin layer to production of non-iridescent structural plumage color: evidence from an amelanotic Steller's jay (*Cyanocitta stelleri)*. *Journal of Experimental Biology*, 209, pp.1245–50, doi:10.1242/jeb.02115.

Shawkey, M. & Hill, G. 2005. Carotenoids need structural colours to shine. *Biology Letters*, 1, pp.121–24, doi:10.1098/rsbl.2004.0289.

Simpson, E. et al. 1994. Aromatase Cytochrome P450: The Enzyme Responsible for Estrogen Biosynthesis. *Endocrine Reviews*, 15, pp.342–55, doi:10.1210/edrv-15-3-342.

Southern, H. 1945. Polymorphism in *Poephila gouldiae*. *Journal of Genetics*, 47, pp.51–7, doi:10.1007/BF02989037.

Stoddard, M. & Prum, R. 2011. How colorful are birds? Evolution of the avian plumage color gamut. *Behavioral Ecology*, 22, pp.1042–52, doi:10.1086/587526.

Takeuchi, S. et al. 1996. Possible involvement of melanocortin 1-receptor in regulating feather color pigmentation in the chicken. *Biochimica et Biophysica Acta - Gene Structure and Expression*, 1308, pp.164–68, doi:10.1016/0167-4781(96)00100-5.

Tarafder, A. et al. 2014. Rab11b mediates melanin transfer between donor melanocytes and acceptor keratinocytes via coupled exo/endocytosis. *The Journal of Investigative Dermatology*, 134, pp.1056–66, doi:10.1038/jid.2013.432.

Theron, E. et al. 2001. The molecular basis of an avian plumage polymorphism in the wild: a point mutation in the melanocortin-1 receptor is perfectly associated with melanism in the bananaquit (*Coereba flaveola*). *Current Biology*, 11, pp.550–57, doi:10.1016/S0960-9822(01)00158-0.

Toews, D. et al. 2016. Plumage Genes and Little Else Distinguish the Genomes of Hybridizing Warblers. *Current Biology*, 26, pp.2313–18, doi:10.1016/j.cub.2016.06.034.

Toomey, M. et al. 2017. High-density lipoprotein receptor SCARB1 is required for carotenoid coloration in birds. *Proceedings of the National Academy of Sciences*, 114, pp.5219–24, doi:10.1073/pnas.1700751114.

Trapnell, C. et al. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, 7, pp.562–78, doi:10.1038/nprot.2012.016.

Tworoger, S. et al. 2004. Association of *CYP17, CYP19, CYP1B1, and COMT* Polymorphisms with Serum and Urinary Sex Hormone Concentrations in Postmenopausal Women. *Cancer Epidemiology and Prevention Biomarkers*, 13, pp.94–101, doi:10.1158/1055-9965.EPI-03-0026.

Tyczkowski, J. & Hamilton, P. 1986. Evidence for differential absorption of zeacarotene, cryptoxanthin, and lutein in young broiler chickens. *Poultry Science*, 65, pp.1137–40, doi:10.3382/ps.0651137.

Vaez, M. et al. 2008. A single point-mutation within the melanophilin gene causes the lavender plumage colour dilution phenotype in the chicken. *BMC Genetics*, 9, [no pagination], doi:10.1186/1471-2156-9-7.

Vrieling, H. et al. 1994. Differences in dorsal and ventral pigmentation result from regional expression of the mouse agouti gene. *Proceedings of the National Academy of Sciences*, 91, pp.5667–71, doi:10.1073/pnas.91.12.5667.

Wang, X.-W. et al. 2010. De novo characterization of a whitefly transcriptome and analysis of its gene expression during development. *BMC Genomics*, 11, [no pagination], doi:10.1186/1471-2164-11-400.

Warren, W. et al. 2010. The genome of a songbird. *Nature*, 464, pp.757–62, doi:10.1038/nature08819.

Watson, H. et al. 2017. Transcriptome analysis of a wild bird reveals physiological responses to the urban environment. *Scientific Reports*, 7, [no pagination], doi:10.1038/srep44180.

Wyss, A. 2004. Carotene oxygenases: a new family of double bond cleavage enzymes. *The Journal of Nutrition*, 134, pp.246S–50S.

Yeh, P. 2004. Rapid evolution of a sexually selected trait following population establishment in a novel habitat. *Evolution*, 58, pp.166–74, doi: 10.1111/j.0014-3820.2004.tb01583.x.

Yoshihara, C. et al. 2012. Elaborate color patterns of individual chicken feathers may be formed by the agouti signaling protein. *General and Comparative Endocrinology*, 175, pp.495–99, doi:10.1016/j.ygcen.2011.12.009.

Yu, M. et al. 2004. The developmental biology of feather follicles. *The International Journal of Developmental Biology*, 48, pp.181–91, doi:10.1387/ijdb.031776my.

Zhang, L. et al. 2017. Genetic evidence from mitochondrial DNA corroborates the origin of Tibetan chickens. *PLoS ONE*, 12:e0172945, doi: 10.1371/journal.pone.0172945.

Zhao, Y. et al. 2012. Bio-inspired variable structural color materials. *Chemical Society Reviews*, 41, pp.3297–317, doi: 10.1039/C2CS15267C.

Zink, R. et al. 2004. Mitochondrial DNA Variation, Species Limits, and Rapid Evolution of Plumage Coloration and Size in the Savannah Sparrow. *The Condor*, 107, pp.21–8, doi:10.1650/7550.

Zinzow-Kramer, W. et al. 2015. Genes located in a chromosomal inversion are correlated with territorial song in white-throated sparrows. *Genes, Brain and Behavior*, 14, pp.641-54, doi:10.1111/gbb.12252.

Appendix

The excel file entitled *APPENDIX1.xlsx,* in Microsoft Office 2010 .xlsx format, contains the following list of tables:

Table S1 - Gene Ontology (GO) terms for upregulated differentially expressed genes between red and yellow colour morphs.

Table S2 - Gene Ontology (GO) terms for downregulated differentially expressed genes between red and yellow colour morphs.

A list of GO terms for melanin vs. carotenoid-based phenotypes is not included, as all de differentially expressed genes are highlighted in the main text.