Evolutionary genetics of *CYP2J19* in red carotenoid pigmentation

Hanlu Twyman

Peterhouse

September 2017

This dissertation is submitted for the degree of Doctor of Philosophy

Evolutionary genetics of *CYP2J19* in red carotenoid pigmentation

Hanlu Twyman

Carotenoids are responsible for much of the bright yellow to red colours in animals and have been extensively studied as condition dependent signals in sexual selection. In addition to their function in coloration, carotenoids also play a crucial role in colour vision within certain lineages. Despite this, little is known about the genetic mechanisms underlying carotenoid based pigmentation. Recently, the gene *CYP2J19* was strongly implicated in red ketocarotenoid pigmentation for coloration and colour vision within two lineages of song birds (the zebra finch and the red factor canary). Here, I extend the investigation of the function of *CYP2J19* in colour vision and red coloration amongst reptiles. I suggest that the original function of *CYP2J19* was in colour vision and that it has been independently co-opted for red coloration within certain red lineages. Using a combination of phylogenetic and expression analysis, I study the role of *CYP2J19* as the avian ketolase involved in red ketocarotenoid generation within a clade of well-studied seed-eating passerines, the weaverbirds, and demonstrate a direct association between levels of *CYP2J19* expression and red ketocarotenoid-based coloration. Next, I consider the evolution of *CYP2J19* across multiple avian lineages. I find evidence for positive selection acting on the gene coding sequence despite its conserved function in colour vision. This finding, though surprising, appears to be common across avian *CYP* loci in general. Finally, by considering the genomic organisation of *CYP2J19* in the zebra finch, I find that the gene underwent a duplication event near the base of the estrildid lineage, which was followed by significant gene conversion post-duplication. Overall, the findings provide strong support for the role of *CYP2J19* in red ketocarotenoid pigmentation and demonstrate how an understanding of evolutionary genetics benefit the study of the evolution of adaptive phenotypes.

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. This dissertation does not exceed the prescribed word limit for the relevant Degree Committee.

Acknowledgements

I would like to firstly thank my supervisor Nick Mundy for his amazing support, advice and guidance throughout the PhD. I have been very privileged to collaborate with many distinguished scientists who have made this thesis possible. In particular, I would like to thank the people who painstakingly collected the various bird and turtle samples, Prof Staffan Andersson, Dr Maria Prager, Dr Nicole Valenzuela, Dr Robert Literman, Dr Claire Spottiswoode, Dr Simon Griffith and Gabriel Jamie. I would also like to thank my advisors, Dr Andrea Manica and Dr John Davey for their thoughtful comments and discussion, Dr Kang-Wook Kim and members of the Jiggins group for their help and advice with bioinformatics related work, and the Zoology department technical staff for their hard work, especially in making the smooth running of the various experiments possible.

My family and friends have been a great support throughout, and I would like to thank my parents, Jin and Peter as well as all my family members near and far for their incredible support and kindness. In particular, would like to thank Paula, Beatrice, Simmy, Tong and Rebekah from Cambridge, and also David, Angelika, Eunice, JoyLynn, Pauline, Jason and Emily, Jane, Charlie and Serena.

This PhD is funded by BBSRC Doctorial Training Partnerships.

Table of Contents

Chapter 1: Introduction……7

Carotenoids and avian coloration……8

Carotenoids in sexual selection……12

Carotenoids as a handicap signal……12

Carotenoids as an index signal……16

Summary……17

Avian colour vision – carotenoids in oil droplets……18

Genetic basis of adaptive evolution……21

Genetics of carotenoid metabolism in animals……24

Thesis outline……26

Chapter 2: General experimental methods……27

Chapter 3: Seeing red to being red: conserved genetic mechanism for red cone oil droplets and co-option for red coloration in birds and turtles……30

Abstract……31

Introduction……32

Methods……34

Results……38

Evolution of CYP2J genes in amniotes……38

Specific expression of CYP2J19 in tissues containing red ketocarotenoids in western painted turtles……41

Discussion……43

Chapter 4: *CYP2J19* underlies convergent evolution of sexually-selected red carotenoid coloration in weaverbirds (Ploceidae)……48

Abstract……49

Introduction……50

Materials and Methods……52

Results……58

CYP2J19 presence, copy number and variation……58

Patterns of CYP2J19 expression……58

Association between CYP2J19 and ketocarotenoid pigmentation.....63

Discussion……..……..……64

Chapter 5: Evolution of *CYP2J19,* a gene for red coloration and colour vision in birds: Positive selection in the face of conservation and pleiotropy……..67

Abstract..……67

Introduction……68

Materials and Methods…...……70

Results……75

Discussion……..……79

Chapter 6: The evolution of *CYP2J19* copy number and function in indigobirds, whydahs (Viduidae) and estrildid finches (Estrildidae)……...……82

Abstract...……..82

Introduction...……83

Materials and Methods…….……85

Results……92

CYP2J19 copy number determination……..92

Expression patterns of CYP2J19...……...95

Evolution of CYP2J19 in estrildid finches……..….100

Discussion...……...……102

Chapter 7: Concluding thoughts...……..……..……..108

Genetics of *CYP2J19*…….……108

Site of ketoconversion.....……..……..110

References...……...……..112

Chapter 1

Introduction

Carotenoids have been well studied in birds for their role in colour signalling and in colour vision. Despite this, surprisingly little is known about the genetics underlying carotenoid pigmentation. Here, I consider the ways in which carotenoids function as honest signals for quality, and review what is known about the roles of carotenoids in the avian visual system. I describe how genetics can influence the evolution of adaptive phenotypes and focus on what is known about the genetics of carotenoid-based pigmentation. This forms the basis for my thesis which aims to investigate the evolution and function of the genes underlying red carotenoid metabolism in birds.

Carotenoids and avian coloration

There is tremendous diversity in colour variation across the animal kingdom. Colour can arise through different mechanisms: as a result of the physical interactions of light waves with nanostructured surfaces that cause scattering of light (structural coloration) (reviewed by (Kinoshita and Yoshioka 2005)) and/or from the presence of pigments that absorb wavelengths of light from specific regions in the electromagnetic spectrum.

One such group of pigments that has received much attention over the years for conferring bright yellow, orange and red hues are carotenoids - organic compounds produced by plants and other photosynthetic organisms including bacteria and fungi, where their primary function is in energy capture, transfer and dissipation (Vershinin 1999). Carotenoids are used for coloration in most invertebrate phyla and in all vertebrate classes where they are present in nearly all forms of integumentary tissues with the exception of mammalian hairs (Hill and McGraw 2006). In addition, carotenoids are the second most prevalent pigment found in avian integument (the first being melanin) (Hill and McGraw 2006) and they have been extensively studied with regards to colour signalling in birds (see below).

Carotenoids are tetraterpenoids and contain eight 5-carbon isoprene subunits ($CH_2=C(CH_3)CH=CH_2$) that collectively make up the 40-carbon molecule. The hydrocarbon chain can be circularised at either one or both ends to form ionone rings, with one double bond per ring. The end rings are known as β-ionone rings, when the double bond is located in the C5-C6 position, or ε-ionone rings when the double bond is found in the C4-C5 position (Britton 1995). Although over 750 carotenoids have been described in nature (Britton *et al.* 2004), far fewer carotenoids have been characterised in birds (Hill and McGraw 2006; LaFountain *et al.* 2010; LaFountain *et al.* 2013; Prum *et al.* 2014) (Figure 1.1). The classification of carotenoids depends on the different functional groups that are substituted onto the end rings. Unsubstituted carotenoids are known as carotenes (e.g. β-carotene), whereas carotenoids that contain oxygen are collectively known as xanthophylls. Depending on the oxygen containing group, xanthophylls can be further subdivided into hydroxyl carotenoids (e.g. lutein) or ketocarotenoids (e.g. astaxanthin) (Britton 1995). All carotenoids share a common feature, namely the chromophore (a system of alternating double and single bonds) which absorbs wavelengths of light of 400-

500nm (blue-green) and gives the molecule its distinctive yellow, orange and red colour depending on the length of the conjugated bond system. The longer the conjugated bond system, the longer the wavelengths of light that are absorbed, and the redder the resultant colour. As a result, the colour of the carotenoid containing tissue is dependent on 1) the type of carotenoids it contains, 2) the concentration of the pigments (Andersson *et al.* 2007), 3) interactions with other biological pigments such as melanin, for example in the red feathers of the red-winged blackbird (McGraw *et al.* 2004), 4) interactions with structural colours, such as the green feathers of the green broadbill (Prum *et al.* 2014), and 5) carotenoid-protein interactions (Stradi *et al.* 1995; Mendes-Pinto *et al.* 2012).

Vertebrates cannot synthesize carotenoids *de novo* (i.e. from other biological molecules such as fats and proteins) but must obtain them from their diets. Therefore, the abundance of carotenoids in the environment, the eating habits of the animal, and the amount of carotenoids the animal can absorb will critically affect how much carotenoid the animal can utilise (Hill and McGraw 2006). In birds, carotenoids have been described to pigment various tissues, including egg yolks, bare body parts (such as beaks and tarsi), and feathers. The use of carotenoids for integumental coloration (including feathers and bare body parts) is widespread among extant birds, and notably, plumage carotenoids appear to have evolved multiple times independently from an unpigmented ancestor (Thomas *et al.* 2014), although similar studies are awaited for carotenoid coloured bare body parts.

Commonly, C4-ketocarotenoids such as astaxanthin, canthaxanthin and adonirubin are used for red integumental coloration, whereas lutein, zeaxanthin and canary xanthophylls are generally used to confer yellow (Hill and McGraw 2006) (Figure 1.1). Both dietary (e.g. lutein and zeaxanthin) and derived carotenoids, carotenoids that are synthesized from the metabolic conversion of dietary carotenoids (e.g. adonirubin) are used as colorants, and there has been an ongoing debate as to the primary anatomical site of metabolism for derived carotenoid. Indeed, studies across several avian lineages point towards either the liver (central conversion) or the peripheral tissues (peripheral conversion) as the main metabolic site. Evidence for the central conversion of carotenoids comes from the detection of derived ornamental carotenoids in the liver and the blood plasma of individual bird species (del Val *et al.* 2009; Prager *et al.* 2009). This was supported by a study that

found high concentrations of synthetic ketocarotenoids in the liver cells of wild male house finches, which is in line with the central metabolism of these ketocarotenoids (Ge *et al.* 2015). In contrast, analysis of several species of passerines (McGraw 2004) and non-passerines (Garcia-de Blas *et al.* 2015; Pérez-Rodríguez *et al.* 2016) found an absence of metabolically derived carotenoids in internal tissues that were otherwise present in the carotenoid coloured tissues, which suggests direct transformation at the integuments, consistent with the peripheral conversion theory. It appears that birds employ both central and peripheral metabolism for integumental coloration, although the adaptive value of each is still unclear.

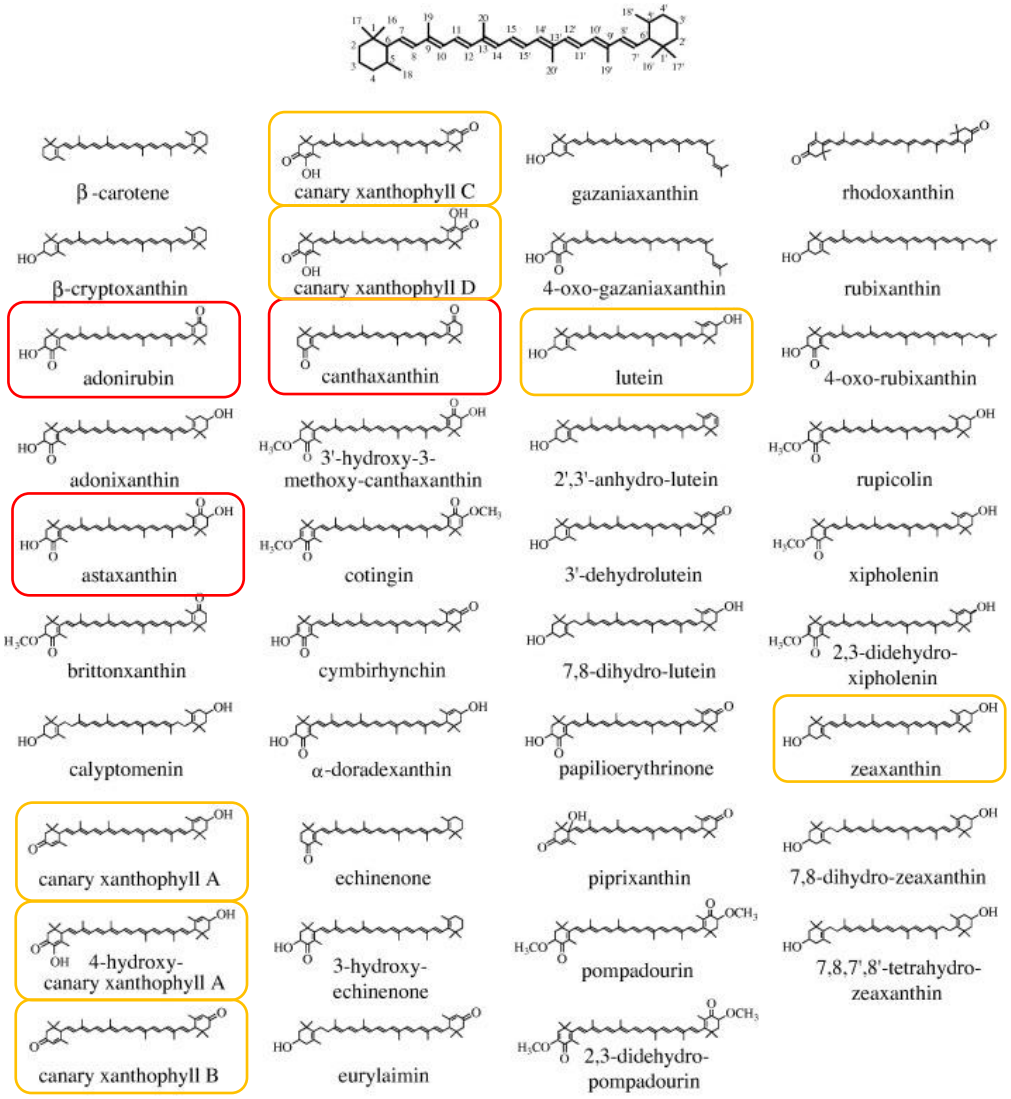

Figure 1.1. Chemical structure of some of the avian carotenoids used for coloration. Red and yellow boxes highlight common avian carotenoids used in red and yellow coloration. The carbon numbering scheme for a generic carotenoid molecule is shown at the top. Modified image taken from LaFountain *et al.* (2015)

Carotenoids in sexual selection

Much of the work on carotenoids in animals focuses on their role in signalling. Biological signals can be honest, whereby the signal will select for an appropriate response in the receiver, or dishonest, when the sender signals to manipulate the behaviour of the receiver in order to benefit themselves only (Davies *et al.* 2012). Thus, signal honesty is essential for effective mutual communication, and various mechanisms have been proposed to explain the maintenance of signal honesty. One such mechanism is the 'handicap principle' (Zahavi 1975), which states that the signal must have a cost to the signaller, such that only high-quality signallers can afford the signal despite the costs. Hence, the feature is an indirect signal to the receiver about the quality of the sender. Critically, it assumes that the costs associated with ornamentation are lower in high quality individuals compared with low quality individuals, making it possible only for high quality individuals to carry the ornaments as the benefits for them would outweigh the costs. This contrasts with the direct signalling mechanism in which the intensity of the signal is causally related to the quality of the sender, and is therefore impossible to fake (Davies *et al.* 2012).

In birds, carotenoids have been extensively studied as sexually selected signals involved in both intrasexual (Pryke *et al.* 2001; Pryke and Andersson 2003; Murphy *et al.* 2009) and intersexual competition (Hill 1999), and here I review how avian carotenoids can be used both indirectly (handicap principle) and directly (index hypothesis) as honest signals for individual quality.

Carotenoids as a handicap signal

According to the handicap principle, carotenoids are only reliable as a signal if they confer a cost (Olson and Owens 1998). Such costs can be viewed in terms of the costs for trait expression (e.g. how conspicuous signals relates to increased predation risks), as well as the costs of trait production (e.g. internal physiological costs). Increasingly research focus has shifted towards the internal costs of signal production which includes 1) trade-offs between the use of carotenoids in ornamentation and carotenoids in health benefiting systems (Lozano 1994), 2) the metabolic costs of conversion as well as 3) the potential toxicity costs of carotenoids.

1) Trade-offs between signalling and health

In order for carotenoids to function as a trade-off signal, carotenoids must a) be in short supply, such that birds cannot obtain enough for all their needs, and b) have fitness enhancing health qualities that benefit the signaller in addition to those gained from signalling (Svensson and Wong 2011). One of the more obvious factors limiting carotenoid supply is that carotenoids can only be obtained from the avian diet. The feeding behaviour (whether the diet is mainly based on plants or animals) can strongly influence the dietary supply of carotenoids. Lutein and zeaxanthin (both hydroxyl carotenoids) are among the more abundant forms of dietary carotenoids found in plants and terrestrial herbivorous prey (Hill and McGraw 2006), in contrast with C4-ketocarotenoids (such as astaxanthin) which are the dominant carotenoid in aquatic animals (Matsuno 2001). The amount of carotenoids that enters circulation and made available to the animal is further limited by the physical matrix from which carotenoids are extracted, and the polarity and stability of the carotenoid molecules (Hill and McGraw 2006; Svensson and Wong 2011).

Despite this, there are reasons why carotenoids may not be limited in most animal diets, and one of them relates to the idea of bioaccumulation (when an organism absorbs a substance at a higher rate than the substance is lost), a process in which selection favours mechanisms that maximise carotenoid acquisition and uptake, and lead to the concentration of carotenoids up the food web (LaFountain *et al.* 2015). In addition, given the solely dietary origin of carotenoids, it follows that birds would alter their behaviours to actively seek out carotenoid rich foods, which assumes that they are able to detect small differences in the nutrient content of foods so as to self-select a nutritionally balanced diet. Accordingly, an investigation on four species of tanagers found that three of the four were able to detect differences in sugar and lipid content as well as between different types of proteins when presented with foods containing the same absolute protein levels (Schaefer *et al.* 2003). Moreover, lab choice tests on both captive and wild great tits showed a preference for carotenoid enriched foods, which did not differ in any other aspect except for carotenoid content with experimental controls. This demonstrates that these birds have the ability to detect and select for food items rich in carotenoids, and is suggestive of a specific appetite for these molecules (Senar *et al.* 2010). Given this ability, the presence of carotenoid coloration may not reflect a true trade-

off between allocation of carotenoids away from the physiological processes towards signal production, but rather reflect the ability of the animal to acquire carotenoids in the first place, and therefore the associated cost of carotenoid acquisition is what maintains signal honesty.

The second aspect of the trade-off hypothesis requires that carotenoids have fitness enhancing physiological benefits, and many biological functions have been described for carotenoids (Krinsky 1993), which include its function as an antioxidant (Miki 1991; Svensson and Wong 2011), in the immune system (Blount *et al.* 2003), as a vitamin A precursor (Moore 1930), and in cell-cell communication (Zhang *et al.* 1992). Of these functions, the role of carotenoids as an antioxidant and their function in the immune system have been well studied. The highly conjugated chromophore of carotenoids enables them to readily accept unpaired electrons from reactive oxygen species (ROS), which are generated as a part of the normal aerobic metabolism, and serve to minimise cellular damage, particularly lipid peroxidation (oxidative damage of lipids, especially in the cell membrane) (Svensson and Wong 2011). Xanthophylls such as astaxanthin appear to be more effective antioxidants than β-carotene and vitamin E (Miki 1991), and in addition, carotenoids can cooperate with other endogenous antioxidants such as vitamin E and vitamin C, where they can function to repair vitamin E radicals and be themselves repaired by vitamin C (Svensson and Wong 2011). Connected with the function of carotenoids as antioxidants, carotenoids can function in the immune system by quenching ROS produced by white blood cells when they encounter foreign microbes, thus protecting the white blood cells and surrounding tissues from oxidative damage (Svensson and Wong 2011). Despite the beneficial roles of carotenoids, studies of avian systems found significant heterogeneities over the role of carotenoids as important antioxidants, and it would appear that the relative importance of carotenoids as antioxidants varies between different taxa (Costantini and Moller 2008).

2) Costs associated with carotenoid metabolism

Metabolic transformation of ingested carotenoids is an energy demanding process (Hill 2000). As a result of the costs associated with metabolism, the use of derived carotenoids in ornamentation serves as an indirect indicator for quality. Two

general transformations have been described in birds *in vivo*: the adding or removing of functional groups, and the adding and removing of double bonds (LaFountain *et al.* 2015). An example of a modification involving the addition of oxygenated groups is C4-oxygenation (C4-ketolation): the introduction of a keto group (C=O) to the 4' carbon of the carotenoid β-ring. This results in the lengthening of the conjugated bond system and is responsible for red coloration (Prager *et al.* 2009). The diversity of avian carotenoid transformations has some interesting characteristics: 1) not all modifications are possible on all precursors, which mainly relates to the functional groups present on the precursor, 2) the same modifications can be performed on many different precursor carotenoids (e.g. a single C4-oxygenation on lutein and adonixanthin turns it into α-doradexanthin and astaxanthin respectively), 3) the same set of transformations can produce very different end products by changing the order in which the transformations occur, 4) some modifications facilitate subsequent transformations (LaFountain *et al.* 2015). However, not all birds use metabolically synthesized carotenoids for ornamentation. The use of synthetic carotenoids appears to be a derived trait in certain avian lineages, such as in the bishops and widowbirds (*Euplectes* spp.) (Prager *et al.* 2009), as well as in the fringillid finches (Ligon *et al.* 2016).

3) Potential toxicity costs of carotenoids

At very high concentrations, carotenoids can have pro-oxidant effects and induce oxidative stress. Carotenoids can autoxidise or be subject to attack from other radical species to form carotenoid radicals, which unless they are quenched by other antioxidants (e.g. vitamin C), may be further oxidised and irreversibly fragmented to form toxic aldehydes known as apo-carotenals (Hartley and Kennedy 2004). In such a case, carotenoid based coloration can be viewed as a handicap whereby only the fittest males can afford bright coloration and endure the toxicity together with the cost of detoxification (Hill and McGraw 2006). The detrimental effects of carotenoids can be seen in avian carotenoid supplementation studies where high supplemental doses of carotenoids result in increased production of ROS and elevated oxidative stress in kestrels (Costantini *et al.* 2007), as well as increased muscle breakdown and impaired flight performance in American goldfinches

(Huggins *et al.* 2010), although the extent to which carotenoids are harmful in nature is unclear.

Recent considerations have suggested that the handicap principle may not be sufficient to explain signal honesty in some cases. The explanation for this relates to the co-evolution of compensatory mechanisms (such as changes in structure, behaviour or physiology) together with the selected trait, which ameliorates the costs associated with trait expression or trait production (Husak and Swallow 2011). Given this, current thinking is shifting towards the presence of a direct link between the extent of ornamentation and the signaller's intrinsic physiological condition.

Carotenoids as an index signal

Carotenoids can function as a direct signal for individual quality. In general, the focus has been on metabolically derived carotenoids and how the process of carotenoid metabolism is associated with fundamental cellular processes. Indeed, metabolically derived ornamental carotenoids can serve as a direct indicator for cellular redox state and vitamin A pools (Hill and D. 2012), as well as a signal for the efficiency of cellular respiration (Johnson and E. 2013). The latter can relate to the similarities between carotenoid metabolism and ubiquinone (UQ) biosynthesis. UQ is a membrane bound molecule well known for its role in aerobic respiration. UQ and carotenoids are highly similar in structure and function (as anti- and prooxidants), and the pathway leading to UQ synthesis involves the same type of modifications as found in carotenoid metabolism. Three metabolic transformations described for carotenoids, namely C4-oxygenation, O-methylation (addition of a methyl group (CH_3) onto a hydroxyl group (OH)) and C(2,3)-didehydrogenation (introduction of a double bond between carbon positions 2 and 3), are also described for UQ biosynthesis (Johnson and E. 2013). The enzymes involved in UQ biosynthesis are found in the inner mitochondrial membrane (IMM), and if the same enzymes are also used for carotenoid metabolism, then it would imply that carotenoid pigmentation could potentially signal for the most fundamental biochemical process of living organisms. For this to be plausible, it assumes that the mitochondria are the primary subcellular site for ornamental carotenoid synthesis. In birds, carotenoids have been known to be transported into the mitochondria (Mayne and Parker 1986), and a

recent study on wild red moulting male house finches provided the first demonstration of the possible hepatic mitochondrial origin of ornamental carotenoids (Ge *et al.* 2015). Hence, if carotenoid ketolation is connected with the mitochondrial membrane, then oxidation of carotenoids will likely be in some ways associated with members of the electron transport chain. Thus, the efficiency of carotenoid ketolation, and therefore the presence or abundance of red ketocarotenoids will likely reflect the efficiency of electron transport (Ge *et al.* 2015). However, demonstration of a plausible links between carotenoid ornamentation and aerobic respiration does not necessarily prove that ornamentation evolved specifically as a biological signal for this process.

Summary

In summary, carotenoids can function as honest signals for quality in a variety of contexts including mate attraction and agonistic signals during Intrasexual competition (Pryke *et al.* 2001). The mechanisms through which carotenoids can signal fall loosely into two categories depending on whether carotenoids act as direct or indirect quality indicators. In this regard, there has been a shift in the way researchers view the relative importance of the two mechanisms. Indeed, the evolution of compensatory mechanisms that reduces the costs necessary to maintain signal honesty during trade-off can be seen in birds whereby lineages that have carotenoids in their plumage also possess higher levels of carotenoids in their plasma (Simons *et al.* 2014). Such a result suggests that increases in plasma carotenoid levels evolved in response to the evolution of carotenoid-based plumage coloration, thereby meeting demand with supply, which challenges the resource trade-off hypothesis in so far as plumage carotenoids are concerned. By contrast, many hypotheses have been put forward for carotenoid-based coloration as direct signals for cellular functions. These are largely based on predicted similarities between the mechanisms for carotenoid metabolism and those found in other cellular systems. Consequently, a more holistic understanding of the genetics of carotenoid metabolism together with the associated cellular systems will be needed to further the understanding of how signal honesty is maintained and ultimately why carotenoids evolved as signals for selection.

Avian colour vision – carotenoids in oil droplets

Visual processing starts with the absorption of photons of light by the visual pigments located in the outer segments of photoreceptors cells in the retina. Two types of photoreceptor cells have been described in vertebrates: rods and cones. Rods typically dominate the retinas of nocturnal animals as they show greater light sensitivity, and are used for vision under dim light conditions. Cones, on the other hand, dominate the retinas of diurnal animals as they can operate under brighter levels of light, and are used for colour vision (Furukawa *et al.* 2014). Importantly, it is the spectral absorption properties of the visual pigments within the rods and cones that determine the wavelengths of light animals can detect. These visual pigments are made up of an opsin protein together with a vitamin A-derived chromophore. Opsins are G-protein coupled receptors composed of around 350 amino acids that form seven α-helical transmembrane regions (Bowmaker 2008). The chromophore is covalently bonded to a lysine residue in the seventh helix, and interactions between the opsin and the chromophore determine the spectral properties of the visual pigment. Different combinations of opsins and chromophores may be used interchangeably, altering the range of wavelengths absorbed (Bowmaker 2008).

The avian retina contains a single type of rod cell together with 4 distinct types of single cones and a single type of double cone. Importantly, it is only the single cones that are used in colour vision, whereas the double cones are involved in non-colour based tasks such as pattern recognition and movement detection. The chromophore that is bound to avian opsins is always 11-cis-retinal, and it is therefore the amino acid sequences of the opsins that determine the spectral absorbance properties of the visual pigment (Hill and McGraw 2006). The cone types that facilitate avian tetrachromatic colour vision are distinguished according to the wavelength of peak absorbance (λ_{max}) of the opsins contained in each cone (Hill and McGraw 2006). The LWS cone contains LWS opsins with λ_{max} values of around 565-570nm. The MWS contains RH2 opsins with λ_{max} values of 500-510nm. SWS cones contain SWS2 opsin with λ_{max} values between 440-460nm, and finally the cones that contains SWS1 opsin are known as VS or UVS cones because the λ_{max} values of the SWS1 opsin appear to fall broadly into two intervals: 400-420nm (VS cones) and 365-375nm (UVS cones) (Bowmaker *et al.* 1997; Hart 2001; Hill and McGraw 2006) (Figure 1.2).

A particular feature of avian cones (single and double cones) is that they possess coloured oil droplets (Bowmaker *et al.* 1993; Wright and Bowmaker 2001). Five types of droplets have been described in birds, four of which are pigmented with carotenoids: LWS cones possess red oil droplets pigmented predominantly with astaxanthin; MWS cones possess yellow droplets pigmented predominantly with zeaxanthin; SWS cones with clear oil droplets pigmented with apocarotenoids including fringillixanthin and galloxanthin; SWS1 opsin containing cones with transparent oil droplets lacking in carotenoids, and finally, double cones (which consist of a principal cone along with an accessory cone) with pale droplets containing a mixture of carotenoids including galloxanthin, lutein and zeaxanthin (Goldsmith *et al.* 1984; Stavenga and Wilts 2014; Toomey *et al.* 2015) (Figure 1.2).

As a result of carotenoid pigmentation, oil droplets function as cut-off filters, preventing certain wavelengths of light from reaching the visual pigments. This filtering process has several consequences: 1) the cut-off wavelength of the oil droplets is often similar to the λ_{max} of the visual pigments, and this has the effect of reducing overall sensitivity of the cones to light, 2) the effective λ_{max} of the photoreceptor shifts towards longer wavelengths (Vorobyev *et al.* 1998), 3) a reduction of the amount of overlap between the spectral sensitivities of the cones, which has the predicted benefit of increasing colour discrimination under sufficient light conditions (Vorobyev 2003) and 4) oil droplets may act as microlenses, effectively channelling the filtered wavelengths of light towards the visual pigments (Stavenga and Wilts 2014). Thus, pigmentation of oil droplets may represent a trade-off between absolute light sensitivity and colour discrimination (Vorobyev 2003; Hart *et al.* 2006), although the predicted channelling effect may to a certain extent compensate for the reduced sensitivity to light (Stavenga and Wilts 2014). By contrast, the VS and UVS cones with transparent unpigmented oil droplets allows transmission of all wavelengths of light, including those in the UV spectrum through to the SWS1 visual pigments.

Given that birds can only obtain carotenoids from their diets, early dietary manipulations found a dramatic effect on retinal carotenoid levels in the Japanese quail, in which carotenoid-free diets resulted in the removal of all detectable carotenoid pigmentation of retinal oil droplets (Bowmaker *et al.* 1993). Notably, retinal carotenoids depleted the slowest in a study of wild house finches compared

with other tissues when the birds were placed on a very low carotenoid diet (Toomey and McGraw 2010). Interestingly, supplementation with astaxanthin, the ketocarotenoid found in the red oil droplets of LWS cones, did not change retinal astaxanthin levels. This suggests that the birds use highly specific mechanisms to regulate carotenoid content in oil droplets, such as via carotenoid metabolism (Toomey and McGraw 2010). Taken together, carotenoids have been shown to play a central role in the avian colour vision, and experimental manipulations highlight the importance of oil droplet pigmentation in birds.

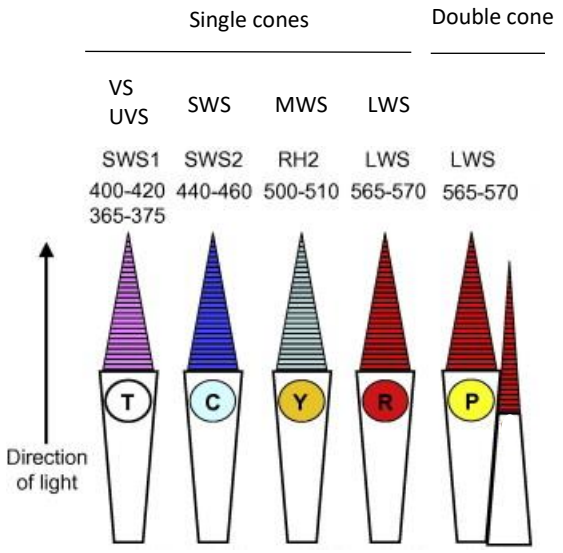

Figure 1.2. Diagram showing the classifications of avian cones. λ_{max} of the opsins and the oil droplets contained within each of the cones are shown. T, C, Y, R and P correspond to transparent, clear, yellow, red and pale oil droplets respectively. Modified image taken from Bowmaker 2008.

Genetic basis of adaptive evolution

Adaptive evolution, or adaptation can be defined as the movement of a population towards a phenotype that best fits its present environment (Orr 2005). Three main types of genetic mutations that contribute to adaptive trait evolution can be distinguished: Coding mutations, Cis-regulatory mutations as well as changes that alter both coding and cis-regulatory regions (e.g. certain gene duplications).

Coding mutations by definition occur in the coding region of the gene, namely genetic sequences that encode the final gene product, whether protein or RNA. In general, for protein-coding genes, it is predicted that non-synonymous mutations, that is mutations that cause a change in the amino acid sequence, play a larger role in phenotypic evolution than synonymous mutations, genetic mutations that do not cause a change in the amino acid sequence. The prediction mainly stems from the very nature of non-synonymous mutations in that they are more likely to affect protein structure, interactions, activity and stability (Stern and Orgogozo 2008) than synonymous mutations. Nevertheless, increasing experimental and comparative studies are demonstrating the ability of synonymous mutations to regulate many aspects of gene expression including transcription (Kudla *et al.* 2009), translation initiation and/or elongation rates (Shabalina *et al.* 2013) and protein folding (Zhang *et al.* 2009), and as a result drive adaptive evolution (Bailey *et al.* 2014).

In contrast, cis-regulatory mutations, defined here as mutations that occur only within cis-regulatory regions (e.g. promotor, enhancer, 5’ and 3’ UTR, introns) that regulate expression of a gene in *cis,* which does not include mutations in coding regions that regulate levels of gene expression. As promoters are needed to initiate basal levels of mRNA transcription in eukaryotes, enhancers are often thought to play a greater role in the control of phenotypic divergence (Wittkopp and Kalay 2012). Arguments that propose that the majority of adaptive phenotypic change results from cis-regulatory mutations reason that mutations in cis-regulatory regions are less constrained and have fewer negative pleiotropic effects than mutations in coding regions (Stern and Orgogozo 2008). Although there is growing experimental evidence to demonstrate the impact of cis-regulatory mutations on the evolution of certain phenotypes (Wray 2007), including on colour signalling (Reed *et al.* 2011), studies of coding sequences, especially of nonsynonymous mutations, are

nevertheless easier to identify and characterise, and as a result, more cases of protein coding changes are known to be involved in phenotypic evolution than cis-regulatory changes (Stern and Orgogozo 2008).

Large scale genetic changes can alter both coding and cis-regulatory sequences and affect phenotypic evolution. An example of such a change is gene duplication, and several models of evolution have been proposed to explain the preservation of new gene copies post duplication. These can be classified based on the selective forces acting on the duplicates immediately after duplication (Innan and Kondrashov 2010). Some of the early and perhaps more well-known models assume that the duplicates confer no selective advantage or disadvantage, and thus evolve, at least initially, through genetic drift. These include the neofunctionalization model, the duplication-degeneration-complementation (DDC) model and the specialisation or escape from adaptive conflict model. The neofunctionalization model assumes that a single gene copy is sufficient to fulfil the function of the gene and therefore duplicate copies would be redundant and accumulate neutral mutations, which are most likely deleterious and results in pseudogenization. However, beneficial mutations may arise within the duplicate, purely by chance, and result in the acquisition of a new gene function which then can then be maintained by selection (Ohno 1970). The problem with this model is that the newly duplicated gene would need to remain in the population for a sufficient amount of time and at sufficiently high frequencies for rare beneficial mutations to accumulate, and in order to explain why higher proportions of duplicates are preserved than would be predicted under neofunctionalisation, the DDC model was put forward (Force *et al.* 1999). The DDC model assumes that the unduplicated gene has more than one function and that following duplication, both copies will accumulate regulatory mutations that impair gene function. The accumulation of such mutations, however, facilitates the preservation of the duplicate copies as the original functions of the unduplicated gene are partitioned between the duplicates via the complementary loss of duplicated cis-regulatory elements (Force *et al.* 1999). In this model, no improvement of the original gene functions occur which contrasts with the escape from adaptive conflict scenario whereby as a result of pleiotropy, the ancestral bifunctional gene cannot improve its functions independently. After duplication, the two functions become divided between the duplicates in a neutral manner and are then driven by

natural selection to improve each of the functions separately (Hughes 1994). Hence gene duplication can provide important sources of innovation and facilitate adaptive evolution (Kondrashov 2012).

To conclude, an understanding of the genetic mechanisms is crucial in the study of the evolution of adaptive phenotypes. Surprisingly, relatively little is known about the genetics underlying sexually selected traits in birds. One exception comes from studies of melanin based pigmentation, where the gene MC1R, which encodes a G-protein couple receptor, has received considerable attention and appears to be responsible for the convergent evolution of melanism in three divergent avian species (Mundy 2005). In this regard, I summarize our current understanding of the genetics of carotenoid based coloration in the next section, which forms the basis for this thesis.

Genetics of carotenoid metabolism in animals

Despite the wide interest in carotenoid pigmentation, relatively little is known about the genetic mechanisms involved in carotenoid utilisation (Walsh *et al.* 2012). This is primarily due to the lack of a model system for the study of carotenoid genetics. Here, I review current understanding of the genes underlying carotenoid uptake and metabolism, two areas in which some molecular functions have been described (Toews *et al.* 2017).

Due to the highly hydrophobic nature of carotenoids, special transport proteins, specifically, class B scavenger proteins (SCARB), are needed to carry carotenoids into the cell (Kiefer *et al.* 2002). Recent studies across divergent animal lineages demonstrate the conserved role of SCARB in carotenoid uptake. The first comes from a study of wild Atlantic salmon where carotenoids obtained from crustacean prey are transported by high density lipoproteins to the midgut and deposited unmodified in the skeletal muscles (Sundvold *et al.* 2011). *SCARB1* was found to be highly expressed in the salmon mid-gut compared with other tissues (liver and skeletal muscle), which is consistent with its role in carotenoid transport (Sundvold *et al.* 2011). Following from this, studies on mutant lines of domesticated silkworm (*Bombyx mori*) identified *SCARB* proteins to be involved in the selective movement of carotenoids, specifically β-carotene and lutein into the silk glands of developing caterpillar (Sakudoh *et al.* 2013). In addition, whole transcriptome comparisons between orange and brown noble scallops (*Chlamys nobilis*), which differ in the concentration of carotenoids, found that only the scavenger receptor like gene (*SRB-like-3*) correlated with the presence of carotenoids in the various tissues (Liu *et al.* 2015), with *SRB-like-3* expression only found in orange but not brown scallops. Functional analysis using RNAi against *SRB-like-3* suppressed expression up to 50% and resulted in a significant reduction in the blood carotenoid concentrations (Liu *et al.* 2015). More recently, a comparative study of the white recessive canary mutant identified coding mutations in *SCARB1* that caused the abnormal splicing of the *SCARB1* transcript, and which appears to be associated with the loss of wild type yellow carotenoid based plumage (Toomey *et al.* 2017). Taken together, these studies independently highlight the role of scavenger proteins in the control of carotenoid uptake.

Two well-known carotenoid cleavage enzymes have been identified in birds and mammals: *BCO1* and *BCO2*. *BCO1* promotes the symmetric cleavage of provitamin A carotenoids such as β- and α-carotene to produce two molecules of retinal (Olson and Hayaishi 1965). In contrast, *BCO2* is responsible for the asymmetric cleavage of carotenes such as β-carotene, and also xanthophyll carotenoids such as lutein and zeaxanthin (Mein *et al.* 2011). *BCO2* is localised in the mitochondria of humans and mice (Amengual *et al.* 2011), and has been described to function in both colour vision and carotenoid based coloration. In the avian retina, *BCO2* is responsible for the initial cleavage of dietary carotenoids into apocarotenoids (compounds derived from the cleavage of carotenoids by carotenoid oxygenases) during the production of galloxanthin, which are found in the clear oil droplets of SWS cones and in the pale oil droplets of double cones (Toomey *et al.* 2016). The function of *BCO2* in coloration has been documented in divergent animal lineages and can involve either cis-regulatory or coding mutations. An example of cis-regulatory changes comes from a study in domesticated chicken where regulatory mutations that inhibit *BCO2* expression in the skin were found to be responsible for the yellow skin phenotype (Eriksson *et al.* 2008). By contrast, mutations in the gene coding sequence was found to cause an increase in the levels of carotenoids in bovine adipose tissue and milk (Tian *et al.* 2010) as well as being responsible for the yellow fat phenotype of sheep (Vage and Boman 2010).

Very recently, independent studies on two mutant bird systems: the 'yellowbeak' zebra finch mutant (Mundy *et al.* 2016) and the 'red factor' canary (Lopes *et al.* 2016), identified a gene, *CYP2J19* as a candidate ketolase responsible for the conversion of yellow dietary carotenoids into red ketocarotenoids. *CYP2J19* is a member of the Cytochrome P450 family of monooxygenases which has diverse roles in a range of cellular systems, including detoxification, hormone and vitamin D metabolism (Werck-Reichhart and Feyereisen 2000). Interestingly, within the zebra finches, two copies of *CYP2J19* were identified: *CYP2J19A* and *CYP2J19B* which show differential tissue specific expression patterns (Mundy *et al.* 2016). *CYP2J19A* is exclusively expressed in the retina, whereas *CYP2J19B* is expressed primarily in the peripheral red tissues (beak and tarsus), but also weakly in the retina. Hence *CYP2J19* appears to have a dual function in colour vision and in ketocarotenoid based red coloration, which fits well with its proposed ketolation role in birds.

In this thesis, I further the investigation on the gene *CYP2J19* as the candidate ketolase responsible for ketocarotenoid pigmentation. First, I consider the evolutionary origin of the gene and suggest its role in colour vision as the conserved ancestral gene function. Next, I consider the evolution of *CYP2J19* across the avian lineage as a whole, and finally, I investigate associations between expression levels and ketocarotenoid pigmentation across two passerine families.

Thesis outline

- **Chapter 2**: General materials and methods
- **Chapter 3**: I investigate the evolution of *CYP2J19* in colour vision and red coloration across the reptiles using a combination of genomic and expression analysis.
- **Chapter 4**: Using expression data and phylogenetic analysis, I investigate the role of *CYP2J19* in explaining the variation in red/yellow male breeding plumage of wild weaverbird species.
- **Chapter 5**: I perform phylogenetic analysis across the avian lineage to determine the selective forces acting on *CYP2J19*. I focus on the weaverbirds from Chapter 3 as a specific example and describe whether selection on coding sequences are associated with the red vs. yellow phenotype.
- **Chapter 6**: I further the investigation of the association between *CYP2J19* expression and red carotenoid based coloration in selected estrildids and viduids and consider the evolution of *CYP2J19* duplication (Mundy *et al.* 2016).
- **Chapter 7**: I discuss the significance of my results and identify areas of potential future investigation.

Chapter 2

General Experimental Methods

Here I describe a common set of experimental procedures that are relevant for several of the following data chapters. Listed below are the methods used for primer design, genomic DNA extraction, long range PCR analysis, RNA extractions, first strand synthesis and qualitative RT-PCR.

Primer design

The available Passeriform genomes on Genbank (*Taeniopygia guttata, Serinus canaria, Ficedula albicollis, Parus major, Geospiza fortis, Pseudopodoces humilis, Zonotrichia albicollis, Manacus vitellinus, Acanthisitta chloris, Corvus brachyrhynchos, Corvus cornix, Zosterops lateralis*) were BLASTn searched for highly similar gene sequences. Sequences were aligned in MEGA version 6 (Tamura *et al.* 2013) using MUSCLE (Edgar 2004) and primers were designed from the appropriate sequence alignments using Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) (Untergasser *et al.* 2012).

Genomic DNA extraction

Genomic DNA was extracted from all avian samples using QIAamp DNA mini kit (QIAGEN) according to the standard protocol. In all species, the livers were used for DNA extraction except in cases where no liver samples were available. Tissues were mechanically disrupted and the DNA was eluted in 200µl Buffer AE.

Long range PCR

Long range PCR was carried out using 1X Extensor Long PCR Master Mix (ThermoFisher) together with 0.4µM of forward and reverse primers to obtain a per reaction volume of 25µl. Reactions were performed in a G-Storm GS1 Thermal Cycler (Life Science Research) according to the following cycle conditions: 94°C for 2 minutes followed by 30-33 cycles of 94°C for 10 seconds, 60-65°C for 30 seconds and 68°C for 8-10 minutes, with a final extension at 68°C for 10 minutes.

Long range PCR amplicons were resolved on 0.6-0.8% agarose gel using Quick-Load 1kb Extend DNA ladder (BioLabs), which allows for identification of band sizes up to 48.5kb.

PCR products were purified using QIAquick PCR Purification Kit according to standard protocol and the DNA was eluted in 30µl of water prior to Next Generation Sequencing.

RNA extraction

Total RNA was extracted from all tissue samples using QIAGEN RNeasy Mini kits. Dissected tissues were placed into 1.5ml microfuge tubes and submerged in liquid nitrogen prior to manual homogenization with an Eppendorf homogenizer and addition of Buffer RLT. The lysate was transferred into QIAshredder spin columns and centrifuged for 2 minutes at 13,000rpm before proceeding with subsequent 3-minute centrifugation according to RNeasy Mini protocol. On-column DNase digestion was carried out using the QIAGEN RNase-Free DNase set.

First strand synthesis and qualitative RT-PCR

First strand synthesis was performed with 10µl total RNA and N_6 primer (0.5µM) using SuperScriptII RT (Life technology Invitrogen) according to the manufacturer's instructions. RT-PCR reactions contained 1 x NH_4 Buffer, $MgCl_2$ (1.5mM), each dNTP (2.5mM), each primer (0.4µM), BioTaq DNA polymerase (Bioline) (0.5U) and cDNA (~50ng).

Qualitative RT-PCR Reactions were run in a G-Storm GS1 Thermal Cycler (Life Science Research) under the following conditions: 2 minutes at 94°C followed by 35–40 cycles of heating for 30 seconds at 94°C, 45 seconds at 58-60°C and 90 seconds at 72°C with a final extension of 5 minutes at 72°C. The amplified full length fragment was purified using ExoSap-IT (Affymetrix) and sequenced on both strands via Sanger sequencing to confirm gene identity.

Chapter 3

Seeing red to being red: conserved genetic mechanism for red cone oil droplets and co-option for red coloration in birds and turtles

This project was conducted in collaboration with Dr Nicole Valenzuela and Dr Robert Literman based at the Department of Ecology, Evolution, and Organismal Biology (EEOB), Iowa State University, and Professor Staffan Anderson at the Department of Biological and Environmental Sciences, University of Gotherburg. Dr Valenzuela and Dr Literman supplied the turtle samples and helped draft the manuscript. Professor Andersson helped conceive the study and edit the manuscript. I designed the experiments, carried out all molecular laboratory work, analysed the data and drafted the manuscript.

Modified version of paper published in Proceedings of the Royal Society B on 3rd August 2016.

Abstract

Avian ketocarotenoid pigments occur in both the red retinal oil droplets that contribute to colour vision and bright red coloration used in signalling. Turtles are the only other tetrapods with red retinal oil droplets, and some also display red carotenoid-based coloration. Recently, the *CYP2J19* gene was strongly implicated in ketocarotenoid synthesis in birds. Here, we investigate *CYP2J19* evolution in relation to colour vision and red coloration in reptiles using genomic and expression data. We show that turtles, but not crocodiles or lepidosaurs, possess a *CYP2J19* orthologue, which arose via gene duplication before turtles and archosaurs split, and which is strongly and specifically expressed in the ketocarotenoid-containing retina and red integument. We infer that *CYP2J19* initially functioned in colour vision in archelosaurs and conclude that red ketocarotenoid-based coloration evolved independently in birds and turtles via gene regulatory changes of *CYP2J19*. Our results suggest that red oil droplets contributed to colour vision in dinosaurs and pterosaurs.

Introduction

Ketocarotenoids are red xanthophyll carotenoids whose terminal ketone groups, when on the C4 position on one or both end rings, lead to the absorption of light at longer wavelengths than in other xanthophylls (Britton 1995). They have been extensively studied in birds where they are the predominant mechanism of saturated red colours that are frequently sexually or socially selected signals (Burley and Coopersmith 1987; Hill 1990; Pryke *et al.* 2002; Ninnes *et al.* 2017). In addition, ketocarotenoids perform another key function in birds: the ketocarotenoid astaxanthin (3,3'-dihydroxy-β-carotene-4,4'-dione) is the red pigment found in the oil droplets in longwave sensitive (LWS) cones of the avian retina (see Introduction). Ketocarotenoids in most landbirds are not obtained directly from the diet but instead synthesised by ketolation (addition of double-bonded oxygen at the C4 position of the end ring) of dietary yellow carotenoids such as zeaxanthin (Britton 1995). The mechanism was for a long time obscure, but recent studies have strongly implicated a locus encoding a cytochrome P450 monooxygenase (*CYP2J19*) as the main or only avian ketolase enzyme that catalyses the conversion of dietary xanthophylls to the ketocarotenoids that are deployed in both red integumentary coloration and red retinal oil droplets (Lopes *et al.* 2016; Mundy *et al.* 2016). This raises the intriguing question of the origin of *CYP2J19* function in birds, and its ancestry in vertebrates more generally.

Turtles (Testudines) are the only group of tetrapods apart from birds that possess red retinal oil droplets (Walls and Judd 1933; Goldsmith *et al.* 1984; Bowmaker 2008), and these likely contain ketocarotenoids (Liebman and Granda 1975; Lipetz 1984). By contrast, other extant reptile lineages either lack retinal oil droplets entirely (crocodiles, snakes) or have coloured oil droplets varying from clear to yellow and green, but not red (many lizards) (Bowmaker 2008). Like birds, turtles have excellent tetrachromatic colour vision (Arnold and Neumeyer 1987; Zana *et al.* 2001).

Red coloration of the integument is uncommon in turtles, but red coloration specifically due to the presence of ketocarotenoids has been reported (Steffen *et al.* 2015) and shown to be related to sexual selection (Ibanez *et al.* 2014). Importantly, aquatic turtles, unlike most landbirds, have the potential to obtain red ketocarotenoids directly from their diet (Clark and Gibbons 1969; Nagaoka *et al.*

2012), and whether turtles have the capacity for carotenoid ketolation remains unknown.

Several interesting questions arise: Is there a shared genetic basis for ketocarotenoid synthesis in birds and turtles? Alternatively, was ketocarotenoid synthesis independently derived in the two groups, or even absent in turtles? If ketocarotenoid synthesis does have a shared genetic basis, was the original function of *CYP2J19* for colour vision or coloration?

The phylogenetic position of turtles within the reptiles has long been a source of debate (Lee 2001) mainly due to the highly derived morphologies of turtles and the somewhat contradictory molecular evidence (Hedges and Poling 1999; Cao *et al.* 2000). More recently, a number of large scale molecular studies using genomic data support a sister relationship between turtles and the archosaurs, forming the turtle-archosaur clade (Archelosauria) independent from the lepidosaurs (lizards, snakes and tuatara) (Shen *et al.* 2011; Chiari *et al.* 2012; Crawford *et al.* 2012; Shaffer *et al.* 2013; Wang *et al.* 2013; Crawford *et al.* 2015). With turtles now placed as an outgroup to archosaurs, the evolution of a locus related to colour vision and coloration in turtles has the potential to inform the occurrence of these traits in the dinosaurs and pterosaurs.

Here I address the genetic basis of ketocarotenoids in the retina and integument of turtles and birds using two approaches. First, I mine recent genomic data and perform phylogenetic analyses to reconstruct the evolutionary history of *CYP2J19* within reptiles and birds. Second, I examine expression of *CYP2J*-like loci in the western painted turtle (*Chrysemys picta bellii*), which possesses both red retinal oil droplets and red ketocarotenoid-based integumentary coloration, to obtain evidence for functional involvement of *CYP2J19* in ketocarotenoid synthesis in different tissues in Testudines. I find that *CYP2J19* is conserved in birds and turtles but has apparently been lost in crocodiles, and that patterns of *CYP2J19* expression are consistent with a function in both colour vision and coloration in turtles, as in birds. We interpret these results to suggest that the carotenoid ketolase *CYP2J19* originally functioned in colour vision in the Testudines-archosaur clade ancestor, and has since been independently and repeatedly recruited for red integumental coloration by changes in gene regulation in turtles and birds. Finally, our results imply the presence of red retinal oil droplets in lineages related to the birds, including dinosaurs and pterosaurs, which provides new insights to the reconstruction of the evolution of colour vision in these groups.

Methods

(a) Data mining and phylogenetic analysis of amniote CYP2J loci

BLASTn searches were conducted in all available reptilian genomes, belonging to the testudine (4), crocodilian (4) and lepidosaur (9) lineages (Table 3.1). We also performed BLASTn searches in the common ostrich (*Struthio camelus*) (Palaeognathae), which together with chicken (Galloanserae) and zebra finch (Neoaves) represent all three major extant avian lineages. Searches were conducted using *CYP2J, CYP2R1* and *CYP2D6* sequences taken from five Ensembl release 83 genomes: zebra finch (*Taeniopygia guttata*), chicken (*Gallus gallus*), human (*Homo sapiens*), anole lizard (*Anolis carolinensis*) and Chinese softshell turtle (*Pelodiscus sinensis*) (Cunningham *et al.* 2015). Nomenclature for avian and reptile *CYP2J* genes was taken from a study on avian CYPs (Watanabe *et al.* 2013). Previous work has shown that there are two lineages of avian *CYP2J* loci, *CYP2J19* and *CYP2J40* (Watanabe *et al.* 2013), whereas humans and other mammals have a single *CYP2J* locus, *CYP2J2*. *CYP2R1* and *CYP2D6* are single copy CYP families that are closely related to *CYP2J* and were used as outgroups (Watanabe *et al.* 2013).

Nucleotide sequences were aligned in MEGA version 6 (Tamura *et al.* 2013) using MUSCLE (Edgar 2004). Phylogenetic reconstructions of protein sequences derived from the nucleotide alignment were performed by maximum likelihood in PhyML-SMS (Smart Model Selection), using model selection based on Bayesian Information Criterion (http://www.atgc-montpellier.fr/phyml/) (Guindon *et al.* 2010). The selected model was JTT+G6 (fixed at 1.997) +I (fixed at 0.089). Branch support was assessed with 1000 bootstrap pseudoreplicates.

Table 3.1. GenBank Accessions (or Ensembl ID) for sequences used in this study.

		CYP genes		
	Species	**CYP2J**	**CYP2D6-like**	**CYP2R1**
Birds	*Taeniopygia guttata*	*CYP2J19A: KX024636*	*XM_002192485.3*	*XM_012573828.1*
		CYP2J40: KX184728		
	Gallus gallus	*CYP2J19: XM_422553.4*	*NM_001195557.1*	*XM_420996.5*
		CYP2J40-like: NM_001195557.1		
	Struthio camelus	*CYP2J19: XM_009689172.1*	*XM_009665288.1*	*XM_009684719.1*
		CYP2J40: XM_009688490.1		
Crocodilians	*Alligator sinensis*	*CYP2J40-like: XM_006036794.2*	*XM_014525977.1*	*XM_006018242.1*
	Alligator mississippiensis	*CYP2J40-like: XM_014598627.1*	*XM_014610884.1 and XM_014610883.1*	*XM_006275041.2*
Testudines	*Chrysemys picta bellii*	*CYP2J19-like: XM_008167974.1*	*XM_005293043.1*	*XM_005303305.1*
		CYP2J40-like: XM_005284738.2		
	Chelonia mydas	*CYP2J19-like: XM_007059062.1*	*XM_007053331.1*	*XM_007057131.1*
	Pelodiscus sinensis	*CYP2J19-like: XM_014578205.1*	*XM_006120509.2*	*XM_006126301.2*
		CYP2J40-like: XM_006133507.2		
Lepidosaurs	*Gekko japonicas*	*CYP2J2-like: XM_015412224.1*	*XM_015426099.1*	*XM_015422757.1*
	Anolis carolinensis	*CYP2J2-like: XM_008109366.1*	*ENSACAG00000003666*	*XM_008107219.1*
Human	*Homo sapiens*	*CYP2J2: NM_000775.3*	*NM_000106.5*	*NM_024514.4*

Species with no Blast hits to *CYP2J* sequences: Crocodilians (*Gavialis gangeticus, Crocodylus porosus*), Testudines (*Apalone spinifera*), Lepidosaurs (*Crotalus mitchellii, Vipera berus, Pantherophis guttatus, Protobothrops mucrosquamatus, Python bivittatus, Thamnophis sirtalis, Ophiophagus Hannah*), Lungfish (*Neoceratodus forsteri*).

(b) Turtle samples

Captive-bred male (N = 3) and female (N = 3) 3-month-old hatchlings of the western painted turtle were obtained by artificial incubation of eggs at constant temperatures that produce a single sex (26°C and 31°C, respectively), as previously described (Valenzuela 2009). At this life stage, red coloration is present in the plastron but has yet to develop on the neck. Individuals were euthanized by an overdose of the anaesthetic propofol, and stored in RNAlater. All procedures were approved by the IACUC of Iowa State University. The following tissues were studied: eye, neck skin, neck muscle, plastron, tail skin and liver.

(c) Primer design

All primers were designed in Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) (Untergasser *et al.* 2012) using the western painted turtle genome (Shaffer *et al.* 2013).

qRT-PCR primers:

PTGAPDHF	CTCCTTTGGCCAAGGTCATC
PTGAPDHR	ATGTTCTGGGCAGCACCTCT
PTHPRT1F	GAAGAGCATGGGGGAACATC
PTHPRT1R	CAACCTGATGAAGTCCACAGTCA
PTTBPF	CAGCAAGCAACACAGGGAAC
PTTBPR	CAGGTGTTGCAGGCGTTATT

Further Primers for *CYP2J19* (full length amplification and sequencing):

PT2ex1F	TGTGAAAAAGAGACGTCCCAAG
PT2ex6R	AGAGCCCAGCGTATGGTTGT
PT2ex7F	GGATGACAAAGAGAATATGCCGTA
PT2ex9R	CGAGAAACAGCACAAATCTGG

Primers for *CYP2J40*:

PT1ex1F	CTGGCAGAACATGCCCTTTC
PT1ex3F	TCCTCTAATGGGCATGTCTGG
PT1ex4R	TTAATTTTAAAGTGAGGGTCAAAAGGT
PT1ex6R	GATCCAGTGTGGAAAATACGAGA
PT1ex7F	AGCGGAGATTGATGCAGTGA
PT1ex9R	TAGCGAGACAGCGCACAGAT

(d) RNA extraction and First strand synthesis

See General Experimental Methods for details.

(e) Quantitative RT-PCR

Quantitative real-time RT-PCR was carried out in an MJ Opticon2 (Research Engines) thermal cycler using the Quantitech SYBRGreen kit (Qiagen) for male and female retina, liver, red plastron, black plastron and yellow neck regions. We used three technical replicates for each condition, and three reference loci (*TBP*, *GAPDH* and *HPRT1*). The geNorm application for the evaluation of expression stability in the control genes was applied to assess the suitability of the reference loci (Vandesompele *et al.* 2002). M values (denoted as the average pairwise variation of a control gene with all other control genes) for *TBP*, *GAPDH* and *HPRT1* were 1.051, 1.085 and 0.991 respectively, indicating suitability for their use. Differences in gene expression among tissues were assessed via Analysis of Variance using the “car” package in RStudio version 3.2 (R Core Team 2009; Fox and Weisberg 2011). The Box-Cox power transformation for normality was applied, and lambda was fixed at 0.3 for subsequent statistical analysis. The Shapiro-Wilk test for normality of residuals ($p > 0.071$), Bartlett’s test of equality of variance ($p > 0.870$) and Runs test for independence of residuals ($p > 0.457$) were upheld.

Results

(a) Evolution of CYP2J genes in amniotes

BLASTn searches revealed the presence of two *CYP2J* family loci in the three most complete turtle genomes out of the four available (western painted turtle, Chinese soft-shell turtle, and green sea turtle) (Shaffer *et al.* 2013; Wang *et al.* 2013). In contrast, we found at most a single *CYP2J* sequence in crocodilians (American alligator, Chinese alligator) and lepidosaurs (green anole, Schlegel's Japanese gecko). Strikingly, phylogenetic analyses revealed that one of the *CYP2J* loci in turtles is orthologous to avian *CYP2J19*, since turtle sequences form a well-supported (100% bootstrap support) monophyletic clade with *CYP2J19* sequences from birds (Figure 3.1 and Figure 3.2). No *CYP2J19* orthologues were identified in either crocodilians or lepidosaurs. The remaining reptile and avian *CYP2J* loci group together by order but support for other relationships is variable. Reconstructions suggest that the crocodilian and second turtle *CYP2J* locus are related to avian *CYP2J40*, but the bootstrap support is weak. The single *CYP2J* locus in lepidosaurs is placed as an outgroup to the Testudines and archosaur (crocodile + bird) sequences, with moderate bootstrap support (60%). The single human *CYP2J* locus (*CYP2J2*) is clearly an outgroup to all of the reptile and avian sequences (99% bootstrap support).

Overall, therefore, we find strong evidence for the conservation of the *CYP2J19* locus in avian and turtle (Testudines) lineages, but no evidence for the presence of *CYP2J19* in crocodilians or lepidosaurs. Furthermore, our phylogenetic analysis indicates that *CYP2J19* arose from a duplication event that occurred after the divergence of reptiles and mammals and before the split between Testudines and archosaurs and *CYP2J19* was subsequently lost in crocodilians.

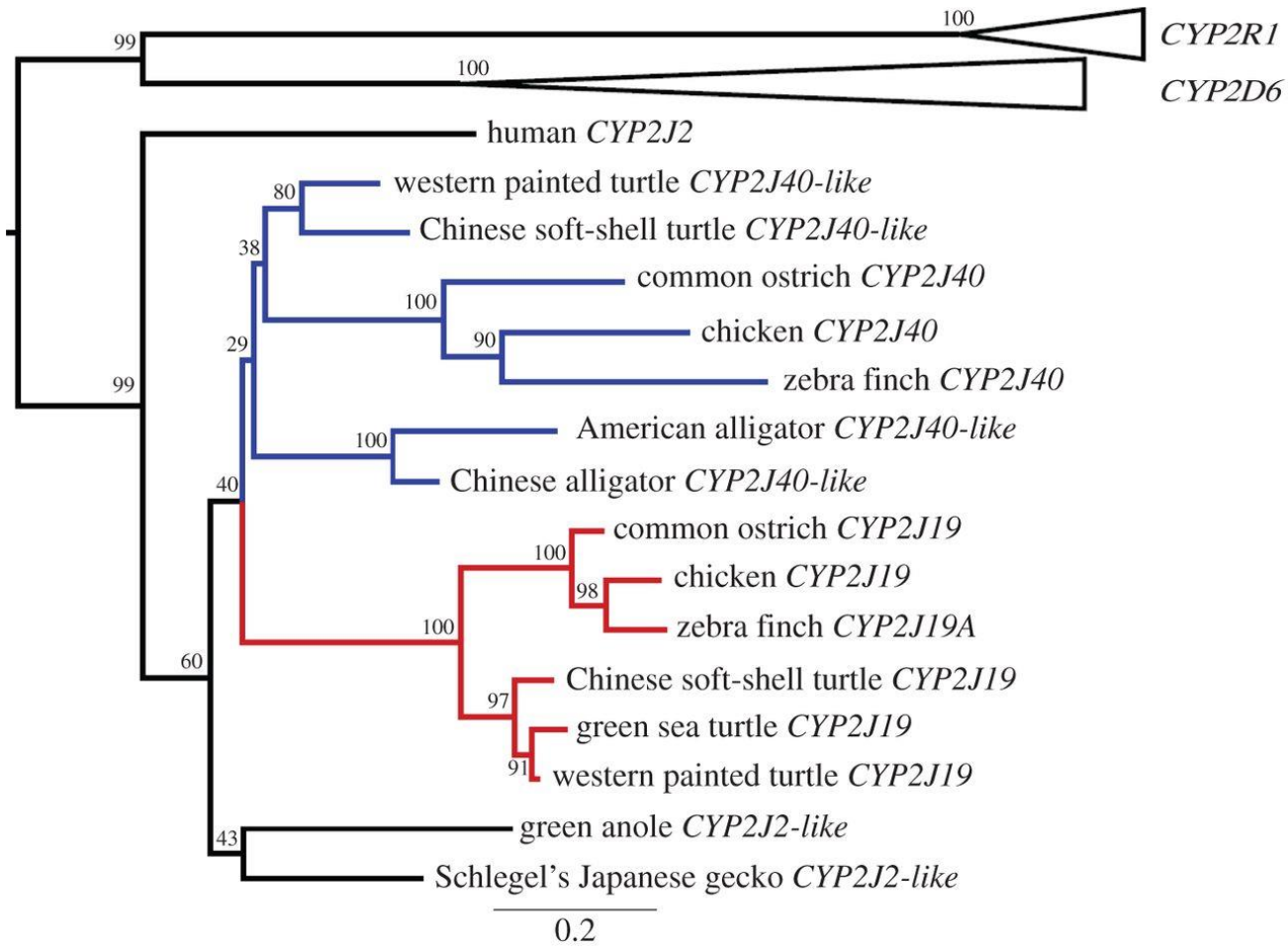

Figure 3.1. JTT model-based protein phylogeny based on all *CYP2J*, *CYP2R1* and *CYP2D6* sequences obtained. Bootstrap values are based on 1000 pseudoreplicates. *CYP2J19*-like sequences and *CYP2J40*-like sequences are outlined in red and blue, respectively. *CYP2R1* and *CYP2D6* clades have been collapsed.

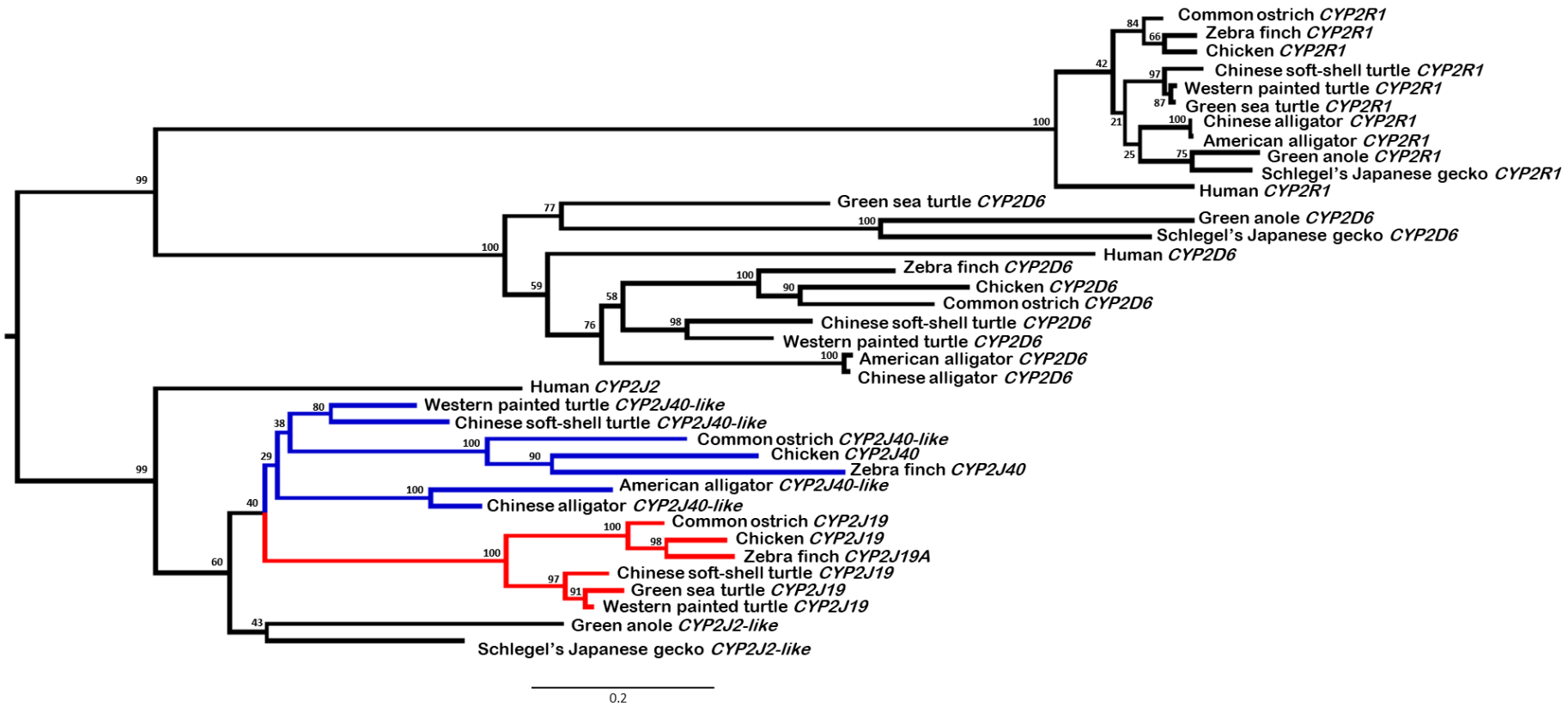

Figure 3.2. Fully expanded JTT model-based protein phylogeny of *CYP2J*, *CYP2R1* and *CYP2D6* from Figure 3.1. Bootstrap values represent 1000 pseudoreplicates. *CYP2J19*-like sequences and *CYP2J40*-like sequences are outlined in red and blue respectively.

(b) Specific expression of CYP2J19 in tissues containing red ketocarotenoids in the western painted turtle

Preliminary experiments revealed high *CYP2J19* expression in the retina and red plastron of western painted turtles of both sexes, and lower, variable expression in the liver, black plastron, yellow tail, black tail, yellow neck and black neck regions (not shown). In contrast, *CYP2J40* expression was detected in all tissue types except for the plastron (not shown).

Quantification of expression via qRT-PCR confirmed the presence of high *CYP2J19* expression in the ketocarotenoid-containing tissues (retina, red plastron) of both sexes, and low expression in other tissues (Figure 3.3). There was significant heterogeneity of *CYP2J19* expression level among different tissue types ($F_{4,25}$ = 35.83, $p < 0.001$) but not among sexes ($F_{1,28}$ = 0.4524, $p = 0.51$) (Table 3.2). As no sex-biased expression was detected, a comparison between all tissue types pooled for both sexes was carried out. The red plastron had the highest *CYP2J19* expression relative to all other tissues ($p < 0.05$), while the retina had higher expression when compared to all tissues ($p < 0.001$) except for the red plastron (Table 3.2).

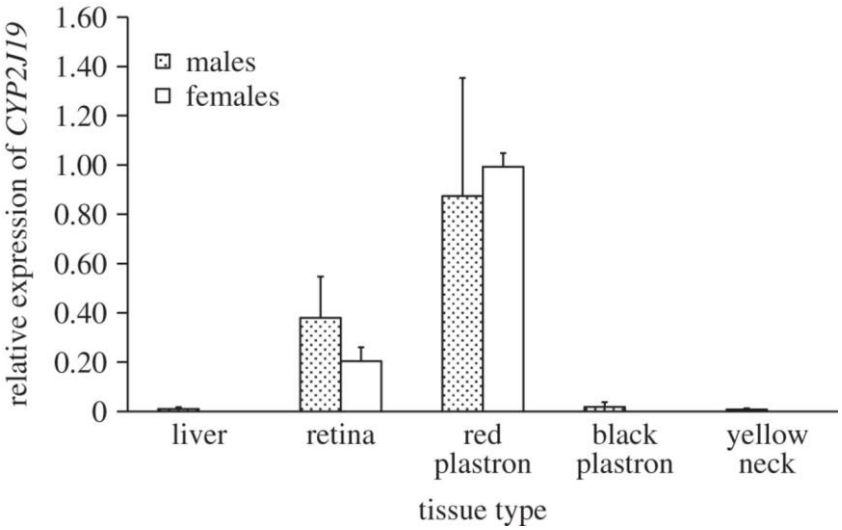

Figure 3.3. Results of qRT-PCR experiments quantifying male and female tissue-specific expression for *CYP2J19* normalised against *TBP*, *GAPDH* and *HPRT1*. Error bars represent SEM from three individual males and females

Table 3.2. Tukey's pairwise tests of *CYP2J19* expression

Pairwise tissue comparisons	p adj
Liver – Retina	0.0000420 ***
Yellow neck – Retina	0.0000284 ***
Black plastron - Retina	0.0000188 ***
Red plastron – Retina	0.0434566 *
Yellow neck – Liver	0.9998532
Black plastron – Liver	0.9975022
Red plastron – Liver	0.0000000 ***
Black plastron – Yellow neck	0.9998199
Red plastron – Yellow neck	0.0000000 ***
Red plastron – Black plastron	0.0000000 ***

Discussion

We provide compelling evidence for the presence of a *CYP2J19* orthologue in three divergent Testudine lineages that suggests conservation of *CYP2J19* function in turtles and birds, despite ~250 million year divergence between these lineages (Chiari *et al.* 2012). Expression patterns of *CYP2J19* in painted turtles are consistent with this gene playing a role in ketocarotenoid synthesis in both retina and red-coloured integument, as recently described in birds (Lopes *et al.* 2016; Mundy *et al.* 2016). This raises the question of whether the original function of *CYP2J19* was for colour vision or coloration. Consideration of the likely ancestral traits, given the extant distributions, strongly favours colour vision as the original function, as we now discuss (Figure 3.4).

Colour vision involving ketocarotenoid-containing oil droplets in LWS cones appears to be a pervasive feature in almost all birds and turtles (Liebman and Granda 1975; Goldsmith *et al.* 1984; Lipetz 1984), implying that red oil droplets were present in both ancestral birds and ancestral turtles. In contrast, red ketocarotenoid coloration has a patchy distribution in both groups. In turtles, red coloration is sparsely distributed amongst both the pleurodirans and cryptodirans. Redness appears to be rare or absent in most turtle families except for two cryptodiran families, Emydidae and Geomydidae, where red coloration is more common, as well as several taxa within the pleurodiran family Chelidae. In addition, the pigmentary basis for redness in turtles has not been widely characterised. In birds, reconstruction of carotenoid-based plumage coloration has shown that this trait is not ancestral but was acquired multiple times during avian evolution (Thomas *et al.* 2014). Evolutionary reconstructions of carotenoid coloration in bare body parts (such as bills) in birds are still awaited. An independent line of evidence suggesting that carotenoid coloration in the integument was not ancestral in birds concerns their lack of specific pigment cells containing carotenoids. Reptiles, like amphibians and teleosts, express carotenoids (along with pteridines) in specialised dermal xanthophores/erythrophores (Hadley 1972), whereas birds express integumentary carotenoids in epidermal keratinocytes. Although studies have found pigment cells in the avian iris that resemble reptilian xanthophores and iridophores (Oliphant 1987; Tillotson and Oliphant 1990), the absence of specialised pigment cells for

carotenoids in the integument of birds seems to imply a period in their ancestry when carotenoids were not used for coloration.

Since tetrachromatic colour vision involving red retinal oil droplets was likely present in the ancestral turtle and avian lineages, while red carotenoid-based coloration was likely absent in both, we conclude that the ancestral function of *CYP2J19* was for production of ketocarotenoids in red retinal oil droplets and that the role of *CYP2J19* in red carotenoid-based coloration evolved independently in the two lineages (Figure 3.4). Many examples of phenotypic convergence via similar genetic changes have been observed across the vertebrates (Wood *et al.* 2005; Stern 2013; Storz 2016). Several examples involve melanin pigmentation, and include the *MC1R* locus which has been involved in convergent evolution of pale/dark coloration in birds and reptiles (Mundy 2005; Hubbard *et al.* 2010; Rosenblum *et al.* 2010). Here, we highlight a remarkable example of convergent evolution whereby *CYP2J19* would have been independently recruited in the bird and turtle lineages, through changes in gene expression, to function in ketocarotenoid-based coloration from an ancestral retina-specific function. The ecological driver for this convergence in many instances may have been sexual selection (Prager and Andersson 2010; Ibanez *et al.* 2014).

Our results imply that red retinal oil droplets themselves were present in the common ancestor of the archelosaurs (archosaurs and turtles) (Crawford *et al.* 2015), so that red oil droplets had a single origin in tetrapods (Figure 3.4). An alternative would be if *CYP2J19* originally arose for a different function than red oil droplets, but this is unlikely since no other functions for ketocarotenoids have been described. The adaptive function of red retinal oil droplets in these early reptiles is an interesting and open question. Their acquisition presumably led to increased colour discrimination at long wavelengths, particularly in bright light (Vorobyev *et al.* 1998; Vorobyev 2003), but the ecological context for this is obscure.

The inferred timing of the gene duplication that produced *CYP2J19* after the split with lepidosaurs is concordant with the absence of red oil droplets in this clade. The apparent absence of *CYP2J19* in crocodilian genomes could be explained by the derived loss of all oil droplets from the retina that has occurred in crocodilians, possibly due to nocturnality (Walls 1942), as in snakes where an ancestral period of nocturnality may have resulted in loss of retinal oil droplets (Ali and Klyne 1985; Schott *et al.* 2016). Although we cannot completely rule out the presence of a divergent *CYP2J19* locus in crocodiles and/or lepidosaurs which was not identified in our genome searches, it is notable that our methods did successfully identify one

CYP2J locus, as well as *CYP2D6* and *CYP2R1* in two crocodilian and two lepidosaur genomes (Figure 3.2). The function of *CYP2J40* is unknown, but the observed expression amongst all the tissues studied (in both turtles and birds (Mundy *et al.* 2016)) does not suggest any direct involvement of this locus in red coloration.

As birds are phylogenetically nested within the saurischian dinosaurs (Gauthier 1986), the continuous presence of red retinal oil droplets in the avian lineage since the split with Testudines implies the presence of red oil droplets in the retina of dinosaurs and other lineages (including pterosaurs) that are more closely related to the living birds than crocodiles. Building on the previous inference of colour vision in dinosaurs based on four cone opsin types (Sillman *et al.* 1991), the presence of red retinal oil droplets in dinosaurs suggests that they possessed similar tetrachromacy to the birds and turtles, with good colour discrimination at long wavelengths. This is concordant with the proposed use of coloration in intraspecific signalling in dinosaurs, where putative melanosomes have been used to infer melanin-based coloration and iridescence (Li *et al.* 2010; Li *et al.* 2012; Foth *et al.* 2014; Koschowitz *et al.* 2014). The co-option of *CYP2J19* for red ketocarotenoid coloration in dinosaurs, as has occurred in birds and turtles, may also have been possible.

The implied presence of a functional carotenoid ketolase in turtles, which are primarily aquatic, is interesting in view of the availability of ketocarotenoids in potential prey items in aquatic environments, e.g. crustaceans, which means that in principle turtles might have lost ketolase activity. If true, this may be a legacy of their terrestrial ancestry in basal archelosaurs. Another possibility is that tight control of carotenoid concentration in retinal oil droplets is required for good colour vision, and that there is too much variation in dietary ketocarotenoids to rely on this source. In this context, it is interesting to note that whereas, as expected, severe dietary carotenoid deprivation is harmful to colour vision in birds, dietary supplementation of carotenoids does not affect the carotenoid content in oil droplets used for colour vision (Bowmaker *et al.* 1993; Knott *et al.* 2010; Toomey and McGraw 2010).

In birds, the anatomical site of ketolation for red coloration has long been contentious but the pattern of *CYP2J19* expression has now shown that it is variable among taxa. Some species, e.g. zebra finch, perform ketolation in the peripheral tissues such as beak where ketocarotenoids are deposited (McGraw 2004; Mundy *et al.* 2016), whereas in other species such as ploceids, ketolation occurs centrally in the liver, and ketocarotenoids are transported in blood to sites of deposition (Twyman

et al., in review – Chapter 4). In turtles, the presence of *CYP2J19* expression in red integument, but not in liver, is indicative of the peripheral conversion model in this group.

Apart from birds and turtles, the only other vertebrate in which red retinal oil droplets have been reported is the lungfish *Neoceratodus* (Bailes *et al.* 2006). Microspectrophotometry measurements on these red droplets showed that they had similar spectral properties to the red droplets of birds (Hart *et al.* 2008). However, acquisition of red oil droplets in this lineage was likely independent to that of archelosaurs since *CYP2J19* arose a long time after the split between lungfish and tetrapods, and this is supported by our failure to find a *CYP2J19-like* sequence in the *Neoceratodus* genome using BLAST searches.

In summary, we have uncovered a remarkable case where a gene with a strongly conserved function in colour vision has been independently co-opted for red ketocarotenoid-based integumentary coloration in turtles and birds via changes in patterns of gene expression. Since *CYP2J19* arose within reptile evolution, the genetic basis for red ketocarotenoid coloration in amphibians and ray-finned fish warrants further research.

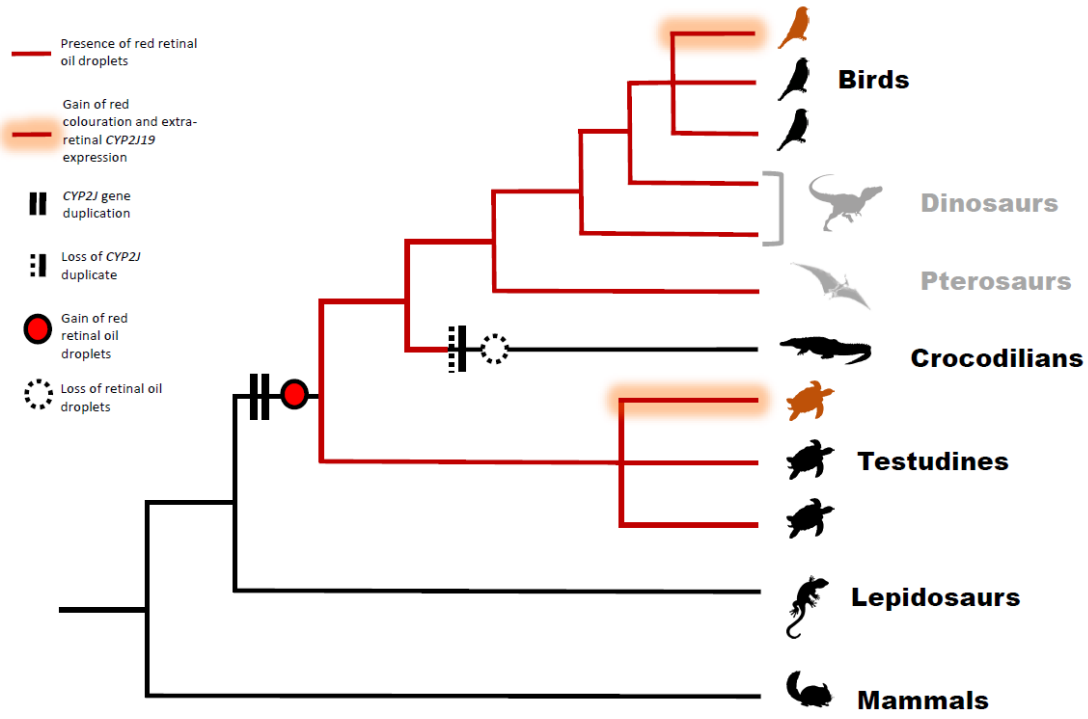

Figure 3.4. Reconstructed scenario for evolution of *CYP2J19* function in red retinal oil droplets and red ketocarotenoid integumentary coloration in reptiles. Red lines denote the presence of red retinal oil droplets. The inferred presence of ketocarotenoid-containing retinal oil droplets coincides with the inferred duplication of *CYP2J2* prior to the turtle–archosaur split, followed by subsequent loss of oil droplets and *CYP2J19* in the crocodilian lineage. Red highlighted branches show the independent gain of red ketocarotenoid-based coloration in certain turtle and bird lineages associated with co-option of *CYP2J19* expression for coloration.

Chapter 4

CYP2J19 underlies convergent evolution of sexually-selected red carotenoid coloration in weaverbirds (Ploceidae)

This project was conducted in collaboration with Professor Staffan Andersson and Dr Maria Prager based at the Department of Biological and Environmental Sciences, University of Gothenburg. Professor Andersson collected a majority of the avian samples and together with Dr Prager performed the test for phylogenetic association between *CYP2J19* expression and ketocarotenoid presence.

Abstract

Red carotenoid colours in birds are widely assumed to be sexually selected quality indicators. Recently, the gene *CYP2J19* has been implicated as the avian carotenoid ketolase, catalysing the synthesis of red C4-ketocarotenoids from yellow dietary precursors, and potentially a major mechanism behind red coloration in birds. Here I investigate the role of *CYP2J19* in the variation of red/yellow male breeding plumages of 16 species of weaverbird (Ploceidae): eight red, seven yellow, and one without carotenoid coloration. All species were found to possess a single copy of *CYP2J19*, unlike the duplication in the zebra finch, with high expression in the retina, confirming the function to colour red oil droplets. Expression was weak or undetectable in the skin, feather follicles, beak and tarsus, including the red C4-ketocarotenoid-pigmented beak and tarsus of the red-billed quelea. Hepatic (liver) expression of *CYP2J19* was substantially higher (greater than 14 fold) in the red C4-ketocarotenoid-pigmented species compared to the rest (including one red species), an association strongly supported by a phylogenetic comparative analysis. The results support a critical role of the candidate ketolase, CYP2J19, in the evolution of red C4-ketocarotenoid colour variation in ploceids. Since ancestral state reconstruction suggests that ketocarotenoid coloration has evolved twice in this group (once in *Euplectes* and once in the *Quelea*/*Foudia* clade), I argue that while CYP2J19 has retained its ancestral role in the retina, it has likely been co-opted for red coloration independently in the two lineages, via increased expression in the liver.

Introduction

Bright red colours in birds are usually carotenoid-based and have been extensively studied as sexually selected signals for quality (see Introduction Chapter). Despite this, the presence of red coloration has a surprisingly patchy distribution across the avian phylogeny. Even in clades where redness is relatively common, such as in cardueline finches (Ligon *et al.* 2016), New World blackbirds (Friedman *et al.* 2014) and widowbirds and bishops (Prager and Andersson 2010), it is often absent in one or more lineages with no obvious ecological or behavioural dissimilarities from their red-coloured relatives. This appears to suggest that there may be some genetic constraints on the evolution of red carotenoid coloration, and that the ketolation ability of the recently proposed candidate ketolase, *CYP2J19*, likely constitutes such a constraint. The two studies that described the association between *CYP2J19* and ketocarotenoid pigmentation differ in two respects, 1) *CYP2J19* gene copy number (two in zebra finch and one in the red factor canary, in which red coloration was introgressed from the red siskin) and 2) tissue-specific expression (’peripherally’ in the integument in zebra finch and both ’peripherally and centrally’ in the feather follicles and liver in the ‘red factor’ canary). There are thus many remaining questions concerning the generality, nature and location of the *CYP2J19* mechanism and function in red coloration. In particular, whether differential *CYP2J19* expression can account for variation in red coloration across an avian clade remains unknown.

Weaverbirds (Aves: Ploceidae, 116 species) are a clade of predominantly African, seed-eating passerines, which are ideal for studying the mechanisms and evolution of carotenoid coloration. Whereas conspicuous yellow plumage colours dominate, especially in the more species rich lineages of ‘true weavers’ (Ploceus spp), red carotenoid coloration occurs in several genera, and some lineages lack integumentary carotenoid pigmentation altogether. The underlying mechanisms (e.g. dietary vs metabolically derived pigments) have been established for several species of widowbirds and bishops (*Euplectes*) (Andersson *et al.* 2007; Prager *et al.* 2009) where red colour hues function as agonistic (threat) signals in male competition. Redness appears to have evolved at least twice from a yellow coloured ancestor (Prager and Andersson 2010), and apparently driven by a pre-existing receiver bias (Ninnes *et al.* 2017). Outside *Euplectes*, weaverbird colour signalling functions are largely unexplored, except in the red-billed quelea (*Quelea quelea*), where the red

beak likely is sexually selected (through either female mate choice or male-male contests) whereas the polymorphic red plumage coloration may be involved in individual recognition (Dale 2000).

In most ploceids where it has been analysed, red colour patches contain red C4-ketocarotenoids, primarily α-doradexanthin and canthaxanthin, co-deposited with the dietary yellow pigments (Andersson *et al.* 2007; Andersson & Prager unpublished). By comparisons with phylogenetically and ecologically closely related yellow-coloured species, this provides an excellent opportunity to test the significance of *CYP2J19* for red carotenoid coloration. Moreover, in some species, such as in the fan-tailed widowbird (*E. axillaris*), red coloration has been found to be achieved without C4-ketocarotenoids (Andersson *et al.* 2007; Prager *et al.* 2009), which provides an additional test of the proposed function (C4-ketolation) of *CYP2J19*.

In this study of 16 red or yellow weaverbird species, we investigate the role of *CYP2J19* in the evolution of carotenoid pigmentation in weaverbirds. First, we establish whether the gene is present in ploceids and, if so, in the gene copy number. Second, we identify the anatomical site(s) of *CYP2J19* expression in this group. Finally, we determine whether *CYP2J19* expression is associated with the occurrence of red C4-ketocarotenoid pigmentation across the ploceids.

Materials and Methods

(a) Samples

Tissue samples from male ploceids in breeding plumage were largely obtained from natural populations in Africa, in addition to a few samples from aviaries in southern Spain and Sweden (Table 3.1), under all applicable national and international permits. Euthanised birds were freshly dissected and tissues placed in RNAlater (Qiagen) or DNA/RNAshield (Zymo) until DNA/RNA extraction.

Table 3.1. List of all tissue samples used in molecular analysis. B = beak, F= feather, Ff = feather follicle, L = liver, R = retina, S+Ff = skin + feather follicle, T = tarsus. ‡ Samples used for qRT-PCR normalised against 3 reference loci (*β-Actin, GAPDH* and *HPRT1*). † Samples used for qRT-PCR normalised against 1 reference locus (*β-Actin*).

Species	Red C4-ketocarotenoid pigmentation	Individual	Tissue	Origin	Date	Sourced by
Philetairus socius	×	1	††‡L	Benfontein Nature Reserve, South Africa	Nov-2014	Staffan Andersson
Ploceus subaureus	×	1	††‡L	Salima, Malawi	Dec-2014	Staffan Andersson
Ploceus melanocephalus	×	1	††‡L, R	Sanlúcar la Mayor, Spain	Sep-2012	
Ploceus capensis	×	1	††‡L, R, B	Port Elizabeth, South Africa	Oct-2014	Staffan Andersson
Ploceus velatus	×	1	††‡L, R, B	Port Elizabeth, South Africa	Oct-2014	Staffan Andersson
		2	L, R	Blouberg Nature Reserve, Limpopo, South Africa	Oct-2013	Fuchs/Wogan
Foudia madagascariensis	✓	1	††‡L	Assumption Island, Seychelles	Apr-2012	Staffan Andersson
		2	†L	Assumption Island, Seychelles	Apr-2012	Staffan Andersson
		3	†L, R, B	Mahé, Seychelles	Nov-2014	Staffan Andersson
		4-5	L, R, B	Mahé, Seychelles	Nov-2014	Staffan Andersson
Quelea erythrops	✓	1	††‡L, S+F	São Tomé, S.T. and Príncipe	Jan-2008	Staffan Andersson
Quelea quelea	✓	1-2	†L, †R, †B, †T	Zambia	Nov-2012	Claire Spottiswoode
		3	††‡L, †R, †B, †T	Zambia	Nov-2012	Claire Spottiswoode
Euplectes afer	×	1	††‡L, R	Sanlucar la Mayor, Spain	Sep-2012	
		2	S+F	Unknown (commercially obtained)	May-2006	Staffan Andersson
		3	S+F	Unknown (commercially obtained)	Feb-2006	Staffan Andersson
Euplectes aureus	×	1	††‡L, S+F	São Tomé, S.T. and Príncipe	Nov-2007	Staffan Andersson
		2-3	F	São Tomé, S.T. and Príncipe	Nov-2007	Staffan Andersson
		4	F	São Tomé, S.T. and Príncipe	Jan-2008	Staffan Andersson & Nicholas I Mundy
Euplectes axillaris	×	1	††‡L, R	Pietermaritzburg, South Africa	Oct-2013	Staffan Andersson
		2	L, R	Pietermaritzburg, South Africa	Oct-2013	Staffan Andersson
Euplectes macrourus	×	1	†L	Buea, Cameroon	Jul-2012	Staffan Andersson
		2	††‡L, R, B	Choma, Zambia	Nov-2012	Claire Spottiswoode
		3	†L, R, B	Choma, Zambia	Nov-2012	Claire Spottiswoode
Euplectes ardens	✓	1	†L	Iringa, Tanzania	Feb-2011	Staffan Andersson
		2	†L	KwaZulu-Natal, South Africa	May-2006	Staffan Andersson
		3	††‡L, R	Cedara, South Africa	Oct-2013	Staffan Andersson
		4	R	Cedara, South Africa	Oct-2013	Staffan Andersson
		5-6	F	Iringa, Tanzania	Feb-2011	Staffan Andersson
Euplectes hordeaceus	✓	1	‡L	São Tomé, S.T. and Príncipe	Nov-2007	Staffan Andersson
		2	†L, F	São Tomé, S.T. and Príncipe	Nov-2007	Staffan Andersson
		3-4	S+F	São Tomé, S.T. and Príncipe	Nov-2007	Staffan Andersson
Euplectes nigroventris	✓	1	††‡L, S+F	Unknown (commercially obtained)	Apr-2006	Staffan Andersson
Euplectes orix	✓	1	R	Cedara, South Africa	Oct-2013	Staffan Andersson
		2-3	†L	Cedara, South Africa	Oct-2013	Staffan Andersson
		4	††‡L, B, T	Cedara, South Africa	Oct-2013	Staffan Andersson

(b) Primer design

See General Experimental Methods (Chapter 2) for details of primer design.

Primers used for amplification of cDNA (5'-3')

β-actin	◆*ACTF2	CTCCCTGATGGCCAGGTCAT
	◆*ACTR2	TGGATACCACAGGACTCCAT
GAPDH	◆GAPDF1	GGGTAGTGAAGGCTGCTGCT
	◆GAPDR1	ATTCAGTGCAATGCCAGCAC
HPRT1	◆HPRT1F	GACCTCTCAACCTTGACTGGAAA
	◆HPRT1R	GCCACTTTCACCATCTTTGGA
CYP2J19	CJ2ex1EF	GAGAAGATCTAAAATGGAGCTACAGTTTTGGCCTGA
	◆*CYP2J2-2R	TGGTAGTCAAAGCGGTTCCCA
	◆*CYP2J2-2F	CGGCCAAATATTCCACTCCTCC
	*CJ2Aex9R	GCTGAACCTGGGTTGATGGGA
	FGCJ2A1ex9R	GAGCCCTTTCCCTACCACCC
	FGCJ2A2ex9R	GGACACCTCAGGAGCCCTTG
	PLO1ex9R	GTGGTCAGGGGAAGGGGATT
	PLO2ex9R	TTCCCGGTTTGTGAGGCTGC
	ESTex9R	CCTACCACCCAGATGCCGG

Primers used for amplification of *CYP2J19* from genomic DNA (5'-3')

*CJ2up9F	CACACCAGGAAGGCTGGAAGA
CJ1Pex1R	GTCTGGAGGGAGATGCTGTCC
CJ1Dex1R	GGAATGTCTGGAGGGAGATGCT
CJ2upsF	GCCTGCTATCCTGATTTAGCC
*CJ2ex1R2	GGAAGGGGGTGGGAGGGAAATTA

* Primers taken from Mundy *et al.* 2016
◆ Primers used in qRT-PCR

(c) CYP2J19 gene copy number determination

For genomic DNA extraction and long range PCR, see General Experimental Methods (Chapter 2). Illumina MiSeq sequencing of PCR amplicons was performed to >1,000-fold coverage at the University of Sheffield and *de novo* assembly was conducted using Seqman NGen (Linux) v.12 (DNASTAR) for *Q. quelea* and *E. orix*.

(d) CYP2J19 expression

RNA extraction and first strand synthesis methods are outlined in General Experimental Methods (Chapter 2).

Quantitative real-time RT-PCR was carried out in an MJ Opticon2 (Research Engines) thermal cycler using the Quantitech SYBRGreen kit (Qiagen). Three technical replicates were used for each condition. Tissue specific *CYP2J19* expression differences in 4 tissue types of *Q. quelea* (N = 3) were quantified against *β-Actin.* Statistical analysis was assessed using Analysis of Variance under the 'car' package in RStudio v. 3.2 (R Core Team 2009; Fox and Weisberg 2011). The Box-Cox power transformation for normality was applied, and lambda was fixed at 0.1. The Shapiro-Wilk test for normality of residuals ($p > 0.8$), Bartlett's test for equality of variance ($p > 0.1$) and Runs test for independence of residuals ($p > 0.06$) were upheld.

For phylogenetic comparative tests, normalisation was done using the geometric mean of three reference loci (*β-Actin, GAPDH* and *HPRT1*) (Pfaffl 2001). The geNorm application for the evaluation of expression stability in the control genes was applied to assess the suitability of the reference loci (Vandesompele *et al.* 2002). The M values (denoted as the average pairwise variation of a control gene with all other control genes) for *β-Actin, GAPDH* and *HPRT1* were 1.185, 1.142, and 1.107 respectively, indicating suitability for their use.

(e) Association between CYP2J19 and red ketocarotenoid pigmentation

To account for phylogenetic non-independence between weaverbird species, the association between hepatic *CYP2J19* expression and red ketocarotenoid pigmentation was assessed using the discrete Markov chain Monte Carlo (MCMC) method in BayesTraits V2 (Pagel 1994; Pagel *et al.* 2004, www.evolution.rdg.ac.uk), while sampling from 10,000 Ploceidae trees downloaded from BirdTree.org (Jetz *et al.* 2012, 'Ericson All Species' source).

Log transformed normalized values of liver *CYP2J19* expression were first discretized using k-means clustering in R version 3.3.1 (R Core Team 2016) with two cluster centres ('high' and 'low'), excluding *E. aureus*, *E. axillaris* and *E. macrourus* where expression was undetectable.

The presence or absence of integumentary C4-ketocarotenoid (KC) pigments was scored based on published data from six of the included species: *Euplectes ardens*, *E. axillaris*, and *E. macroura* (Andersson *et al.* 2007), *E. afer* and *E. orix* (Prager *et al.* 2009) and *Quelea quelea* (Walsh *et al.* 2012). For the remaining species, C4-ketocarotenoid presence or absence was determined by Staffan Andersson and Maria Prager from unpublished high-performance liquid chromatography (HPLC) analyses of carotenoids extracted from feathers.

In BayesTraitsV2, an 'independence model' estimating four separate evolutionary rates (gain and loss for each trait) was compared to a 'dependence model' which allowed for a maximum of eight separate rates (Figure 3.1). In the independence scenario, we assume that changes in one trait do not affect changes in the second trait, such that expression level of *CYP2J19* is independent from presence or absence of ketocarotenoids. This contrasts with the dependence model where we assume ketocarotenoid pigmentation is contingent on high liver expression of *CYP2J19*. Hence, under this assumption, only three possible states exist: 1) high expression, presence of ketocarotenoids, 2) high expression, absence of ketocarotenoids, and 3) low expression, absence of ketocarotenoids (Figure 3.1). Marginal likelihoods of alternative models were approximated using harmonic means of log likelihoods from 1 million generations of discrete Markov Chain Monte Carlo (MCMC) runs, discarding the first 100,000 generations as burn-in, and compared using Bayes Factors (Kass and Raftery 1995). As MCMC methods can be sensitive to the choice of priors, we tested five different prior distributions on transition rates: 1) the default setting of uniform (0, 100), 2) a conservative prior of uniform (0,1) which

favors the null model (independence model), and 3-5) exponential distributions centered at 0.1 and 1, which favors the null, and 10, which give some prior advantage to our hypothesis (dependence model).

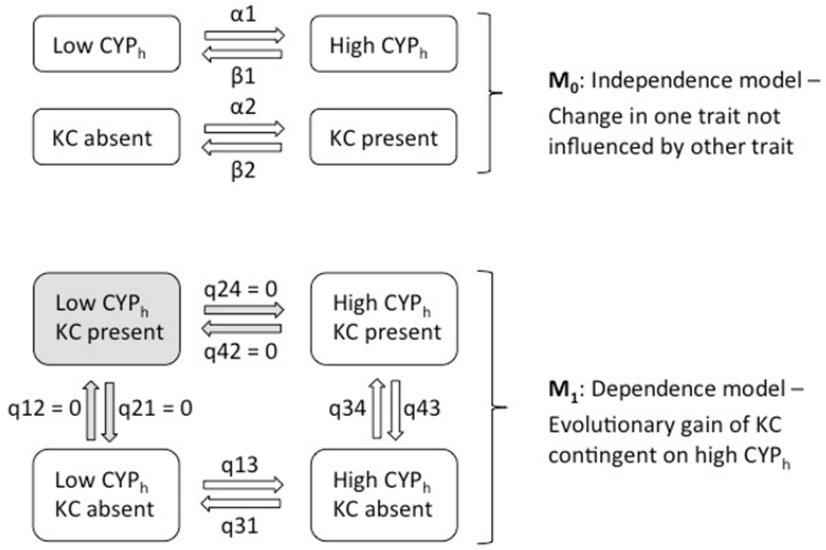

Figure 3.1. Alternative models for the evolution of hepatic *CYP2J19* expression (CYP_h) and red ketocarotenoid pigmentation (KC) in weaverbirds. Arrows show evolutionary transition rates that were either estimated (white) or restricted to zero (grey).

Results

(a) CYP2J19 presence, copy number and variation

Tissue samples from 16 species of weaverbirds (seven *Euplectes*, five *Ploceus*, two *Quelea*, one *Foudia*, one *Philetairus*) were analysed (see Table 3.1). Based on a long range PCR assay on genomic DNA, all 16 species were found to have a single *CYP2J19* gene copy of ~10-15kbp, which was confirmed by Illumina Miseq sequencing in two species (*E. orix* and *Q. quelea*). Given the possibility of differential expression of copies in different tissues (as in the zebra finch), we further confirmed a single copy in ploceids by showing that full length sequences of *CYP2J19* cDNA from different tissues (retina and liver) of the same individual were identical in three species (*E. ardens, F. madagascariensis,* and *Q. quelea*).

Full length *CYP2J19* cDNA sequences revealed that there were no shared amino acid substitutions present among species that possess red C4-ketocarotenoid plumage compared to species without C4-ketocarotenoids.

(b) Patterns of CYP2J19 expression

Analysis of expression using qualitative RT-PCR showed strong *CYP2J19* expression in the retinas of all species examined (N = 10 species), variable expression in the liver, and weak or undetectable expression in all peripheral tissues (skin and feather follicles (N = 5 species), beak (N = 6), and tarsus (N = 2), Table 3.2). Using *Q. quelea* as an example since it is the only sampled species with red bare body parts (beak and tarsus), significant heterogeneity of *CYP2J19* expression was found ($F_{3,8} = 87.96$, $p < 0.001$) with the liver and retina having higher expression than the beak and tarsus (Figure 3.2, Table 3.3).

Table 3.2. Qualitative analysis of CYP2J19 expression in the retina, beak, tarsus, skin and feather follicles of 16 ploceid species. Strong (●), weak (ʘ) and undetectable (-) expression levels are shown. Gaps in table were not determined.

SPECIES	TISSUES			
	Retina	Beak	Tarsus	Skin and feather follicle
Ploceus melanocephalus	●			
Ploceus capensis	●	ʘ		
Ploceus velatus	●	ʘ		
Foudia madagascariensis	●	-		
Quelea erythrops				ʘ
Quelea quelea	●	ʘ	ʘ	
Euplectes afer	●			ʘ
Euplectes aureus				ʘ
Euplectes axillaris	●			
Euplectes macrourus	●	-		
Euplectes ardens	●			ʘ
Euplectes hordeaceus				ʘ
Euplectes nigroventris				ʘ
Euplectes orix	●	ʘ	ʘ	

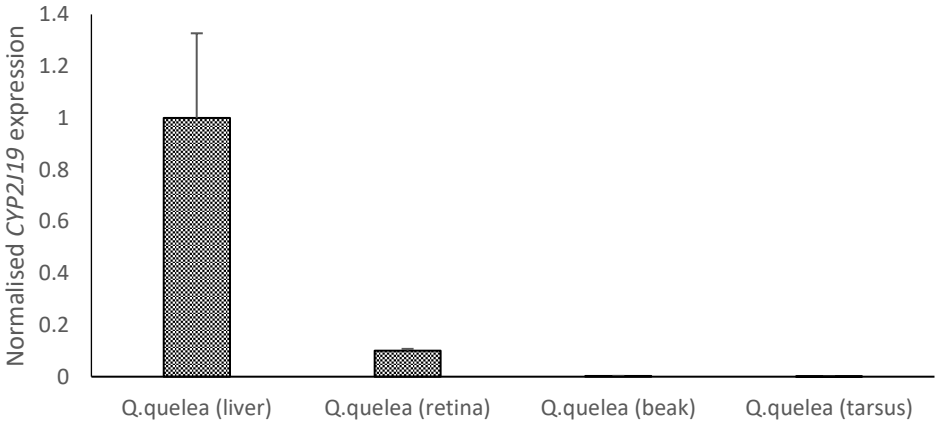

Figure 3.2. *CYP2J19* expression quantified by qRT-PCR in *Q. quelea*, normalised against *β-actin* (N = 3). Error bars represent SEM.

Table 3.3. Tukey's pairwise tests of *CYP2J19* expression for four Q. *quelea* tissues (N = 3) showing adjusted p-values.

Pairwise tissue comparisons	p adj
Liver – Beak	0.0000088 ***
Retina – Beak	0.0005862 ***
Tarsus – Beak	0.1668680
Retina – Liver	0.0030217 **
Tarsus – Liver	0.0000023 ***
Tarsus – Retina	0.0000725 ***

Initial qRT-PCR quantification of hepatic expression of *CYP2J19* using a single control locus (*β-actin*) and samples of 1-3 breeding males across the 16 species showed high levels of *CYP2J19* in four members of the *Euplectes* clade (*E. orix*, *E. hordeaceus*, *E. nigroventris*, *E. ardens*), two queleas and a fody, with levels more than 100-fold greater than all other species (Figure 3.3). We confirmed these findings by performing qRT-PCRs using three control loci on a randomly chosen subset of samples (one individual per species) (Figure 3.4). These gave similar results, with the same seven species showing high (0.1 – 8.6) levels of hepatic *CYP2J19* compared to the remaining species (<0.007) (more than 14 fold difference).

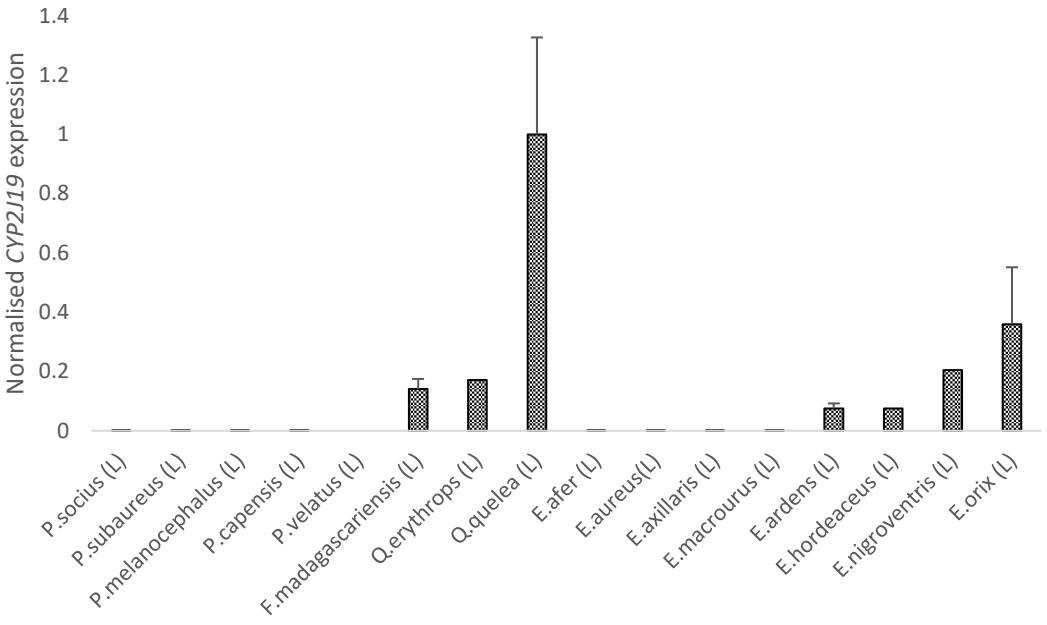

Figure 3.3. Normalised liver expression of *CYP2J19* in 16 ploceids in relation to *β-Actin.* SEM of five species (*F. madagascariensis, Q. quelea, E. macrourus, E. ardens* and *E. orix*) are presented.

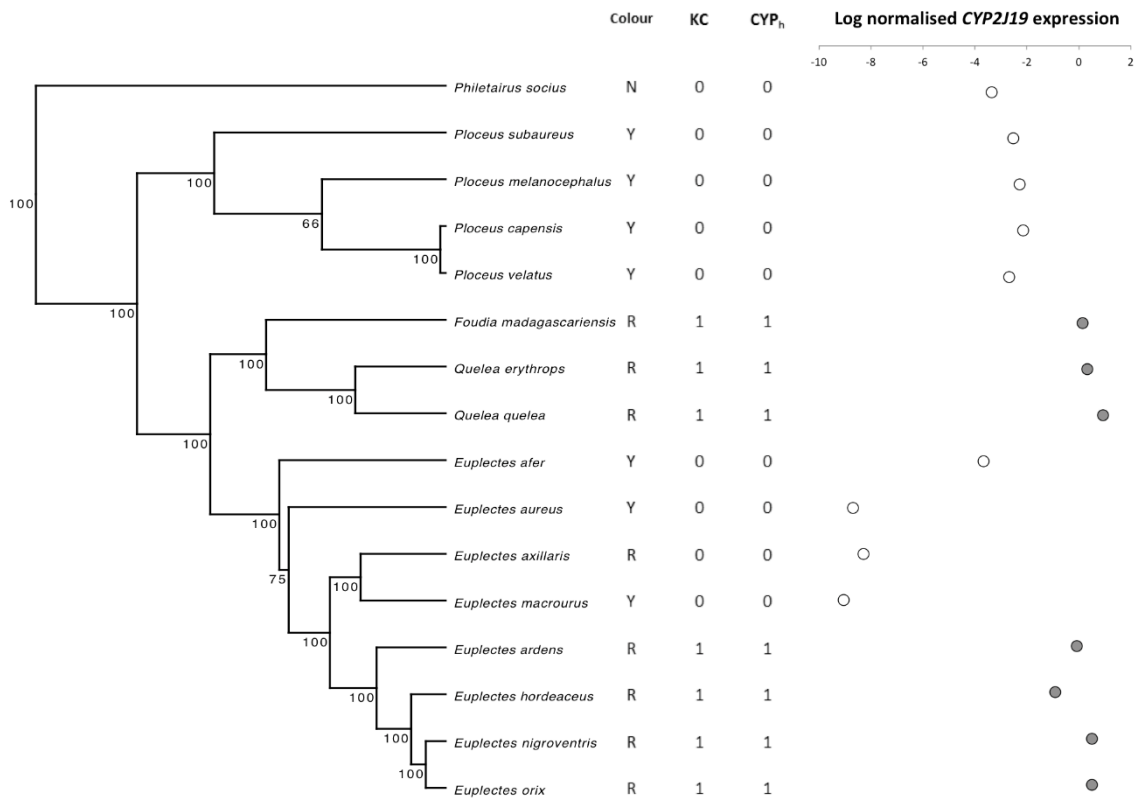

Figure 3.4: Hepatic expression of *CYP2J19* in 16 weaverbirds, in relation to phylogeny, coloration and ketocarotenoid presence. Gene expression was normalised against *β-Actin, GAPDH* and *HPRT1*. Expression of *CYP2J19* in three species (*E. aureus, E. axillaris* and *E. macrourus*) were undetectable after 50 PCR cycles. The phylogeny is a 50% majority-rule consensus (MRC) tree constructed in Mesquite 3.03 based on 10,000 trees downloaded from birdtree.org, numbers showing clade credibility (Bayesian posterior probability) in percent. Discrete scores of hepatic *CYP2J19* expression level (CYP_h: 0 = 'low', 1 = 'high') and red ketocarotenoid pigmentation (KC: 0 = 'absent', 1 = 'present') used in evolutionary association tests are shown. The carotenoid-based coloration of the species (Red: 'R', Yellow: 'Y', Carotenoid absent: 'N'), is also shown.

(c) Association between CYP2J19 and red ketocarotenoid pigmentation

There is a perfect association between high hepatic *CYP2J19* expression and the presence of red C4-ketocarotenoids: breeding males of the seven species with high liver expression of *CYP2J19* all have red C4-ketocarotenoid based coloration (Figure 3.4). In contrast, the nine species without C4-ketocarotenoids (eight of which have yellow carotenoids, one with no carotenoids) all have low hepatic *CYP2J19* expression, and this includes *E. axillaris*, the only species sampled here that produces a red colour hue using 'yellow' carotenoids only, i.e. without C4-ketocarotenoids (Andersson *et al.* 2007).

Phylogenetic comparative tests of correlated evolution between hepatic *CYP2J19* expression and red C4-ketocarotenoid pigmentation were performed in BayesTraits V2. Estimated marginal likelihoods based on five different prior assumptions of transition rates (Table 3.3) consistently support a 'dependence model', where the evolution of red C4-ketocarotenoid pigmentation is contingent on high hepatic expression of *CYP2J19*, over an 'independence model', where the rate of change in one trait is unaffected by the state of the other trait (Figure 3.1). Even with priors that favoured the 'independence model', Bayes Factor test statistics (calculated as 2*[lnL(dependent model) - lnL(independent model)] exceeded 10, which usually is interpreted as very strong support for an association (Kass and Raftery 1995).

Table 3.3. Estimated marginal log-likelihoods (lnL) of 'dependency' (dep) versus 'independency' (indep) models, and the Bayes Factor test statistic (2lnBF), given different prior assumptions of evolutionary transition rates.

Rate distribution prior	lnL(dep)	lnL(indep)	2lnBF
Uniform (0, 100)	-11.6	-21.8	20.5
Uniform (0, 1)	-12.0	-19.4	14.8
Exponential (1/λ=0.1)	-11.2	-16.6	10.8
Exponential (1/λ=1)	-11.2	-20.2	18.1
Exponential (1/λ=10)	-10.9	-22.0	22.1

Discussion

Molecular analysis established that weaverbirds have a single copy of *CYP2J19,* and that high liver expression of *CYP2J19* in breeding males is associated with the presence of red C4-ketocarotenoid coloration. The single copy of *CYP2J19* is interesting since the zebra finch, an estrildid finch belonging to the nearest outgroup clade to Ploceidae (De Silva *et al.* 2017), has two copies, *CYP2J19A* and *CYP2J19B*, which appear to be specialised for a retinal function and red coloration, respectively (Mundy *et al.* 2016). It therefore appears that the estrildid *CYP2J19* duplication occurred after the split between ploceids and estrildids. More broadly, a single copy of *CYP2J19* was reported also in the red factor canary (Lopes *et al.* 2016), as well as in chicken and ostrich (Twyman *et al.* 2016) and GenBank searches reveal only a single copy in the vast majority of available avian genomes (Chapter 5), which means that a single *CYP2J19* copy probably is the "normal" situation for birds.

Tissue-specific expression data for *CYP2J19* strongly implicate the liver as the main site of carotenoid ketolation in ploceids, making ploceids "central" ketoconverters. There was a perfect association between high *CYP2J19* expression and the presence of red C4-ketocarotenoid pigmentation, and very strong support for dependence of ketocarotenoid presence on *CYP2J19* expression in the Bayesian phylogenetic comparative analysis. Notably, the hepatic *CYP2J19* expression was very low in *E. axillaris*, a species with red carotenoid coloration that does not involve C4-ketocarotenoids, supporting a direct link (i.e. not via colour) between *CYP2J19* and C4-ketocarotenoids. In contrast to variable liver *CYP2J19* expression, we detected low or no *CYP2J19* expression in peripheral tissues (skin, feathers, beak, tarsus), including the red beak and tarsus of the red-billed quelea, which have a high concentration of ketocarotenoids. Nevertheless, more extensive sampling will be required to rule out the possibility of a minor role of *CYP2J19* in feather follicles, as has been described in the mutant red factory canary (Lopes *et al.* 2016). The major role for liver *CYP2J19* expression in performing ketolation for red coloration is concordant with the previously described high plasma concentration of red ketocarotenoids in ploceids (Prager *et al.* 2009), required to transport ketocarotenoids to their peripheral site of deposition.

There was substantial variation in hepatic *CYP2J19* expression even among species with high expression. Interestingly, the highest expression was found in the red-billed quelea, the only species with a red beak and tarsus. We speculate that this may relate to a need for a more or less constant supply of pigments to this continually renewing tissue, which may also require a greater amount of pigment to achieve its colour, compared to plumage. The lower but still highly variable *CYP2J19* expression among species with red plumage likely relates to timing of sampling in relation to the pre-nuptial plumage moult, but also to a host of other genetic, social and environmental factors, such as quantity and quality of carotenoids in the recent diet.

Historically there has been considerable debate over the anatomical site of ketolation (McGraw 2004; del Val *et al.* 2009). The contrast between the red-billed quelea and zebra finch, which both have red beak and tarsus, is particularly striking: the former has high *CYP2J19* expression in liver and low/absent expression in beak and tarsus, while the zebra finch shows the opposite pattern. Overall, ploceids have a single copy of *CYP2J19* and perform central ketoconversion to colour plumage and bare body parts, while the estrildid zebra finch has two gene copies and uses peripheral ketoconversion to colour bare body parts (beak and tarsus), and the fringillid red factor canary has a single copy and uses both central (liver) and peripheral (feather follicle) ketoconversion to colour plumage (although it is unknown whether this is a natural occurrence in fringillids). The adaptive value of variation in *CYP2J19* copy number and ketoconversion site is an interesting question, and a much broader sampling across avian clades will be useful to address this issue.

Based on a previous ancestral character state reconstruction (Prager and Andersson 2010), the two clades with high hepatic *CYP2J19* expression (*Foudia/Quelea* and *E. ardens/E. hordeaceus/E. nigroventris/E. orix*) likely acquired red coloration convergently. We therefore hypothesize that convergent evolution of red coloration occurred in these two clades via increases in hepatic *CYP2J19* expression. Due to the highly conserved function of *CYP2J19* for red retinal oil droplets, this would have required specific acquisition of high liver expression of *CYP2J19* while maintaining high retinal expression, which may have occurred via evolution of *cis*-regulation of *CYP2J19* and/or *trans*-regulating factors. Given the relatively rare but phylogenetically widespread occurrence of red carotenoid coloration, the co-option of *CYP2J19* has probably occurred many times in birds and also in turtles (Chapter 3, Twyman *et al.* 2016). Loss of red C4-ketocarotenoid

coloration could have also occurred in some lineages, perhaps via regulatory changes at *CYP2J19,* or changes in other carotenoid utilisation mechanisms such as uptake, transport, metabolism and deposition.

Cis-regulatory evolution is often regarded as a major motor behind adaptive change (Stern and Orgogozo 2008). Factors contributing to this include the high evolvability of cis-regulatory regions, which have been suggested to be less constrained and can avoid potential negative pleiotropic effects associated with mutations in the gene coding regions (Stern and Orgogozo 2008). The evolution of the role of *CYP2J19* in red coloration fits well with the *cis*-regulatory mechanism, since it is a change in gene expression that seems to be the key to enable red coloration. Given this, one would expect a more widespread occurrence of red ketocarotenoid based coloration, which is not what is seen. The relatively rare and patchy distribution of red coloration in birds may indicate that additional constraints are present in the control of *CYP2J19* expression. These may include coordination of expression in relation to factors such as age, sex, body condition and season, which potentially requires the evolution of additional cis-regulatory modules that integrates information from the various cellular networks. Detailed investigation of the genetic and environmental causes and consequences of the co-option of *CYP2J19* for integumentary pigmentation will be crucial to elucidate these issues.

Rapid progress has recently been made in documenting the genetic basis of convergent evolution of naturally selected traits (Stern 2013). For example, in birds, the evolution of melanin-based coloration is frequently due to two loci, *MC1R* and *ASIP* (Mundy 2005; Uy et al. 2016). Here we have uncovered one of the first examples in vertebrates where a single locus is involved in convergent evolution of a sexually-selected trait. Future work on *CYP2J19* promises many novel insights into the diversification and signalling function of red colour patches in birds.

Chapter 5

Evolution of *CYP2J19*, a gene for red coloration and colour vision in birds: Positive selection in the face of conservation and pleiotropy

Abstract

The gene *CYP2J19* has been described to be associated with ketocarotenoid pigmentation in two lineages of passerine birds. However, the extent to which *CYP2J19* functions in carotenoid ketolation across the entire avian clade is unknown. From 70 avian genomes, I retrieved full length *CYP2J19* of 43 species, representing all major avian clades, which is consistent with a conserved function of *CYP2J19* in colour vision, and its co-option for red coloration. In addition, I analysed *CYP2J19* sequences obtained from 13 species of weaverbirds (Ploceidae), which have variable red C4-ketocarotenoid coloration. I find that the locus has been positively selected throughout the radiation of the birds despite the conserved retinal function and pleiotropy via co-option for red coloration. Analysis on a further eight *CYP* loci across 25 matched species show that positive selection is common within members of this gene family. In addition, I did not detect a change in selection pressure on *CYP2J19* following co-option for red coloration in the weaverbirds. The cause of positive selection on *CYP2J19* is unclear, but it may be partly related to selection at the adjacent locus *CYP2J40*.

Introduction

The genetic basis of adaptations for signalling is of fundamental importance in biology. Recently a fascinating and potentially widely important genetic link was found in birds between the generation and sensory discrimination of long wavelength (yellow to red) carotenoid colour hues, which are frequent targets of social and sexual selection (Burley and Coopersmith 1987; Hill 1990; Pryke *et al.* 2002; Hill and McGraw 2006). Red C4-ketocarotenoid pigmentation was shown (in zebra finch and the hybrid red factor canary) to depend on the locus *CYP2J19,* which likely encodes a ketolase that catalyses the conversion of dietary yellow carotenoids to their red C4-ketocarotenoid derivatives (Lopes *et al.* 2016; Mundy *et al.* 2016). In addition to its function in coloration, *CYP2J19* is also expressed in the avian retina (Goldsmith *et al.* 1984; Lopes *et al.* 2016) where it functions to generate the C4-ketocarotenoid astaxanthin in the red oil droplets in long wave sensitive cones (Goldsmith *et al.* 1984).

A study on the deep evolutionary history of *CYP2J19* in reptiles showed that it arose in the common ancestor of turtles and archosaurs, and that cone oil droplet pigmentation in the retina most likely was the original function, from which it subsequently was co-opted for red coloration independently in avian and turtle lineages (Twyman *et al.* 2016) (Chapter 3). Whereas most of the few birds examined to date seem to have a single *CYP2J19* gene, two copies have been reported in the zebra finch, one specialised for colour vision (*CYP2J19A*) and the other for red coloration (*CYP2J19B*) (Mundy *et al.* 2016). Given the importance of duplication events for functional divergence, it is an interesting question if and where *CYP1J19* duplication has occurred in other bird lineages.

Carotenoid-based coloration (yellow, orange, red) has evolved multiple times in birds in a complex pattern (Thomas *et al.* 2014) and within a few avian clades, the evolution of red carotenoid-based coloration has been well studied, notably in the genus of African widowbirds and bishops (Ploceidae; *Euplectes*) together with its sister group including the genera *Quelea* and *Foudia* (Prager and Andersson 2010). In this clade, red C4-ketocarotenoid coloration evolved twice from a yellow ancestor, and has been shown to be strongly associated with high hepatic (liver) expression of *CYP2J19,* whereas both yellow and red species express *CYP2J19* in the retina (Twyman *et al.* in review, Chapter 4). Moreover, in contrast with the zebra finch,

CYP2J19 occurs in a single copy in all Ploceidae studied so far (Twyman *et al.* in review, Chapter 4). Hence this is an excellent group in which to address whether the evolution of pleiotropy in *CYP2J19* through the acquisition of *CYP2J19*-based red coloration is associated with a change in the pattern of selection on *CYP2J19*.

The detection of natural selection in the genome, through the ratio of non-synonymous to synonymous mutations (dN/dS) is a powerful tool identifying loci involved in adaptation (Yang and Bielawski 2000), as illustrated by many compelling cases of the genetics of adaptation (Hughes and Nei 1988; Messler and Stewart 1997; Nadeau *et al.* 2007). Positive selection at the molecular level is often associated with phenotypic adaptation, and two factors that are considered to have a strong effect on whether a gene is likely under positive selection towards an adaptive peak are 1) whether the gene has a well conserved function and 2) whether the gene is pleiotropic (i.e. influences more than one phenotypic traits). In both cases, gene evolution is likely constrained and purifying selection is considered the more likely outcome. However, under situations whereby an adaptive peak has already been reached, positive selection can still operate in order to alleviate deleterious alleles arising through selection for adaptive changes in another interacting locus, such as in the case of compensatory mutations (non-adaptive mutations under positive selection). This can potentially explain the occurrence of positive selection in a gene with a conserved function. Of relevance here is the presence of another *CYP, CYP2J40,* adjacent to *CYP2J19* on chromosome 8. Selection on *CYP2J40* could potentially lead to fixation of deleterious alleles in *CYP2J19* and result in positive selection within the locus itself for compensation. Currently, there is a poor understanding of the prevalence of compensatory mutations.

In this study, I investigated the evolution of *CYP2J19* in birds, by first mining available avian genomic data for presence, copy number and selection of *CYP2J19,* followed by a focused selection analysis of the new *CYP2J19* sequence data from the striking and differentially carotenoid signalling weaverbirds (Ploceidae), where the dual (pleiotropic) effects of *CYP2J19* on both colour production and perception are uniquely well documented. More specifically, I asked (i) whether *CYP2J19* is present in all birds, (ii) what the copy number of *CYP2J19* is in different avian lineages, (iii) what the pattern of selection on *CYP2J19* is across birds and how this compares to other *CYP*s, including *CY2J40,* which is syntenic to *CYP2J19* on chromosome 8, and (iv) whether selection on *CYP2J19* changes after it was co-opted for a pleiotropic effect on red coloration.

Materials and methods

(a) Sequence acquisition

BLASTn searches were performed on 70 Genbank avian genomes using query sequences for zebra finch and chicken *CYP2J19* annotated in Ensembl 83 (Cunningham *et al.* 2015). The full-length coding sequence as well as all 9 exons of *CYP2J19* were blast searched individually against the 70 genomes, and the resultant sequences were examined manually for complete open reading frames. Full length (1431 bp) sequences from 43 species were retained for downstream analyses.

An assembled dataset of other *CYP*s was used in a matched comparative study of *CYP* evolution. The 43 avian genomes with full-length *CYP2J19* were searched for 35 CYPs identified from the chicken genome. In order to maximise the number of *CYP*s and lineages, eight further full length orthologous *CYP*s were obtained from 25 avian genomes, representing all major avian clades (Table 5.1). The full list of nine CYPs and their chromosomal locations in zebra finch are as follows: *CYP2J19* (Chromosome 8, 1431 bp), *CYP2J40* (Chromosome 8, 1482 bp), *CYP19A1* (Chromosome 10, 1506 bp), *CYP7A1* (Chromosome 2, 1536 bp), *CYP8B1* (Chromosome 2, 1524 bp), *CYP4V2* (Chromosome 4, 1548 bp), *CYP3A9* (Chromosome 14, 1449 bp), *CYP7B1* (Chromosome 2, 1404 bp), *CYP20A1* (Chromosome 7, 1317 bp).

Furthermore, full-length *CYP2J19* data from thirteen species of weaverbirds (Aves: Ploceidae), sampled by Staffan Andersson, were used to investigate selection in relation to coloration in this clade (species with C4-ketocarotenoid coloration marked with an asterisk): *Ploceus melanocephalus*, *P. capensis*, *P. velatus*, **Foudia madagascariensis*, **Quelea quelea*, **Q. erythrops*, *Euplectes afer*, **E. orix*, **E. hordeaceus*, **E. nigroventris*, **E. ardens*, *E. axillaris, E. macroura* (Twyman et al. in review). C4-ketocarotenoid coloration evolved independently in two clades of red birds: the *Foudia*/*Quelea* clade and the *E. orix/hordeaceus/nigroventris/ardens* clade (Prager & Andersson 2010).

All gene sequences were aligned in MEGA 6 (Tamura *et al.* 2013) using MUSCLE (Edgar 2004), and all species trees were taken from BirdTree.org (Jetz *et al.* 2012).

Table 5.1. 70 avian species investigated in the study and those used in downstream analysis.

Order	70 avian genomes searched	43 species with full-length *CYP2J19*	25 species used in comparative CYP analysis
STRUTHIONIFORMES	*Struthio camelus* (African ostrich)	✓	
APTERYGIFORMES	*Apteryx australis* (brown kiwi)		
TINAMIFORMES	*Tinamus guttatus* (white-throated tinamou)		
GALLIFORMES	*Colinus virginianus* (northern bobwhite)		
	Coturnix japonica (Japanese quail)		
	Gallus gallus (red junglefowl)	✓	✓
	Lyrurus tetrix (black grouse)		
	Meleagris gallopavo (turkey)		
ANSERIFORMES	*Anas platyrhynchos* (mallard)	✓	✓
	Anser cygnoides (swan goose)	✓	✓
PHOENICOPTERIFORMES	*Phoenicopterus ruber* (American flamingo)		
PODICIPEDIFORMES	*Podiceps cristatus* (great crested grebe)		
COLUMBIFORMES	*Columba livia* (rock pigeon)	✓	✓
MESITORNITHIFORMES	*Mesitornis unicolor* (brown roatelo)	✓	
PTEROCLIDIFORMES	*Pterocles gutturalis* (yellow-throated sandgrouse)		
APODIFORMES	*Calypte anna* (Anna's hummingbird)	✓	
	Chaetura pelagica (chimney swift)	✓	✓
CAPRIMULGIFORMES	*Caprimulgus carolinensis* (chuck-will's-widow)	✓	
CUCULIFORMES	*Cuculus canorus* (common cuckoo)	✓	✓
OTIDIFORMES	*Chlamydotis macqueenii* (MacQueen's bustard)		
	Chlamydotis undulata (houbara bustard)		
MUSOPHAGIFORMES	*Tauraco erythrolophus* (red-crested turaco)	✓	
OPISTHOCOMIFORMES	*Opisthocomus hoazin* (hoatzin)	✓	✓
GRUIFORMES	*Balearica pavonina* (black crowned crane)		
	Balearica regulorum (grey crowned crane)	✓	✓
CHARADRIIFORMES	*Calidris pugnax* (ruff)		
	Charadrius vociferous (killdeer)	✓	
GAVIIFORMES	*Gavia stellate* (red-throated loon)		
PROCELLARIIFORMES	*Fulmarus glacialis* (northern fulmar)	✓	✓
SPHENISCIFORMES	*Aptenodytes forsteri* (emperor penguin)	✓	✓
	Pygoscelis adeliae (Adelie penguin)		
SULIFORMES	*Phalacrocorax carbo* (great cormorant)		
PELECANIFORMES	*Egretta garzetta* (little egret)	✓	
	Nipponia nippon (crested ibis)	✓	✓
	Pelecanus crispus (dalmatian pelican)		
EURYPYGIFORMES	*Eurypyga helias* (sunbittern)		
PHAETHONTIFORMES	*Phaethon lepturus* (white-tailed tropicbird)	✓	✓
CATHARTIFORMES	*Cathartes aura* (turkey vulture)		
ACCIPITRIFORMES	*Haliaeetus albicilla* (white-tailed eagle)	✓	✓
	Haliaeetus leucocephalus (bald eagle)		
	Aquila chrysaetos (golden eagle)	✓	✓
STRIGIFORMES	*Tyto alba* (barn owl)		
COLIIFORMES	*Colius striatus* (speckled mousebird)	✓	✓
LEPTOSOMIFORMES	*Leptosomus discolour* (cuckoo roller)	✓	
TROGONIFORMES	*Apaloderma vittatum* (bar-tailed trogon)	✓	
BUCEROTIFORMES	*Buceros rhinoceros* (rhinoceros hornbill)	✓	

CORACIIFORMES	*Merops nubicus* (northern carmine bee-eater)		
PICIFORMES	*Picoides pubescens* (downy woodpecker)	✓	✓
CARIAMIFORMES	*Cariama cristata* (red-legged seriema)	✓	
FALCONIFORMES	*Falco cherrug* (saker falcon)	✓	✓
	Falco peregrinus (peregrine falcon)	✓	✓
PSITTACIFORMES	*Amazona aestiva* (blue-fronted amazon)		
	Amazona vittata (Puerto Rican parrot)		
	Ara macao (scarlet macaw)		
	Melopsittacus undulatus (budgerigar)	✓	✓
	Nestor notabilis (Kea)	✓	
PASSERIFORMES	*Acanthisitta chloris* (rifleman)	✓	
	Corvus brachyrhynchos (American crow)		
	Corvus cornix (hooded Crow)	✓	✓
	Ficedula albicollis (collared flycatcher)	✓	✓
	Geospiza fortis (medium ground-finch)	✓	✓
	Manacus vitellinus (golden-collared manakin)	✓	
	Parus major (great tit)	✓	
	Phylloscopus plumbeitarsus (two-barred warbler)		
	Pseudopodoces humilis (Tibetan ground-tit)	✓	✓
	Serinus canaria (common canary)	✓	✓
	Sturnus vulgaris (common starling)	✓	
	Taeniopygia guttata (zebra finch)	✓	✓
	Zonotrichia albicollis (white-throated sparrow)	✓	
	Zosterops lateralis (silver-eye)	✓	

(b) Molecular evolution analyses

In order to investigate the presence and the type of selective forces acting on *CYP* genes, the ratio of nonsynonymous to synonymous substitution rates (dN/dS = ω) was estimated using the CodeML program within PAML 4.7 (Phylogenetic Analysis by Maximum Likelihood) (Yang 2007). Omega values of less than one, equal to one, and more than one, correspond to negative, neutral and positive selection respectively. Several model comparisons were performed, including site, branch and branch-site comparisons for the detection of positive selection. Bonferroni correction for multiple testing was carried out in the comparative analysis of CYPs, with n = 9 representing the number of loci tested. All models were run several times (where applicable) applying different initial ω values in order to avoid local likelihood peaks. The Likelihood Ratio Test (LRT) was used to compare between nested models (2 x [lnL(alternative model) - lnL(null model)]), with the number of degrees of freedom (df) calculated as the difference in the number of model parameters between the comparisons. Bayes empirical Bayes method (BEB) was used to identify positively selected sites when significant results were found.

Site models allow the ω to vary among different codon sites (Yang 1997), and comparisons between M1a (Nearly Neutral model) and M2a (Positive selection) and also between M7 (beta) and M8 (beta& ω) models were implemented here. Model M1a allows for sites to fall into two categories: ω < 1 and ω = 1, whereas M2a includes an additional category of ω > 1. The M8 model has 11 site classes and includes an additional category of ω > 1 absent from M7. The critical values for the LRT comparisons between M1a-M2a and M7-M8 were 5.99 at 5% and 9.21 at 1% under 2 degrees of freedom.

To investigate the selective forces acting along certain lineages, branch-specific models were used to assess whether the ω values varied significantly between preselected foreground and background lineages (Yang 1998). In these tests, the null model (M0) assumes a single ω value across all branches. Analysis was carried out on weaverbird *CYP2J19* sequences, and selection of foreground lineages was based on the presence of red ketocarotenoid coloration (Andersson *et al.* 2007; Prager *et al.* 2009; Prager and Andersson 2010).

In order to identify positive selection amongst different codon sites and between lineages, branch-site models were implemented on the ploceid dataset. The

modified alternative model A (Zhang *et al.* 2005) was compared with the null model which fixes ω = 1.

Results

BLASTn searches identified *CYP2J19*-like sequences within all of the 70 avian genomes with the exception of the bald eagle (*Haliaeetus leucocephalus*) and the two-barred warbler (*Phylloscopus plumbeitarsus*). Five of the remaining 68 genomes failed to find all 9 exons of *CYP2J19* (*Amazona aestival, Amazona vittata, Apteryx australis, Ara macao,* and *Lyrurus tetrix*), and a further 20 species were discarded due to the presence of multiple internal stop codons and indels. Partial gene duplications were found for three species: downy woodpecker (*P. pubescens*), collared flycatcher (*F. albicollis*) and scarlet macaw (*A. macao*), with 2, 4 and 5 exons duplicated, respectively. Full-length open reading frames of *CYP2J19* were obtained for the remaining 43 species (Table 5.1).

Selection analysis of full length coding sequence of *CYP2J19* in 43 species revealed strong evidence for positive selection (> 6% positively selected sites with ω > 1.3; LRT > 61.57, p = 0.00) under M7-M8 comparisons (Table 5.3). Five positively selected sites were identified by BEB, and alignment of *CYP2J19* with other avian CYP2 protein sequences described in Almeida *et al.* (2016) showed that three of the sites (37, 122, 474) were located within the predicted functional domain: Substrate Recognition Site 0 (SRS0, site 37), Heme binding domain (HEM, site 122) and Substrate Recognition Site 6 (SRS6, site 474). No evidence for positive selection was found under M1a-M2a tests (Table 5.2).

Table 5.2. Site-specific analysis of *CYP2J19* for 43 avian species.

Underlined sites fall within predicted functional domains for CYP2 proteins annotated in Almeida et al.

Model	Parameter estimates	*lnL*	Model comparison	LRT Statistic	df	P-value	Positively selected sites under BEB P>95% (bold: P>99%)
M0	ω_0 = 0.147	-12990.709					
M1a	ω_0 = 0.056 p_0 = 0.868 ω_1 = 1.000 p_1 = 0.132	-12561.580					
M2a	ω_0 = 0.056 p_0 = 0.868 ω_1 = 1.000 p_1 = 0.077 ω_2 = 1.000 p_2 = 0.055	-12561.580	M1a vs. M2a	0.000	2	1.000	
M7	p = 0.168 q = 0.788	-12533.632					
M8	p_0 = 0.935 (p_1 = 0.065) p = 0.302 q = 2.898 ω_s = 1.383	-12502.845	M7 vs. M8	61.574	2	0.000*	37, 122, 329, **455, 474**

2016. * Significant result.

Given the strong evidence for positive selection on *CYP2J19*, comparisons were made with other *CYP* loci. Using genomic searches, full sequences were obtained for a further eight *CYP* genes across 25 species of birds which are a subset of the 43 species with full *CYP2J19* ORFs (Table 5.1). Selection analysis revealed 6 of the 9 *CYPs* (*CYP2J19*, *CYP2J40, CYP8B1, CYP4V2, CYP3A9* and *CYP7B1*) were under significant positive site-specific selection when comparing between M7 – M8 models, with 0.02 – 0.08 sites with ω of 1.39 – 2.60. Of these, four *CYP*s (*CYP2J40, CYP8B1, CYP4V2* and *CYP3A9*) also showed significant site-specific selection in M1a-M2a model comparisons (Table 5.3).

Amino acid alignment of *CYP2J19* with annotated structures of CYP2 proteins found that certain positively selected sites, identified by BEB, fall within functional domains: SRS0 (site 37)(Almeida *et al.* 2016) and SRS-3 (site 233)(Gotoh 1992), while positively selected sites in *CYP2J40* lie in SRS0 (sites 56, 59, 60, 61 and 80) and SRS-3 (sites 248, 250).

Table 5.3. Matched site-specific comparisons for 25 avian species.

	LRT Statistic		p-value		Corrected p-value		M2a			M8		
CYP gene	**M1a- M2a**	**M7-M8**	**M1a- M2a**	**M7-M8**	**M1a- M2a**	**M7-M8**	**Proportion of sites >1**	**ω>1**	**Positively selected sites under BEB P>95% (bold: P>99%)**	**Proportion of sites >1**	**ω>1**	**Positively selected sites under BEB P>95% (bold: P>99%)**
CYP2J19	4.577	49.376	0.101	0.000		0.000*	0.018	1.944		0.083	1.387	8, 37, 233, 455
CYP2J40	76.397	84.911	0.000	0.000	0.000*	0.000*	0.038	2.974	56, **59**, 60, **61**, **80**, 247, 250, **393**	0.056	2.246	**56**, **59**, **60**, **61**, **80**, 247, 250, **393**
CYP19A1	0.000	3.579	1.000	0.167			0.020	1.000		0.011	1.031	
CYP7A1	0.000	9.871	1.000	0.007		0.065	0.030	1.000		0.062	1.331	
CYP8B1	11.835	46.986	0.003	0.000	0.024*	0.000*	0.028	2.170		0.101	1.506	104
CYP4V2	57.130	64.122	0.000	0.000	0.000*	0.000*	0.031	3.160	31, **67, 69, 75**, 272	0.043	2.603	31, 57, **67, 69, 75**, 145, **272**
CYP3A9	57.980	86.997	0.000	0.000	0.000*	0.000*	0.059	2.504	**88**, 95, 135, **149**, 203, 279, 466	0.081	2.095	4, 87, **88, 95**, 135, **149**, 203, 223, **279**, 309, 466, 470
CYP7B1	8.154	25.295	0.017	0.000	0.153	0.000*	0.017	2.468		0.023	2.220	**89**, 348, 437, 458
CYP20A1	0.000	5.957	1.000	0.051			0.047	1.000		0.014	1.833	

Underlined sites fall within predicted functional domains for CYP2 proteins annotated in Almeida *et al.* 2016. * Significant result after p-value correction for multiple testing.

There was no evidence of positive selection on *CYP2J19* within the Ploceidae (Table 5.4). Neither was there a significant difference in ω in *CYP2J19* between lineages that contain C4-ketocarotenoid-based red coloration and those that do not in the branch-specific analysis. No signal for positive selection was detected when applying branch-site models (Table 5.4).

Table 5.4. Summary of PAML results for *CYP2J19* across 13 ploceid species including site-specific, branch and branch-site analyses.

Model	N	Parameter estimates	*lnL*	Model comparison	LRT Statistic	df	p-value
M0	13	$\omega_0 = 0.241$	-2591.727				
M1a	13	$\omega_0 = 0.024$ $p_0 = 0.782$ $\omega_1 = 1.000$ $p_1 = 0.218$	-2587.623				
M2a	13	$\omega_0 = 0.176$ $p_0 = 0.985$ $\omega_1 = 1.000$ $p_1 = 0.000$ $\omega_2 = 5.935$ $p_2 = 0.015$	-2586.444	M1a vs. M2a	2.358	2	0.308
M7	13	$p = 0.005$ $q = 0.019$	-2587.657				
M8	13	$p_0 = 0.985$ ($p_1 = 0.015$) $p = 21.271$ $q = 99.000$ $\omega_s = 5.960$	-2586.447	M7 vs. M8	2.422	2	0.298
2ω branch model *	13	ω (#0) = 0.165 ω (#1) = 0.370	-2590.044	M0 vs. 2ω branch model	3.366	1	0.067
Branch-site Null *	13	site class 0 1 2a 2b proportion 0.677 0.144 0.148 0.031 background ω 0.000 1.000 0.000 1.000 foreground ω 0.000 1.000 1.000 1.000	-2586.422				
Branch-site Alternative Model A *	13	site class 0 1 2a 2b proportion 0.756 0.147 0.081 0.016 background ω 0.012 1.000 0.012 1.000 foreground ω 0.012 1.000 1.939 1.939	-2586.295	Null vs. Alternative Model A	0.255	1	0.613

*Foreground branches (#1) include all lineages with integumentary ketocarotenoid (*F. madagascariensis, Q. erythrops, Q. quelea, E. ardens, E. nigroventris, E. orix* and *E. hordeaceus*), and background branches (#0) refers to all lineages lacking red ketocarotenoids.

Discussion

A single copy of *CYP2J19* was identified in a phylogenetically broad range of avian lineages, supporting a single *CYP2J19* locus as the ancestral avian state. Considering that red C4-ketocarotenoid coloration has a patchy distribution in birds (including the lineages sampled here) whereas red retinal oil droplets are present in almost all birds examined, the contribution to colour vision is almost certainly the ancestral and likely strongly conserved function of *CYP2J19.* The finding of positive selection on *CYP2J19* across all lineages analysed is somewhat surprising. There was no evidence that the acquisition of a second function in red coloration affected the selective pressure on *CYP2J19*. However, we did find evidence for positive selection on *CYP2J40*, a syntenic gene of unknown function, which may plausibly lead to the evolution of compensatory mutations in *CYP2J19.*

In most avian species examined, there was evidence for the presence of *CYP2J19*. Although full-length open reading frames were only retrieved for 43/70 (61%) of species, these 43 species span most major avian lineages and it seems likely that poor genome quality can account for much of the missing data. Of the 43 species with intact ORFs, only the zebra finch (*Taeniopygia guttata*) has been confirmed to possess red C4-ketocarotenoid based coloration (Burley and Coopersmith 1987; McGraw and Toomey 2010). By contrast, it is thought that almost all birds have red oil droplets in their retinas, including some nocturnal lineages such as the tawny owl (*Strix aluco*) (Bowmaker and Martin 1978). Hence, notwithstanding the possibility of rare loss of functional *CYP2J19* in certain lineages (Emerling 2017), our results support the notion that *CYP2J19* has an ancestral, conserved function in avian colour vision and its function in red coloration appears to have been independently gained along specific bird lineages. This would most likely have been through changes in the patterns of expression of *CYP2J19* as a result of in *cis*-regulatory sequences and/or *trans*-acting factors. This concords with our previous findings investigating the origin of *CYP2J19* in the reptiles (Twyman *et al.* 2016).

Analyses of the 43 species dataset as well as the reduced 25 species dataset consistently revealed that positive selection acted on *CYP2J19* throughout its evolution in birds. In particular, site-specific selection was found to act within the newly defined substrate recognition site (SRS0) of *CYP2J19* for both datasets (Almeida *et al.* 2016). This was surprising given the inferred conserved function of

CYP2J19 in the retina, where it is involved in generating the ketocarotenoid astaxanthin that is the major component of all red cone oil droplets examined (Goldsmith *et al.* 1984). The gain of pleiotropy via co-option of *CYP2J19* function in red coloration might provide a further constraint on *CYP2J19* evolution. However, we found no evidence for a change in mode of selection on *CYP2J19* in ploceids, where we were able to directly compare lineages with exaggerated red C4-ketocarotenoid coloration and those without.

Extending the selection analyses to eight other *CYP* loci showed that selection is common in the gene family. Five of the loci showed evidence for positive selection and these included genes that have been shown in mammals to be involved in bile acid synthesis (oxysterol 7α-hydroxylase (*CYP7B1*) and sterol 12-alpha-hydroxylase (*CYP8B1*)) (Eggertsen *et al.* 1996; Rose *et al.* 1997; Schwarz *et al.* 1997; Gafvels *et al.* 1999), fatty-acid oxidation (fatty acid ω-hydroxylase, *CYP4V2*) (Nakano *et al.* 2009) and the control of sexual dimorphism (*CYP3A9*) (Mahnke *et al.* 1997; Wang and Strobel 1997; Robertson *et al.* 1998; Anakk *et al.* 2003). The remaining locus under positive selection is *CYP2J40,* of unknown function. This locus is of particular interest since it lies adjacent to *CYP2J19* on chromosome 8. We find that it has one of the strongest signatures of positive selection amongst the loci studied, and concordant with previous results (Almeida *et al.* 2016), several positively selected sites fall in the functional domain SRS0. The three *CYP* loci found not to be under positive selection are the aromatase, *CYP19A1*, which plays a crucial role in sex determination and appears to be well conserved in nearly all species studied: birds (Elbrecht and Smith 1992; Peterson *et al.* 2005), turtles (Crews and Bergeron 1994), zebrafish (Lau *et al.* 2016) and vertebrates in general (Naftolin *et al.* 1996), *CYP7A1,* which catalyses the first and rate-liming step in bile synthesis (Ogishima *et al.* 1987; Jelinek *et al.* 1990), and *CYP20A1,* of unknown function.

Overall, therefore, although certain loci with well conserved functions (notably *CYP19A1*) do not show evidence of positive selection, there are many *CYP* loci which do show evidence of positive selection across a broad phylogenetic sampling of birds. Moreover, our sample of 9 *CYP* loci are likely among the most conserved in the avian genome since we targeted *CYP* genes that were identifiable in diverse genomes. A study focusing on the *CYP2* gene family, which are believed to primarily be involved in metabolising toxins, also found evidence for widespread positive selection (6 out of 12 *CYP2* subfamilies, and 11 out of 17 loci; Almeida *et al.* 2016). Our results are thus not unusual for *CYP* loci in general, but while selection of many

of them may be attributed to coevolution with e.g. new toxins, the question remains why some *CYP* loci with more conserved functions, such as *CYP2J19,* are also evolving under positive selection. For *CYP2J19*, an intriguing possibility is that it is affected by selection acting on the adjacent *CYP2J40*, which in the zebra finch is a mere few kbp upstream of *CYP2J19A* on chromosome 8. Specifically, strong selection on *CYP2J40* may lead to the fixation of mildly deleterious alleles on *CYP2J19* and result in positive selection on *CYP2J19* itself for compensatory mutations, possibly in protein functional domains, to alleviate the deleterious effects of selection on *CYP2J40*. Establishing the function of *CYP2J40* would be very helpful in assessing this.

To conclude, we found that *CYP2J19* is widespread in avian lineages which is consistent with a conserved function in colour vision and co-option for red integumentary coloration. Like several other *CYP* loci, including some with conserved functions, *CYP2J19* shows evidence of evolving under positive selection across birds. The cause of the positive selection on *CYP2J19* is unclear. There is no evidence for a change in selection pressure on *CYP2J19* following co-option for red coloration, but one factor may be compensatory mutations related to selection at the adjacent locus *CYP2J40.*

Chapter 6

The evolution of *CYP2J19* copy number and function in indigobirds, whydahs (Viduidae) and estrildid finches (Estrildidae)

Abstract

The recently implicated carotenoid ketolase, *CYP2J19*, shows copy number variation across the passerines. Specifically, two copies of the gene have been found within the estrildid zebra finch in contrast to the single gene copy identified in all other species examined, including the closely related weaverbirds (Ploceidae). Here, I investigate the origin and evolution of the duplication of *CYP2J19* by surveying representative species from Viduidae and Estrildidae, and provide evidence for a single ancestral duplication near the base of the estrildid lineage. Expression analysis supports a role for *CYP2J19* in the peripheral conversion of ketocarotenoids in viduids and estrildids, and highlights a potential role for the carotenoid cleavage enzyme, BCO2, in the control of red coloration.

Introduction

In the previous chapters, I have highlighted the importance of carotenoids, particularly red ketocarotenoids, in avian coloration and colour vision. Many avian carotenoids have been described and several of the metabolic transformations have been outlined (LaFountain *et al.* 2015), but despite this little is known about the genetics underlying carotenoid transport, processing and deposition. Very recently, a candidate locus, *CYP2J19*, was strongly implicated in two mutant bird lineages (zebra finch (Aves: Estrildae) and the red factor canary (Aves: Fringillidae)) as the avian ketolase responsible for the metabolic conversion of yellow dietary xanthophylls into the red ketocarotenoids for colour vision and colouration. To date, no functional evidence has been obtained to establish *CYP2J19* as the avian ketolase, nevertheless the first associative study on a wild bird population, the weaverbirds (Aves: Ploceidae), found compelling evidence for an association between C4-ketocarotenoid presence with high expression levels of *CYP2J19* that is independent of colour (Chapter 4). Although the weaverbird family is closely related to the estrildid zebra finches (De Silva *et al.* 2017), two important differences involving *CYP2J19* can be distinguished: 1) the site of tissue-specific expression of *CYP2J19* for integumentary coloration (centrally in the weaverbirds, peripherally in the zebra finch), and 2) the copy number of *CYP2J19* (single copy in the weaverbirds and two copies in the zebra finch). Notably, *CYP2J19* is present as a single gene copy in a majority of the Aves (Chapter 5), and also in the Testudines (Chapter 3), which strongly suggests that the presence of the two zebra finch loci is the result of a recent duplication of *CYP2J19*.

Gene duplications play a major role in genome evolution and biological diversity (Jaillon *et al.* 2004; Putnam *et al.* 2008); however, post-duplication, many duplicates are lost due to the accumulation of random mutations that results in pseudogenization. The acquisition of differences between the duplicates, either in coding or regulatory regions, allows for the preservation of gene copies, and various models have been used to describe this acquisition process (Innan and Kondrashov 2010). As argued in Chapter 3, gene duplication is responsible for the origin of *CYP2J19* and *CYP2J40*, where the proposed ancestral role for *CYP2J19* was in retinal oil droplet pigmentation and whose function in red ketocarotenoid based

coloration was subsequently co-opted within certain lineages of the birds and the turtles.

In contrast to the function of *CYP2J19* in red carotenoid modification, carotenoid-related genes have also been described for the uptake (Scavenger proteins) and breakdown (carotenoid cleavage enzymes) of yellow carotenoids. Notably, the gene *BCO2,* described for the asymmetric cleavage of these carotenoids, is involved in carotenoid-based coloration and avian vision where it is responsible for the generation of the apocarotenoid galloxanthin present in the clear oil droplets of single cones (see Introduction Chapter for details). Given that *CYP2J19* metabolises yellow carotenoids, the interaction between *BCO2* and *CYP2J19* is an interesting question.

The estrildid finches (Aves: Estrildidae) are a group of brightly coloured seed-eating passerines found across Africa, Asia and Australia (Hoyo *et al.* 2010). The family consists of 134 species and three systematic groups have been recognised: waxbills (mainly African), grassfinches (Australian) and mannikins (spread through the Old World Tropics). Sister to the estrildids are the indigobirds and whydahs (Aves: Viduidae), which consist of 20 species spanning two genera: *Vidua* (19 species) and *Anomalospiza* (Cuckoo finch) (Hoyo *et al.* 2010; De Silva *et al.* 2017). Red coloration is widespread across the estrildids, particularly in the feathers and beak of several genera, with the notable exception of the *Lonchura* clade which lacks any form of redness. By contrast, viduids are less colourful with red coloration almost exclusively found in bare body tissues (beak and tarsus) of a few *Vidua* species. Given the sister relationship between ploceids and the estrildid-viduid clade, together with the copy number variation of *CYP2J19* between the weaverbirds and the estrildid zebra finch, analysis of other viduid and estrildid species will help to elucidate the origin and evolution of *CYP2J19* duplication.

Amongst the estrildids, C4-ketocarotenoids are used for integumentary coloration (McGraw 2004; McGraw and Schuetz 2004) where they can function as honest signals in sexual selection (Burley and Coopersmith 1987). Analysis of blood, liver and integumentary tissues identified C4-ketocarotenoids (astaxanthin, α-doradexanthin, adonirubin and canthaxanthin) only in the integument, establishing the estrildids as peripheral ketoconverters (McGraw 2004; Mundy *et al.* 2016). Moreover, expression of *CYP2J19* within the integument (beak and tarsus) of the zebra finch concords with the peripheral metabolism of red ketocarotenoids (McGraw and Toomey 2010; Mundy *et al.* 2016). Similarly, certain viduids have also been

shown to possess bright ketocarotenoid coloured integuments, and employ peripheral metabolism of dietary carotenoids (McGraw 2004).

In this study, I sample phylogenetically representative species amongst the viduids and estrildids in order to investigate the duplication and function of *CYP2J19*. I first examine whether the gene, *CYP2J19,* is present in the viduids and other estrildid species, and if so, in how many copies. Second, I describe the expression patterns across retinal and extra-retinal tissues and provide support that the integument is the primary anatomic site for ketocarotenoid metabolism in both avian families. Finally, I consider the evolution of *CYP2J19* and its functional relationship with the locus *BCO2*.

Materials and Methods

(a) Samples

19 estrildid species (13 African and 6 Australian) and four viduid species encompassing the two genera *Vidua* and *Anomalospiza* were investigated in this study (Table 6.1). Where possible, both red and non-red integumentary tissues were investigated in this study together with the liver and retina. Unfortunately, due to sampling limitations, feathers of only three estrildid species were analysed here (Table 6.1).

Table 6.1. Sample list of species and tissues investigated in the study. Presence of tissue samples (◆) are shown, and red coloured tissues (◆) are highlighted. *Samples used for Sanger sequencing of *CYP2J19* cDNA. ⁑Samples used for Illumina sequencing of *CYP2J19* genomic amplifications. †Samples used in qRT-PCR.

Passeriform Family	Origin	Species	Presence of red coloration	Individual	Sex	Source	Sourced by	Tissue samples					
								Liver	Retina	Beak	Tarsus	Feather	Other
ESTRILDIDAE	African	Common waxbill (*Estrilda astrild*)	✓	1	M	Cedara, South Africa	Staffan Andersson	◆†⁑	◆*	◆*†			
				2	M	Sao Tomé And Principe	Staffan Andersson	◆†		◆†			
				3	M	South Africa	Staffan Andersson	◆†	◆	◆†	◆		
				4	?	Zambia	Gabriel A. Jamie					◆	
	Africa	Orange-cheeked waxbill (*Estrilda melpoda*)	✓	1	F	Dubreka, Guinea	Fuchs/Douno	◆					
	Africa	Red-billed firefinch (*Lagonosticta senegala*)	✓	1	F	Zambia	Gabriel A. Jamie	◆		◆*			
				2	M	Zambia	Gabriel A. Jamie	◆⁑	◆*	◆*	◆		
	Africa	African firefinch (*Lagonosticta rubricata*)	✓	1	M	Forecariah, Guinea	Fuchs/Douno	◆					
	Africa	Western bluebill (*Spermophaga haematina*)	✓	1	M	Forecariah, Guinea	Fuchs/Douno	◆⁑					
	Africa	Chestnut-breasted nigrita (*Nigrita bicolor*)	×	1	M	Forecariah, Guinea	Fuchs/Douno		◆				
	Africa	Crimson seedcracker (*Pyrenestes sanguineus*)	✓	1	F	Forecariah, Guinea	Fuchs/Douno	◆⁑	◆				
	Africa	Violet-eared waxbill (*Uraeginthus granatinus*)	✓	1	F	Limpopo, South Africa	Fuchs/Wogan/Mandiwana-Neudani	◆	◆				
	Africa	Red-throated twinspot (*Hypargos niveoguttatus*)	✓	1	F	Mozambique	Fuchs/Pons	◆					
	Africa	Cut-throat finch (*Amadina fasciata*)	✓	1	M	Zambia	Gabriel A. Jamie	◆⁑	◆*	◆*	◆		
				2	F	Zambia	Gabriel A. Jamie	◆	◆*	◆*	◆		
	Africa	Blue waxbill (*Uraeginthus angolensis*)	×	1	?	Zambia	Gabriel A. Jamie	◆⁑	◆*	◆*	◆		
	Africa	Bronze mannikin (*Lonchura cucullata*)	×	1	?	Zambia	Gabriel A. Jamie	◆⁑	◆*	◆	◆		
				2	?	Zambia	Staffan Andersson	◆	◆*				
	Africa	Orange-winged pytilia (*Pytilia afra*)	✓	1	?	Zambia	Gabriel A. Jamie	◆⁑				◆ + ◆	
	Australian	Gouldian finch (*Erythrura gouldiae*)	✓	1	M	Australia	Terry Burke						Blood
	Australian	Painted finch (*Emblema pictum*)	✓	1	M	Australia	Simon Griffith	◆	◆*	◆*	◆		Blood
	Australian	Pictorella finch (*Heteromunia pectoralis*)	×	1	M	Australia	Simon Griffith	◆	◆*	◆	◆		Blood
	Australian	Star finch (*Neochmia ruficauda*)	✓	1	M	Australia	Simon Griffith	◆	◆*	◆*	◆	◆	Blood
	Australian	Red-billed long-tail finch (*Poephila acuticauda*)	✓	1	M	Australia	Simon Griffith	◆	◆*	◆*	◆*		Blood
	Australian	Red-eared firetail (*Stagonopleura oculata*)	✓	1	M	Australia	Simon Griffith						Heart
VIDUIDAE	Africa	Village indigobird (*Vidua chalybeata*)	✓	1	M	Zambia	Nicholas I. Mundy & Gabriel A. Jamie	◆†	◆*†	◆*†	◆*†		Muscle⁑
				2	M	Zambia	Nicholas I. Mundy & Gabriel A. Jamie	◆†	◆†	◆†	◆†		
				3	M	Zambia	Nicholas I. Mundy & Gabriel A. Jamie	◆†	◆†	◆†	◆†		
	Africa	Purple indigobird (*Vidua purpurascens*)	×	1	M	Zambia	Gabriel A. Jamie	◆					
	Africa	Pin-tailed whydah (*Vidua macroura*)	✓	1	M	Zambia	Staffan Andersson	◆	◆	◆			Duodenum
	Africa	Cuckoo finch (*Anomalospiza imberbis*)	×	1	?		Claire Spottiswoode						Blood
				2	?		Claire Spottiswoode						Embryo

(b) Primers

See General Experimental Methods (Chapter 2) for details of primer design.

List of primers used for amplification of cDNA sequences:

β-actin	◆*ACTF2	CTCCCTGATGGCCAGGTCAT
	◆*ACTR2	TGGATACCACAGGACTCCAT
CYP2J19	[δ]CJ2ex1EF	GAGAAGATCTAAAATGGAGCTACAGTTTTGGCCTGA
	CJ2ex1R	GCAGTGCAAGATGGAGAGGC
	◆*CYP2J2-2R	TGGTAGTCAAAGCGGTTCCCA
	◆*CYP2J2-2F	CGGCCAAATATTCCACTCCTCC
	*CJ2Aex9R	GCTGAACCTGGGTTGATGGGA
	*CJ2do1R	TTCTGCTGCGCTTCAGGTGAA
	FGCJ2A1ex9R	GAGCCCTTTCCCTACCACCC
	FGCJ2A2ex9R	GGACACCTCAGGAGCCCTTG
	ESTex9R	CCTACCACCCAGATGCCGG
	PLO1ex9R	GTGGTCAGGGGAAGGGGATT
	VC1ex9R	CTGGGGTCGTCAGGAGAAGG
	*CJ2ex9ER	GAGAGAATTCGATTCCCCAGCTTGTGAGGCT
CYP2J40	CJ1ex2F	TGGGCAGCATGAAGTTTGTGA
	CJ1ex4R	TGCCAAAGATGAGGGAGCAGA
	CJ1ex1F3	GGACAGCATCTCCCTTCAGACA
	CJ1ex2F	TGGGCAGCATGAAGTTTGTGA
	*CJ1ex9R2	ATAAACCACCCAGCCTCCCC
	CJ1ex9R4	GGCTTGTTTTCAGCACCCCT
	CJ1ex3F1	AGGCACAGAGGAGGTTCACC
BCO2	BC2shortF	CACAGCTCATCCCCACTACG
	BC2shortR	AGGGGCTGCTCGATGAAAAT
	[◐]BC2ex1F	ATCTGCTGAGGAATGGCCCC
	[◐]BC2ex11R	CAGCGAAGATGCCGTGGAAC

Additional primers used for amplification of gDNA sequences:

CYP2J19	*CJ2up9F	CACACCAGGAAGGCTGGAAGA
	*CJ1ex1R2	GGAAGGGGGTGGGAGGGAAATTA
	CJ1Pex1R	GTCTGGAGGGAGATGCTGTCC
	CJ1Dex1R	GGAATGTCTGGAGGGAGATGCT
	*CJ2up4F	AAGTGGGATGACACAGGGAGC
	CJ2A2upF	GAGACCACTTAGCAGCCCCT
	CJ2A3upF	AAATAGCCAAGGGCTCCTGA
	CJ2ex9R	CGCAGAGCTGGAAGGGTTT
	CJ2ex9F	CCCTGCTGCAGAAGTTCACC

* Primers taken from Mundy *et al.* (2016)
[§] Primers designed by Dr Nicholas I Mundy
[¶] Primers designed by Dr Neil Walsh
◆ Primers used in qRT-PCR

(c) Copy number analysis of CYP2J19

For genomic DNA extraction, long range PCR and product purification, see General Experimental Methods (Chapter 2) for details.

Several primer combinations were applied to investigate *CYP2J19* copy number using the zebra finch genomic organisation as a reference. Primer combinations that spanned the entire *CYP2J19* region (CJ2up9F with CJ1Pex1R/CJ1Dex1R/CJ1ex1R2), *CYP2J19A* only (CJ2up9F with FGCJ2A1ex9R/FGCJ2A2ex9R), *CYP2J19B* only (CJ2A2upF/CJ2A3upF/CJ2up4F with CJ1Pex1R/CJ1Dex1R/CJ1ex1R2) regions were used (Figure 6.1). Nested PCRs using *CYP2J19* exon specific primers (*CYP2J19A*: CJ2ex1EF with FGCJ2A1ex9R/FGCJ2A2ex9R, *CYP2J19B*: CJ2A2upF/CJ2A3upF/CJ2up4F with CJ2ex9R) were carried out on any products >20kbp in order to verify the presence of both copy A and B within the primary amplicon (Figure 6.1).

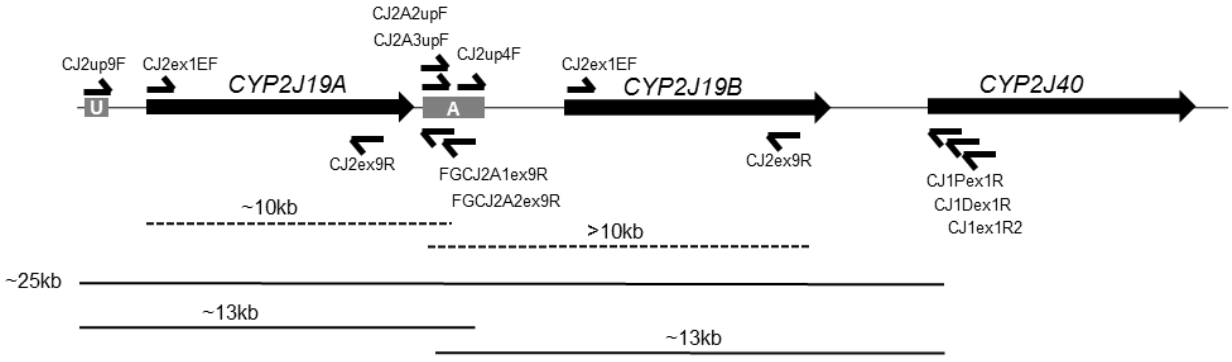

Figure 6.1. Schematic showing the primer locations for gDNA amplification in Estrildidae with two *CYP2J19* copies based on the genomic organisation found in zebra finch (Mundy *et al.* 2016). The expected sizes of long range PCR products are shown, with nested PCR products represented by dashed lines. Grey boxes U and A represent unique sequences upstream and downstream of *CYP2J19A* respectively.

Illumina MiSeq sequencing of purified long range PCR amplicons was carried out in the first instance for *E. astrild CYP2J19A*, *L. senegala CYP2J19A* and *B* and *V. chalybeata CYP2J19*. Sequencing was performed to >1000-fold coverage, and de novo assembly was performed in Seqman NGen (Linux) v.12 (DNASTAR) by Dr Kang-Wook Kim (University of Sheffield). Further sequencing was performed by the DNA Sequencing Facility, Department of Biochemistry at the University of Cambridge. This involved Nextera library preparation, paired end MiSeq sequencing (500-cycles Nano) to >300-fold coverage and sequence assembly, all completed by the Sequencing Facility (Table 6.2).

Table 6.2. List of long range PCR amplicons used in Illumina MiSeq sequencing.

Species	Common name	Primers used for amplification	Genes targeted by primers
Vidua chalybeata	Village indigobird	CJ2up9F/ CY1ex1R2	*CYP2J19*
Estrilda astrild	Common waxbill	CJ2up9F/FGCJ2A2ex9R	*CYP2J19A*
		CJ2A2upF/CJ2ex9R	*CYP2J19B*
Lagonosticta senegala	Red-billed firefinch	CJ2up9F/FGCJ2A2ex9R	*CYP2J19A*
		CJ2A2upF/CJ1ex1R2	*CYP2J19B*
Amadina fasciata	Cut-throat finch	CJ2up9F/FGCJ2A1ex9R	*CYP2J19A*
		CJ2A2upF/CJ2ex9R	*CYP2J19B*
Uraeginthus angolensis	Blue waxbill	CJ2up9F/FGCJ2A1ex9R	*CYP2J19A*
		CJ2A3upF/CJ1Pex1R	*CYP2J19B*
Pytilia afra	Orange-winged pytilia	CJ2up9F/FGCJ2A1ex9R	*CYP2J19A*
		CJ2A3upF/CJ1Dex1R	*CYP2J19B*
Lonchura cucullata	Bronze mannikin	CJ2up9F/CJ1ex1R2	*CYP2J19*
		CJ2up9F/CJ2ex9R	*CYP2J19A*
Pyrenestes sanguineus	Crimson seedcracker	CJ2up9R/FGCJ2A1ex9R	*CYP2J19A*
		CJ2A3upF/CJ2ex9R	*CYP2J19B*
Spermophaga haematina	Western bluebill	CJ2up9F/FGCJ2A1ex9R	*CYP2J19A*
		CJ2A3upF/CJ2ex9R	*CYP2J19B*

(d) Expression analysis

For RNA extraction and first strand synthesis, refer to General Experimental Methods (Chapter 2).

(e) Quantitative RT-PCR

In order to quantitatively assay tissue-specific variation in *CYP2J19* expression, real-time RT-PCR was carried out on two species (one estrildid: *E. astrild* and one viduid: *V.chalybeata*) where tissue samples from three individuals were available. Reactions were performed in an MJ Opticon2 (Research Engines) thermal cycler using the Quantitech SYBRGreen kit (Qiagen) for the livers, beaks, tarsus and retina of three male *V. chalybeata,* as well as for the livers and beaks of three *E. astrild* individuals. Three technical replicates were used for each tissue sample and gene expression was normalised against *β-actin*. The Box-Cox power transformation for normality was applied, and lambda was fixed at 0.15 and 0.3 for statistical analysis of *V. chalybeata* and *E. astrild* datasets respectively. The Shapiro-Wilk test for normality of residuals (*V. chalybeata* $p > 0.6$, *E. astrild* $p > 0.9$), Bartlett's test of equality of variance (*V. chalybeata* $p > 0.9$, *E. astrild* $p > 0.4$) and Runs test for independence of residuals (*V. chalybeata* $p > 0.7$, *E. astrild* $p = 1$) were upheld.

(f) Phylogenetic analysis of viduid and estrildid CYP2J19

Thirty-three sequenced *CYP2J19* coding sequences (1473 bp) across 13 estrildids, one viduid and one ploceid were aligned in MEGA v.6 (Tamura *et al.* 2013) using MUSCLE (Edgar 2004) (30 from this study, see Table 6.1, two zebra finch sequences from Mundy *et al.* 2016, and one ploceid sequence from Chapter 5). Phylogenetic reconstruction of DNA sequences was carried out by maximum-likelihood in PhyML, using Smart Model Selection based on Bayesian information criterion (http://www.atgc-montpellier.fr/phyml/) (Guindon *et al.* 2010). The selected model was K80 + G (0.834) + I (0.635). Branch support was assessed with 1000 bootstrap pseudo-replicates.

Results

(a) CYP2J19 copy number determination

For viduids, four species (three *Vidua* and one *Anomalospiza*) were analysed (Table 6.3). Long-range PCR identified products (~13kbp) of similar lengths to the zebra finch *yellowbeak* mutant, which possesses a single *CYP2J19* copy (Mundy *et al.* 2016) in all four species. In order to verify the amplification of *CYP2J19*, nested secondary PCRs were carried out on the ~13kbp amplicon of the cuckoo finch, the purple indigobird and the pin-tailed whydah, using primers specific to zebra finch *CYP2J19A* (CJ2ex1EF with CJ2Aex9R/FGCJ2A1ex9R, see Materials and Methods). Products of ~10kbp were isolated from the nested PCR reactions, which suggests that the single viduid *CYP2J19* is comparable with zebra finch *CYP2J19A*. In addition, Illumina MiSeq sequencing of the village indigobird ~13kbp product further confirms amplification of the target gene *CYP2J19*. Comparisons of full length *CYP2J19* cDNA found complete sequence match between different tissues of the same individual (retina, beak and tarsus of *V. chalybeata* and retina and beak of *V. macroura*), which supports the presence of a single functional *CYP2J19* gene copy.

Table 6.3. Summary of long range PCR results for three viduid species. + denotes the presence of target product.

Species	**Individual**	**Presence of ~13kbp *CYP2J19* amplicon**	**Nested PCR on ~13kbp product using *CYP2J19A* specific primers**
Village indigobird (*Vidua chalybeata*)	1	**+***	Not determined
Purple indigobird (*Vidua purpurascens*)	1	**+**	**+**
Pin-tailed whydah (*Vidua macroura*)	1	**+**	**+**
Cuckoo finch (*Anomalospiza imberbis*)	1	**+**	**+**
	2	**+**	**+**

*Long range PCR product sequenced using Illumina MiSeq sequencing (see Method section for details).

In contrast to the viduids, long range PCR amplification using primers that span the entire *CYP2J19* region upstream of *CYP2J40* generated ~25kbp amplicons for 12 of the 19 estrildid species, corresponding to two copies of *CYP2J19* (Table 6.4). Of the seven species that failed to yield a ~25kbp product, six did not produce any high molecular weight bands using the primers spanning the entire array, whilst the remaining species (bronze mannikin) yielded a surprising product size of ~15kbp (Table 6.4). Separate amplifications for copies A and B successfully isolated band sizes of ~10-15kbp for all 19 species, and it appears that a combination of poor gDNA quality and/or poor primer pairing likely explains the lack of >20kbp products in the seven species.

To verify the presence of two *CYP2J19* gene copies within the 12 species that produced ~25kbp products, nested secondary PCR reactions were carried out using *CYP2J19A* and *B* specific primers. Strong bands of 10-15kbp, which corresponds to a single *CYP2J19,* were resolved from the ~25kbp amplicons using copy A and B primers independently (Table 6.4).

Sequence confirmation of long-range PCR products was performed for eight species (Table 6.4). All 9 exons of *CYP2J19* were identified from sequenced *CYP2J19A-like* and *B-like* amplicons for all species except the bronze mannikin. Surprisingly, two contigs, similar in both sequence and length (<10kbp) were assembled for each of the two mannikin long-range PCR products sequenced (Table 6.4). Comparisons with sequenced mannikin retinal *CYP2J19* cDNA confirmed assembly of genomic *CYP2J19*, with eight identifiable exons present in the sequenced *CYP2J19A-like* product, and six exons present in the ~15kbp product obtained from the full amplification for >20kbp bands. The presence of two contigs per sequenced genomic amplicon in the bronze mannikin suggests that two copies of *CYP2J19* were amplified using what are ‘copy-specific’ primers in other estrildids.

Given the differential tissue specific expression patterns of *CYP2J19A* and *B* in the zebra finch, I further confirmed the presence of two gene copies in estrildids by demonstrating that the full length *CYP2J19* cDNA from retinal and non-retinal (beak and tarsus) tissues of the same individual, that matched genomic copy A and B respectively, were distinct in four species (*E. astrild*, *L. senegala, A. fasciata* and *U. angolensis*).

Table 6.4. Summary long range PCR results for *CYP2J19* from 13 estrildid species. +/- denotes presence/absence of target amplicon.

Species	**Presence of ~25kbp amplicon**	**Nested PCR on ~25kbp product (where applicable)**		**Presence of *CYP2J19A-like* amplicon**	**Presence of *CYP2J19B-like* amplicon**
		Presence of *CYP2J19A*	Presence of *CYP2J19B*		
Common waxbill (*Estrilda astrild*)	**+**	**+**	**+**	**+***	**+***
Orange-cheeked waxbill (*Estrilda melpoda*)	**-**			**+**	**+**
Red-billed firefinch (*Lagonosticta senegala*)	**+**	**+**	**+**	**+***	**+***
African firefinch (*Lagonosticta rubricata*)	**+**	**+**	**+**	**+**	**+**
Western bluebill (*Spermophaga haematina*)	**+**	**+**	**+**	**+***	**+***
Chestnut-breasted nigrita (*Nigrita bicolor*)	**-**			**+**	**+**
Crimson seedcracker (*Pyrenestes sanguineus*)	**+**	**+**	**+**	**+***	**+***
Violet-eared waxbill (*Uraeginthus granatinus*)	**-**			**+**	**+**
Red-throated twinspot (*Hypargos niveoguttatus*)	**-**			**+**	**+**
Cut-throat finch (*Amadina fasciata*)	**+**	**+**	**+**	**+***	**+***
Blue waxbill (*Uraeginthus angolensis*)	**+**	**+**	**+**	**+***	**+***
Bronze mannikin (*Lonchura cucullata*)	**-** (~15kbp product ***)			**+***	**+**
Orange-winged pytilia (*Pytilia afra*)	**+**	**+**	**+**	**+***	**+***
Gouldian finch (*Erythrura gouldiae*)	**-**			**+**	**+**
Painted finch (*Emblema pictum*)	**+**	**+**	**+**	**+**	**+**
Pictorella finch (*Heteromunia pectoralis*)	**+**	**+**	**+**	**+**	**+**
Star finch (*Neochmia ruficauda*)	**+**	**+**	**+**	**+**	**+**
Red-billed long-tail finch (*Poephila acuticauda*)	**+**	**+**	**+**	**+**	**+**
Red-eared firetail (*Stagonopleura oculata*)	**-**			**+**	**+**

*Long range PCR products sequenced using Illumina MiSeq sequencing.

(b) Expression patterns of CYP2J19

CYP2J19 was expressed strongly in the retina and red peripheral tissues (red feathers, beak and tarsus) of all viduid and estrildid species studied, with low or undetectable expression in the liver (Table 6.5 and 6.6). Weak expression levels were found in the pin-tailed whydah duodenum, which is the site of carotenoid uptake. Highly significant differences in *CYP2J19* expression were observed between the liver and all other tissues (beak, tarsus and retina) of the village indigobird ($F_{3,7} = 80.39$, $p < 0.001$) (Figure 6.2), which possesses a red beak and tarsus; as well as between the liver and the red ketocarotenoid pigmented beaks of the common waxbill (two-tailed t-test, $p = 0.005$) (Figure 6.3).

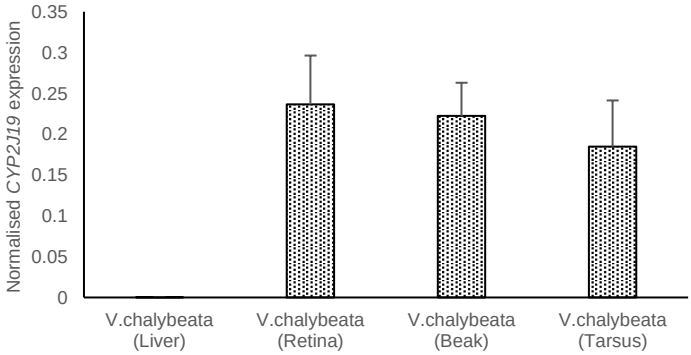

Figure 6.2. qRT-PCR quantification of *CYP2J19* expression in the liver, retina, beak and tarsus of the village indigobird (*V. chalybeata*) normalised against *β-actin*. Error bars represent SEM from three individuals.

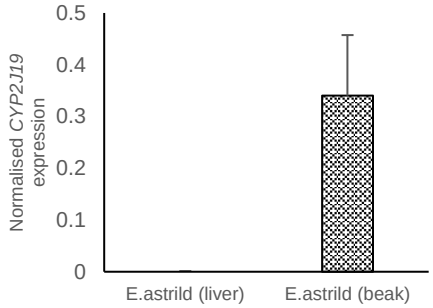

Figure 6.3. qRT-PCR quantification of *CYP2J19* expression in the liver and beak of the common waxbill (*Estrilda astrild*) normalised against *β-actin*. Error bars represent SEM from three individuals.

Contrary to expectations, strong *CYP2J19* expression was also found in two estrildid integumentary tissues that lack apparent red coloration: the black beak of the cut-throat finch (*Amadina fasciata*) and the black feathers of the common waxbill (*Estrilda astrild*) (Table 6.6). Given the uncoupling of high *CYP2J19* expression with redness, I considered the possible involvement of other candidate genes for carotenoid colouration (see Introduction Chapter for details), focusing on genes involved in yellow carotenoid breakdown, namely *BCO2*.

In all red peripheral tissues, I found weak or undetectable levels of *BCO2*, with more variable expression in the non-red peripheral tissues (Table 6.5 and 6.6). By contrast, expression of *BCO2* was unable to explain the lack of redness in the cut-throat finch (*Amadina fasciata*) beaks given strong *CYP2J19.* Here, no detectable expression of *BCO2* was found in the beaks. Comparison of sequenced *CYP2J19* cDNA from the retina and the beak of two *A. fasciata* individuals (one male and one female) revealed a 29 amino acid deletion at the 3' end of exon 7 that is unique to the beak tissue. Alignment of retinal and beak cDNA with sequenced genomic DNA confirmed that the beak specific deletion is a form of post-transcriptional modification of *CYP2J19B*. The deleted region overlaps with the predicted 3rd HEM binding domain of CYP2 proteins (Almeida *et al.* 2016), and presumably results in protein malfunction, and can explain the lack of apparent red given strong *CYP2J19* and weak *BCO2*. As the pink and black regions of the blue waxbill beak were not analysed separately, the observed strong expressions of *BCO2* and *CYP2J19* likely reflect expression patterns contained in both coloured regions. In addition, strong *BCO2* expression was also found in the retina of all species examined (one viduid and ten estrildids), which concords with the role of *BCO2* in the production of galloxanthin (Toomey *et al.* 2016) (Table 6.6). *CYP2J40* was found across all tissue types, which does not suggest any involvement in carotenoid pigmentation (Table 6.5 and 6.6).

Table 6.5. Qualitative expression patterns of *CYP2J19, BCO2* and *CYP2J40* for two viduid species.

Species	**Tissue**	**Gene expression**		
		CYP2J19	***BCO2***	***CYP2J40***
Village indigo-bird (*Vidua chalybeata*)	Liver	○		●
	Retina	●		●
	Beak (red)	●		●
	Tarsus (red)	●		●
Pin-tailed whydah (*Vidua macroura*)	Liver	-	-	●
	Retina	●	●	●
	Beak (red)	●	-	○
	Duodenum	-	-	●

Strong (●), weak (○) and undetectable (-) expression levels are shown. Gaps in table were not determined.

Table 6.6. Summary of qualitative expression patterns of *CYP2J19* (both A and B copies), *CYP2J19A, CYP2J19B, BCO2* and *CYP2J40* across different tissues in 13 estrildid species.

Origin	Species	Sex	Tissue	Gene Expression				
				CYP2J19	*CYP2J19A*	*CYP2J19B*	*BCO2*	*CYP2J40*
AFRICAN	Common waxbill (*Estrilda astrild*)	M	Liver	○			○	●
			Retina	●	●	○	●	●
			Beak (red)	●	-	●	○	●
			Tarsus (black)	○			●	●
		?	Feather (black)	●	-	●	●	●
	Orange-cheeked waxbill (*Estrilda melpoda*)	F	Liver	-				●
	Red-billed firefinch (*Lagonosticta senegala*)	M	Liver	-			-	●
			Retina	●	●	○	●	●
			Beak (red)	●	-	●	-	●
			Tarsus (pale)	○			-	●
		F	Liver	-				●
			Beak (red)	●	-	●		●
			Feather (brown)	-				●
	African firefinch (*Lagonosticta rubricata*)	M	Liver	-				●
	Western bluebill (*Spermophaga haematina*)	M	Liver	-				●
	Chestnut-breasted nigrita *(Nigrita bicolor*)	M	Retina	●	●	-		●
	Crimson seedcracker (*Pyrenestes sanguineus*)	F	Liver	-				●
			Retina	●	●	-		○
	Violet-eared waxbill (*Uraeginthus granatinus*)	F	Liver	-				●
			Retina	●	●	-		●
	Red-throated twinspot (*Hypargos niveoguttatus*)	F	Liver	-				●
	Cut throat finch (*Amadina fasciata*)	M	Liver	-			-	●
			Retina	●	●	○	●	●
			Beak (black)	●	-	●	-	●
			Tarsus (pale)	-			-	●
		F	Liver	-			-	●
			Retina	●	●	○	●	●
			Beak (black)	●	-	●	-	●
			Tarsus (pale)	-			-	●
	Blue waxbill (*Uraeginthus angolensis*)	?	Liver	-			-	●
			Retina	●	●	○	●	●
			Beak (pink and black)	●	-	●	●	●
			Tarsus (black)	○			●	●
	Bronze mannikin (*Lonchura cucullata*)	?	Liver	-			●	●

			Retina	●			●	●
			Beak (black)	-			●	●
			Tarsus (black)	-			●	●
	Orange-winged pytilia (*Pytilia afra*)	?	Feather (red)	●	-	●	-	●
			Feather (black)	○			-	●
AUSTRALIAN	Painted finch (*Emblema pictum*)	M	Liver	○			-	●
			Retina	●	●	○	●	●
			Beak (red)	●	-	●	-	●
			Tarsus (black)	-			-	-
	Pictorella finch (*Heteromunia pectoralis*)	M	Liver	-			-	●
			Retina	●	●	○	●	●
			Beak (black)	-			○	●
			Tarsus	○			-	●
	Star finch (*Neochmia ruficauda*)	M	Liver	-			-	●
			Retina	●	●	○	●	●
			Beak (red)	●	-	●	-	●
			Tarsus (yellow)	-			-	●
			Feather (red)	●				○
	Red-billed long-tail finch (*Poephila acuticauda*)	M	Liver	-			-	●
			Retina	●	●	○	●	●
			Beak (red)	●	-	●	-	●
			Tarsus (red)	●	-	●	-	●

Strong (●), weak (○) and undetectable (-) expression levels are shown. Gaps in table were not determined.

Across all estrildid species studied, a single *CYP2J19* copy (*CYP2J19A*) was found to be expressed in the retina, and the second gene copy, *CYP2J19B*, showed strong expression in the integumentary tissues (beak, tarsus and feathers), with variable expression in the retina (Table 6.6), similar to the pattern in the zebra finch (Mundy *et al.* 2016). Alignment of cDNA with sequenced gDNA from 4 species (*E. astrild, L. senegala, A. fasciata* and *U. angolensis)* confirmed tissue specific patterns of copy expression. Within the bronze mannikin, I found complete sequence match between retinal *CYP2J19* cDNA with one of the two assembled genomic copies, which, based on the pattern of expression can be identified as *CYP2J19A.*

(c) Evolution of CYP2J19 in estrildid finches

Phylogenetic reconstruction showed estrildid *CYP2J19* paralog within a species were more closely related to one another (≥83% bootstrap support), with the exception of *S. haematina*, than they are to their orthologous counterparts in other species (Figure 6.4). In most cases, the reconstructed relationship parallels the species phylogeny for the clade (Figure 6.5), with the majority of African estrildids grouped apart from their Australian relatives (100% bootstrap support), except for the bronze mannikin (*L. cucullata*). The African cut-throat finch (*A. fasciata*) clusters however with the Australian species with weak bootstrap support (43%). Paralogous sequences of the same individual were divergent by 0.20-2.25% in all species except the cut-throat finch with a divergence value of 6.34%, of which 5.93% is attributed to the beak specific deletion of *CYP2J19B* cDNA. The reconstructed relationship strongly supports the monophyly of estrildid *CYP2J19* with the single viduid (*V. chalybeata*) gene represented as the unduplicated *CYP2J19*.

No amino acids were completely associated with retinal or integumentary specific *CYP2J19*. However, two sites (221 and 390) were partially conserved amongst the *CYP2J19B* orthologs. Seven species, six African (*L. senegala, P. sanguineus, S. haematina, U. angolensis, A. fasciata* and *E. astrild*) and one Australian (*N. ruficauda*) possess the amino acid valine (V) at site 390 and methionine (M) at the same site of *CYP2J19A*, and three of these (*L. senegala, P. sanguineus* and *S. haematina)* also have tyrosine (Y) at site 221 instead phenylalanine (F) in *CYP2J19A*. Site 221 falls within the predicted functional domain SRS2_SRS3 outlined in Almeida *et al.* 2016.

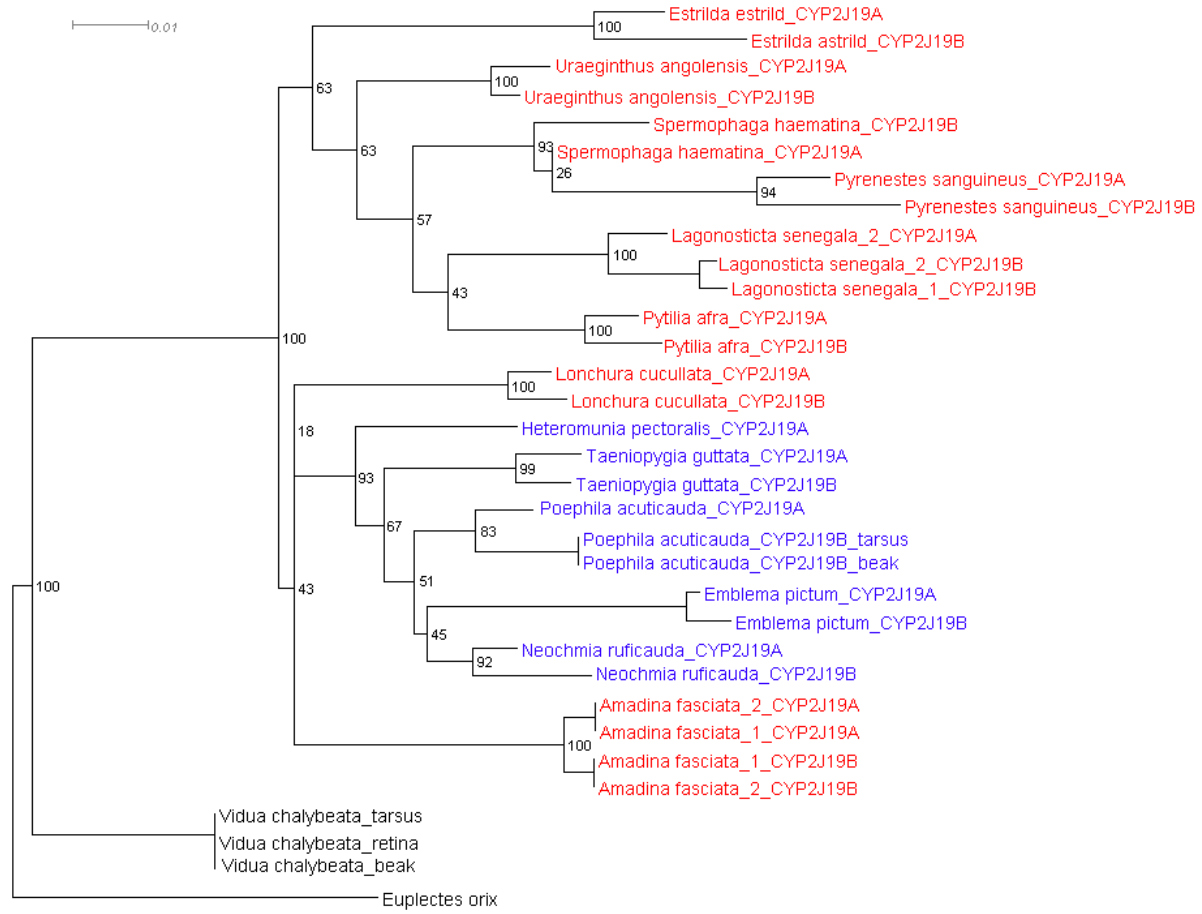

Figure 6.4. Maximum likelihood phylogenetic reconstruction of *CYP2J19,* based on 33 *CYP2J19* coding sequences (30 from this study, two zebra finch sequences from Mundy *et al.* 2016, and *E. orix CYP2J19* from Chapter 5) from 13 estrildid finches (8 African in red and 5 Australian in blue), one viduid (*V. chalybeata*) and one ploceid (*E. orix*). All sequences were obtained from tissue-specific-derived cDNA of *CYP2J19* except for 3 species (*S. haematina, P. anguineus* and *P. afra*) where coding sequences were derived from sequenced genomic *CYP2J19*. Tip numbers represent different individuals of the same species. The identical *CYP2J19B* sequences from tarsus and beak of a single *P. acuticauda* are included.

Discussion

Copy number analysis revealed a single *CYP2J19* locus in all viduids studied, including three species of *Vidua* and *Anomalospiza*. This is similar to the weaverbirds (Chapter 4), and birds in general (Chapter 5). In contrast, two *CYP2J19* copies are present in all estrildids studied (19 species, 13 African and 6 Australian), which represents most major estrildid lineages (Figure 6.5). Given the location of the primers used for copy number analysis, the genomic organisation of the two CYPs is likely the same as in the zebra finch, with copies A and B located in tandem, upstream of *CYP2J40*.

In almost all cases, the two gene copies appear to be more closely related to one another than they are to their orthologues in other species (Figure 6.4), which can be attributed to two potential evolutionary scenarios. The first is a single duplication event followed by extensive gene conversion between the duplicate copies thereby making paralogous sequences more similar to each other than they are to their orthologous counterparts of another species. The second scenario for the observed pattern is multiple independent species-specific duplication events. This is highly unlikely given the conserved amino acid sequences in the paralogues together with the similar genomic organisation of the duplicate CYP loci amongst estrildid species. Overall, these findings support a single estrildid-specific duplication of *CYP2J19* prior to the split between the African and Australian finches, followed by conservation of the paralog post-duplication. As a result of sampling limitations, particularly with regards to the lack of *Amandava* species, which are an outgroup to the African and Australian estrildids, it is yet to be determined whether the duplication of *CYP2J19* occurred at the base of the estrildid clade.

Expression patterns of *CYP2J19* in viduids suggest that the single gene copy is involved in both colour vision and red ketocarotenoid-based coloration (McGraw 2004), and that the gain of red coloration likely occurred via the independent co-option of *CYP2J19* function in colour vision. By contrast, the two estrildid *CYP2J19* show differential tissue specific expression patterns, similar to that of the zebra finch (Mundy et al. 2016) where one copy is solely expressed in the retina for colour vision (*CYP2J19A*) and the other has strong expression for integumentary coloration and weak or undetectable expression in the retina (*CYP2J19B*).

The pattern of tissue-specific expression of *CYP2J19* argues for the peripheral ketoconversion of yellow carotenoids in the estrildids and viduids for coloration. In contrast with the ploceids, where strong hepatic expression of *CYP2J19* is associated with the presence of integumentary red C4-ketocarotenoids (Chapter 4), I found weak or undetectable expression of *CYP2J19* in the livers of all species studied, with strong expression in the retina and all red coloured integumentary tissues, including the bare body parts (beak and tarsus) and feathers. The observation of *CYP2J19* in feathers builds on what has been described in the zebra finch, where only expression in bare body parts was considered (Mundy *et al.* 2016). Significantly lower levels of *CYP2J19* expression were found in the livers of a viduid (*V. chalybeata*) and an estrildid (*E. astrild*) compared with red peripheral tissues, which is similar to what was found in zebra finch (Mundy *et al.* 2016). The role of peripheral *CYP2J19* expression in carrying out ketolation for red integumentary ketocarotenoids fits well with the absence of ornamentation carotenoids from the liver and blood plasma in estrildids and viduids (McGraw 2004). Given that the ploceids are the closest outgroup to the viduid-estrildid clade (De Silva *et al.* 2017), it is surprising to find that the families employ such different mechanisms for ketocarotenoid pigmentation of their integument.

CYP2J19 expression levels are not completely associated with conspicuous red coloration in estrildid finches. Although strong gene expression was detected in all red peripheral tissues, strong *CYP2J19* expression was also found in non-red tissues, namely the black beak of the cut-throat finch and the black feather follicles of the common waxbill. The discordance between strong gene expression and lack of redness in the cut-throat finch beak appears to be the result of a post-transcriptional inactivation of *CYP2J19* function via a deletion within the predicted 3rd HEM binding domain of the *CYP2J19B* transcript (Almeida *et al.* 2016). In addition, the cut-throat finch also has conspicuous red plumage, which likely contains a complete *CYP2J19B* transcript expressed in the feather follicles of red but not non-red feathers. This remains, however, to be confirmed due to sampling limitations. Notwithstanding, given that the beak-specific deletion of *CYP2J19B* transcript is absent from the genomic sequence and that a full-length *CYP2J19B* is presumably needed for red plumage pigmentation, the deleted transcript likely represents a beak-specific splice variant of *CYP2J19B*. Here I present the first evidence of a transcriptomic regulation of *CYP2J19* function that is apart from the level of gene expression. The reason why

CYP2J19 is expressed in the beak at all, given that it is not used in ketoconversion, may indicate a potential loss of beak-specific coloration from a red C4-ketocarotenoid coloured ancestor. As a result, investigation of the mechanism of regulation of *CYP2J19* function has the potential to inform us about the evolution of red carotenoid-based coloration.

By contrast, a complete *CYP2J19* transcript is expressed in the black feather follicles of the common waxbill (*E. astrild*). However, analysis of the carotenoid cleavage enzyme, *BCO2*, revealed strong expression in the same tissue. Since *BCO2* likely cleaves the yellow carotenoids on which *CYP2J19* acts, this could explain the lack of apparent redness in the face of strong expression of intact *CYP2J19* transcript. Assuming the epistatic interactions between the two genes, whereby *BCO2* acts upstream of *CYP2J19*, several predictions can be made: 1) strong *BCO2* results in integument lacking in carotenoids (red and yellow), irrespective of *CYP2J19* expression, 2) low (or a lack of) *BCO2* expression together with strong *CYP2J19* produces red C4-ketocarotenoid containing integument, and 3) low (or a lack of) *BCO2* and low *CYP2J19* results in yellow carotenoid pigmented integument, given the delivery of precursor carotenoids to peripheral tissues. Although no characterisation of carotenoid content has been made for any of the species examined here, except for the C4-ketocarotenoid-containing beaks of the common waxbill (*E. astrild*), star finch (*N. ruficauda*) and the pin-tailed whydah (*V. macroura*) (McGraw 2004), there are, to our knowledge, no documented cases where red coloured beaks and tarsi are not due to the presence of C4-ketocarotenoid. So, it is reasonable to assume that redness of these bare body parts is caused by C4-ketocarotenoid pigmentation. Accordingly, in all red coloured tissues, weak or undetectable *BCO2* was found together with strong *CYP2J19*. Moreover, the integumentary tissues (beak, tarsus and feathers) that have strong *BCO2* do not possess conspicuous red or yellow hues, irrespective of *CYP2J19*. Hence, the results so far provide support for the predicted interactions between *BCO2* and *CYP2J19*. Interestingly, in two species (*L. senegala* and *A. fasciata*), the peripheral tissues (tarsi) that lack strong expression of both *BCO2* and *CYP2J19* are pale, suggesting a lack of delivery of yellow precursor carotenoids into these tissues, which contrasts with the yellow tarsi of the star finch (*N. ruficauda*) where neither gene is expressed. This highlights a potential role for carotenoid transport proteins in the regulation of estrildid integumentary coloration, similar to the carotenoid-based

plumage coloration of the common canary, where the transport protein SCARB1 is an important mediator (Toomey *et al.* 2017). Hence, future characterisation of tissue-specific carotenoid content together with an understanding of the genetics underlying carotenoid transport and metabolism is essential in describing avian carotenoid coloration.

Several models of evolution via gene duplication have been proposed to explain the ways in which gene copies can be preserved post-duplication (Innan and Kondrashov 2010). Given the involvement of *CYP2J19* in red carotenoid-based coloration, distinguishing between the different models in this case will depend on a reconstruction of estrildid colour evolution. If the reconstructed ancestral estrildid lacks red ketocarotenoid-based coloration, then the function of *CYP2J19* in coloration likely evolved via neofunctionalization through the gaining of *CYP2J19B*-specific role in integumentary coloration. Under such a scenario, the duplication of *CYP2J19* can be seen as a means of escape from pleiotropy (where a single gene takes on additional functions), which contrasts with the gain of pleiotropy in all the C4-ketocarotenoid pigmented lineages investigated so far which have a single *CYP2J19* gene (Chapter 3, 4 and 5). On the other hand, if the ancestral estrildid possessed red ketocarotenoid-based coloration, two models of evolution can be distinguished: escape from adaptive conflict (EAC) and subfunctionalisation (Innan and Kondrashov 2010). Under EAC, the bifunctional single gene is constrained by pleiotropy to improve either of its functions independently, and as a result, gene duplication allows the separate gene copies to fully take on one of its ancestral functions and become specialised for said function via selection. This contrasts with the subfunctionalisation scenario where there is no improvement of ancestral functions. Evidence for functional specialisation can be found in the coding sequence of *CYP2J19B* where seven of the 13 estrildids possess a substituted amino acid (V) at site 390, of which three species also carry a second amino acid substitution (Y) at site 221. Given that the unduplicated viduid *CYP2J19* sequence, which matches the amino acid at *CYP2J19A* 390 (M) and 221 (F), the substituted sites of *CYP2J19B* likely correspond to derived changes. Hence the preservation of these mutations in the face of widespread gene conversion likely represents selection for functional specialisation of *CYP2J19B* for integumentary coloration within these lineages. In such a case, the data argue against the subfunctionalisation model and future reconstruction of

estrildid colour evolution will be needed to distinguish between neofunctionalization and EAC.

To conclude, a single copy of *CYP2J19* was found within the viduids, consistent with its function in colour vision and co-option for red ketocarotenoid based coloration. Two gene copies (*CYP2J19A* and *B*) were found across the estrildids and evidence points towards a single estrildid-specific duplication of *CYP2J19* near the base of the estrildid lineage followed by extensive gene conversion between the paralogs in each lineage. The lack of apparent redness given high *CYP2J19* expression can be explained by the interaction between *BCO2* and *CYP2J19*, and I describe the first incidence of a mechanism for the regulation of *CYP2J19* function apart from changes in levels of gene expression. As highlighted, the limitations of this study relate particularly to the variable tissue sampling of different species, along with a lack of characterisation of carotenoid content in the various tissues. Hence, more comprehensive sampling together with detailed analysis of genetic and pigmentary mechanisms will be crucial in the understanding of the evolution of red carotenoid-based coloration.

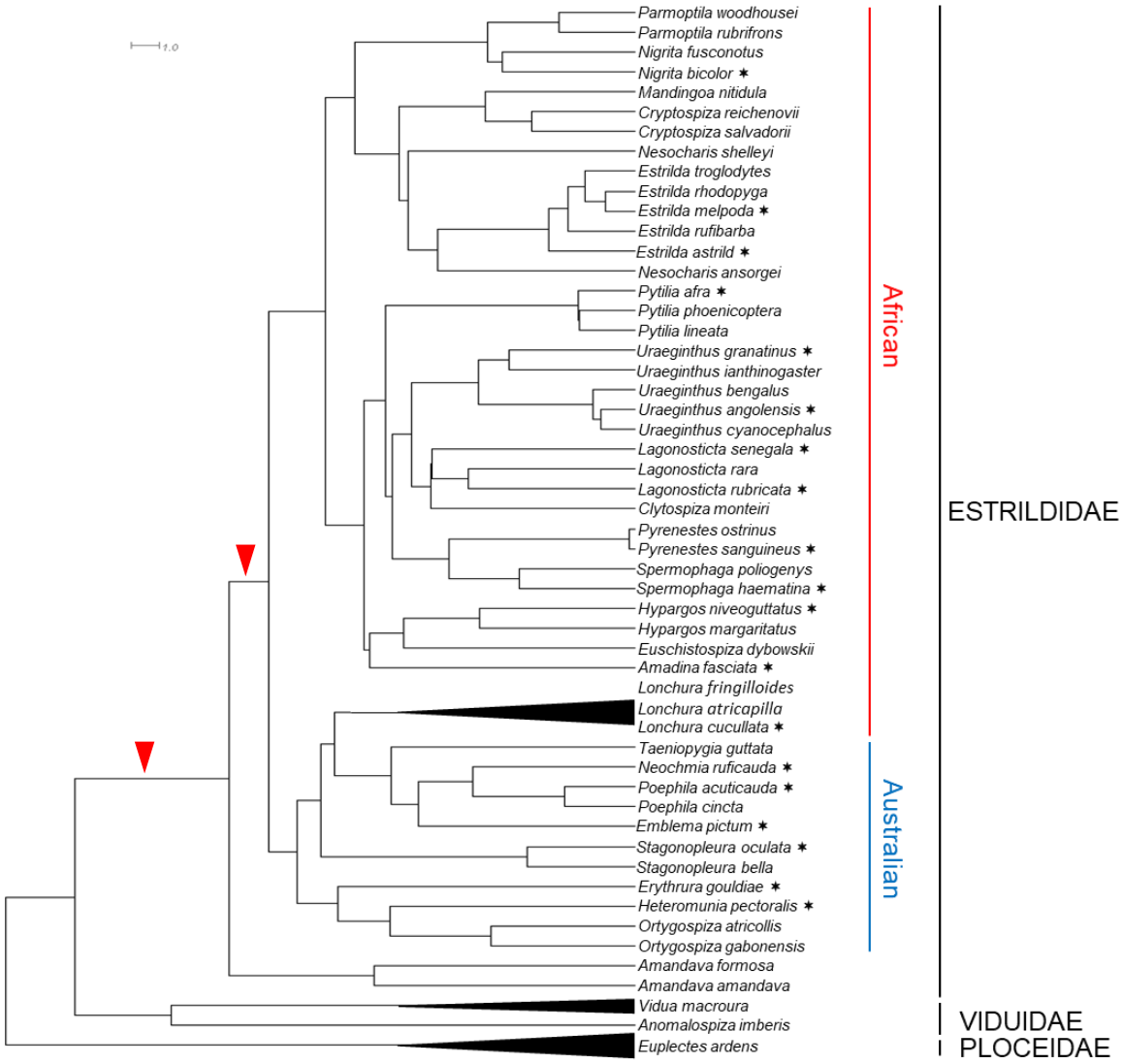

Figure 6.5. Phylogeny of all major estrildid genera taken from birdtree.org (Jetz *et al.* 2012). African and Australian species are highlighted, along with the 19 estrildid species investigated in this study (✶). Red triangles represent possible lineages in which *CYP2J19* duplication occurred.

Chapter 7

Concluding thoughts

This thesis builds on the findings of two recent papers in which the gene, *CYP2J19*, was described to be associated with red C4-ketocarotenoid pigmentation. I attempt to characterise the evolution and function of *CYP2J19* across several reptilian lineages, largely focusing on birds, and provide strong support for the apparent universal involvement of the *CYP2J19* in both red ketocarotenoid-based coloration and red retinal oil droplet pigmentation. The results presented here inevitably raise new questions and provide avenues for future research.

Genetics of CYP2J19

Function

So far, all investigation of *CYP2J19* tests the association between gene expression and red carotenoid pigmentation, and lacks any functional demonstration of the protein's enzymatic activity. This is mainly due to the lack of a model system in which to study carotenoid genetics in birds. An alternative possibility, however, would be to shift to a cell-based system, such as in the single-celled eukaryote *S. cerevisiae* (budding yeast) or *S. pombe* (fission yeast), where ectopic expression of the membrane-bound avian *CYP2J19* followed by introduction of yellow precursor carotenoids may be used to confirm the enzyme as a carotenoid ketolase. An intriguing question that remains to be answered concerns the subcellular localisation of avian CYP2J19 protein, which can relate to the function of carotenoids as a direct signal for the efficiency of cellular respiration (see Introduction). Eukaryotic cytochrome P450s are all membrane-bound proteins found either on the endoplasmic reticulum (ER) or mitochondria, and contain either ER or mitochondrial-targeting sequences in the amino-terminal end of the newly translated precursor peptide. Thus, if CYP2J19 catalyses ketoconversion of yellow carotenoids on the mitochondrial membrane, then it is highly plausible that the presence of red ketocarotenoids serves as a direct signal for mitochondrial efficiency. Moreover, the

subcellular location of CYP2J19 protein could possibly explain the apparent epistatic interaction between *CYP2J19* and *BCO2,* such that imported yellow carotenoids are broken down by BCO2 before CYP2J19 can gain access to them for ketoconversion. Therefore, an understanding of protein dynamics could potentially shed light on several areas of *CYP2J19* function.

Cis-regulation

As argued in Chapter 3, the involvement of *CYP2J19* in colour vision was the most likely ancestral gene function, and its function in C4-ketocarotenoid based coloration was secondarily co-opted. Cis-regulation of *CYP1J19* is predicted to be the main mechanism employed by many bird lineages for this co-option (Chapter 3, 4 and 5), which lends itself to the question of what are the cis-regulatory modules involved? An interesting parallel is with the invertebrate gene *optix,* an intronless homeobox transcription factor gene shown to be involved in fruit fly eye and wing development (Seimiya and Gehring 2000). In *Heliconius* butterflies, *optix* plays a central role in the control of red wing patterning (Reed *et al.* 2011; Martin *et al.* 2014), and is likely also used in butterfly eye development. Small cis-regulatory regions have been found to be associated with the co-option of *optix* from an ancestral role in eye pigmentation, or alternatively from an ancestral role in wing development, given the involvement of *optix* in fly wings (Jiggins *et al.* 2017).

Cis-regulatory modules are generally poorly annotated in the genome. This is largely due to the high degree of variability in the module sequence content, length and proximity to the transcription start site. However, recent advances in next-generation sequencing techniques, together with the availability of whole genome sequences, has led to the development of several computational and experimental approaches in order to identify cis-regulatory modules (Suryamohan and Halfon 2015). As *CYP2J19* functions in both coloration and colour vision, and as the estrildid *CYP2J19A* is solely expressed in the retina for colour vision (Chapter 6), comparisons between the genomic sequences upstream of the duplicate loci for regulatory sequences unique to *CYP2J19A,* together with examination of the identified regions for clusters of transcription factor binding sites, may lead to potential cis-regulatory sequences that drive retina-specific expression. Moreover, as *CYP2J19* is apparently responsible for red oil droplet pigmentation in LWS cones, we would predict that the active binding sites belong to those transcription factors that

are upregulated in the LWS cones relative to all other cone types. Experimental validation of candidate cis-regulatory modules activity is crucial to confirm the identity of the modules in the regulation of gene expression. This can be done by testing the predicted modules using a reporter assay, whereby the candidate module is placed in a reporter construct and transfected into the cell type of interest. Multiple constructs may be injected and RNA-seq determination of transcript abundance can be used to quantify module activity (Suryamohan and Halfon 2015).

A comparative approach may also be used to determine regulatory modules responsible for red carotenoid-based coloration (Suryamohan and Halfon 2015). One possibility is to compare non-coding sequences of red vs. yellow members of the ploceid family; however, given that the species are closely related, sequence similarities amongst red species may not necessarily represent true cis-regulatory modules. Comparison between the upstream sequences of red turtles and birds is another possibility, assuming that the gain of red coloration in both lineages is due to convergent mutations in *CYP2J19* regulatory regions. Again, experimental validation of potential cis-regulatory modules will be essential, particularly to eliminate false-positives (sequences that fail to drive reporter gene expression in report assays). Identification of the regulatory elements underlying patterns of *CYP2J19* expression lends itself to the understanding of possible genetic constraints to the gain of red ketocarotenoid-based coloration.

Site of ketoconversion

Conventionally, characterisation of carotenoid content in the liver and blood plasma has been used to identify the site of ketoconversion (McGraw 2004; del Val *et al.* 2009; Prager *et al.* 2009). However, with the identification of the candidate ketolase, *CYP2J19*, tissue-specific expression patterns can be used to investigate the location of ketoconversion in red carotenoid pigmented lineages. Accordingly, expression of the gene in the liver is consistent with presence of C4-ketocarotenoids in the blood (Chapter 4), whilst integumentary expression is consistent with the absence of these carotenoids (Chapter 6). Notwithstanding the observed expression of *CYP2J19* in both the liver and feather follicles of the hybrid “red factor” canary, gene expression patterns appear able to differentiate between peripheral and central ketoconverters.

It is unclear why certain avian lineages use one mechanism over another for ketoconversion. One argument in favour of peripheral conversion concerns the higher risk of transport of the more costly ketocarotenoids in the bloodstream, where they are more likely to be subject to free-radical damage on route to the integuments (McGraw 2004). By contrast, the hypothesized trade-off of carotenoids in ornamentation and carotenoids in health benefiting systems requires that these molecules are in circulation in the bloodstream, where they can be diverted to function in other physiological contexts (del Val *et al.* 2009), and fits well with central conversion. On the other hand, a trade-off is also consistent with peripheral ketoconversion where circulating yellow precursor carotenoids may be diverted away from their roles in ornamentation to function in other cellular systems. An understanding of the phylogenetic distribution of peripheral versus central ketoconversion across different carotenoid pigmented lineages will be needed to further the understanding of the evolution of avian coloration.

References

Ali, M. A. and Klyne, M. A. (1985). Vision in vertebrates. New York, Plenum Press.

Almeida, D., Maldonado, E., Khan, I., Silva, L., Gilbert, M. T. P., Zhang, G. J., . . . Antunes, A. (2016). "Whole-genome identification, phylogeny, and evolution of the cytochrome P450 family 2 (CYP2) subfamilies in birds." Genome Biology and Evolution **8**(4): 1115-1131.

Amengual, J., Lobo, G. P., Golczak, M., Li, H. N. M., Klimova, T., Hoppel, C. L., . . . von Lintig, J. (2011). "A mitochondrial enzyme degrades carotenoids and protects against oxidative stress." Faseb Journal **25**(3): 948-959.

Anakk, S., Ku, C. Y., Vore, M. and Strobel, H. W. (2003). "Insights into gender bias: Rat cytochrome P450 3A9." Journal of Pharmacology and Experimental Therapeutics **305**(2): 703-709.

Andersson, S., Prager, M. and Johansson, E. I. A. (2007). "Carotenoid content and reflectance of yellow and red nuptial plumages in widowbirds (*Euplectes* spp.)." Functional Ecology **21**(2): 272-281.

Arnold, K. and Neumeyer, C. (1987). "Wavelength discrimination in the turtle *Pseudemys scripta elegans*." Vision Research **27**(9): 1501-1511.

Bailes, H. J., Robinson, S. R., Trezise, A. E. O. and Collin, S. P. (2006). "Morphology, characterization, and distribution of retinal photoreceptors in the Australian lungfish *Neoceratodus forsteri* (Krefft, 1870)." Journal of Comparative Neurology **494**(3): 381-397.

Bailey, S. F., Hinz, A. and Kassen, R. (2014). "Adaptive synonymous mutations in an experimentally evolved *Pseudomonas fluorescens* population." Nature Communications **5**.

Blount, J. D., Metcalfe, N. B., Birkhead, T. R. and Surai, P. F. (2003). "Carotenoid modulation of immune function and sexual attractiveness in zebra finches." Science **300**(5616): 125-127.

Bowmaker, J. K. (2008). "Evolution of vertebrate visual pigments." Vision Research **48**(20): 2022-2041.

Bowmaker, J. K., Heath, L. A., Wilkie, S. E. and Hunt, D. M. (1997). "Visual pigments and oil droplets from six classes of photoreceptor in the retinas of birds." Vision Research **37**(16): 2183-2194.

Bowmaker, J. K., Kovach, J. K., Whitmore, A. V. and Loew, E. R. (1993). "Visual pigments and oil droplets in genetically manipulated and carotenoid deprived quail - a microspectrophotometric study." Vision Research **33**(5-6): 571-578.

Bowmaker, J. K. and Martin, G. R. (1978). "Visual pigments and color-vision in a nocturnal bird, *Strix Aluco* (Tawny Owl)." Vision Research **18**(9): 1125-1130.

Britton, G. (1995). "Structure and properties of carotenoids in relation to function." FASEB Journal **9**(15): 1551-1558.

Britton, G., Liaaen-Jensen, S. and Pfander, H. (2004). Carotenoids handbook. Boston, Birkhäuser Verlag.

Burley, N. and Coopersmith, C. B. (1987). "Bill Color Preferences of Zebra Finches." Ethology **76**(2): 133-151.

Cao, Y., Sorenson, M. D., Kumazawa, Y., Mindell, D. P. and Hasegawa, M. (2000). "Phylogenetic position of turtles among amniotes: evidence from mitochondrial and nuclear genes." Gene **259**(1-2): 139-148.

Chiari, Y., Cahais, V., Galtier, N. and Delsuc, F. (2012). "Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria)." BMC Biology **10**(1): 65.

Clark, D. B. and Gibbons, J. W. (1969). "Dietary shift in turtle *Pseudemys Scripta* (Schoepff) from youth to maturity." Copeia (4): 704-706.

Costantini, D., Coluzza, C., Fanfani, A. and Dell'Omo, G. (2007). "Effects of carotenoid supplementation on colour expression, oxidative stress and body mass in rehabilitated captive adult kestrels (*Falco tinnunculus*)." Journal of Comparative Physiology B: Biochemical Systemic and Environmental Physiology **177**(7): 723-731.

Costantini, D. and Moller, A. P. (2008). "Carotenoids are minor antioxidants for birds." Functional Ecology **22**(2): 367-370.

Crawford, N. G., Faircloth, B. C., McCormack, J. E., Brumfield, R. T., Winker, K. and Glenn, T. C. (2012). "More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs." Biology Letters **8**(5): 783-786.

Crawford, N. G., Parham, J. F., Sellas, A. B., Faircloth, B. C., Glenn, T. C., Papenfuss, T. J., . . . Simison, W. B. (2015). "A phylogenomic analysis of turtles." Molecular Phylogenetics and Evolution **83**: 250-257.

Crews, D. and Bergeron, J. M. (1994). "Role of reductase and aromatase in sex determination in the red-eared slider (*Trachemys Scripta*), a turtle with temperature-dependent sex determination." Journal of Endocrinology **143**(2): 279-289.

Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., . . . Flicek, P. (2015). "Ensembl 2015." Nucleic Acids Research **43**(D1): D662-D669.

Dale, J. (2000). "Ornamental plumage does not signal male quality in red-billed queleas." Philosophical Transactions of the Royal Society B: Biological Sciences **267**(1458): 2143-2149.

Davies, N. B., Krebs, J. R. and West, S. A. (2012). An introduction to behavioural ecology, Wiley-Blackwell Publications.

De Silva, T. N., Peterson, A. T., Bates, J. M., Fernando, S. W. and Girard, M. G. (2017). "Phylogenetic relationships of weaverbirds (Aves: Ploceidae): A first robust phylogeny based on mitochondrial and nuclear markers." Molecular Phylogenetics and Evolution **109**: 21-32.

del Val, E., Senar, J. C., Garrido-Fernández, J., Jarén, M., Borràs, A., Cabrera, J. and Negro, J. J. (2009). "The liver but not the skin is the site for conversion of a red carotenoid in a passerine bird." Naturwissenschaften **96**(7): 797-801.

Edgar, R. C. (2004). "MUSCLE: Multiple sequence alignment with high accuracy and high throughput." Nucleic Acids Research **32**(5): 1792-1797.

Eggertsen, G., Olin, M., Andersson, U., Ishida, H., Kubota, S., Hellman, U., . . . Björkhem, I. (1996). "Molecular cloning and expression of rabbit sterol 12α-hydroxylase." Journal of Biological Chemistry **271**(50): 32269-32275.

Elbrecht, A. and Smith, R. G. (1992). "Aromatase enzyme-activity and sex determination in chickens." Science **255**(5043): 467-470.

Emerling, C. A. (2017). "Independent pseudogenization of CYP2J19 in penguins, owls and kiwis implicates gene in red carotenoid synthesis." bioRxiv.

Eriksson, J., Larson, G., Gunnarsson, U., Bed'hom, B., Tixier-Boichard, M., Stromstedt, L., . . . Andersson, L. (2008). "Identification of the Yellow skin gene reveals a hybrid origin of the domestic chicken." PLOS Genetics **4**(2).

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L. and Postlethwait, J. (1999). "Preservation of duplicate genes by complementary, degenerative mutations." Genetics **151**(4): 1531-1545.

Foth, C., Tischlinger, H. and Rauhut, O. W. M. (2014). "New specimen of *Archaeopteryx* provides insights into the evolution of pennaceous feathers." Nature **511**(7507): 79-82.

Fox, J. and Weisberg, S. (2011). An R companion to Applied Regression. United States of America, Thousand Oaks California: SAGE.

Friedman, N. R., McGraw, K. J. and Omland, K. E. (2014). "Evolution of carotenoid pigmentation in caciques and meadowlarks (Icteridae): repeated gains of red plumage coloration by carotenoid C4-oxygenation." Evolution **68**(3): 791-801.

Furukawa, T., Hurley, J. B. and Kawamura, S. (2014). Vertebrate photoreceptors : functional molecular bases. New York, Springer.

Gafvels, M., Olin, M., Chowdhary, B. P., Raudsepp, T., Andersson, U., Persson, B., . . . Eggertsen, G. (1999). "Structure and chromosomal assignment of the sterol 12 alpha-hydroxylase gene (CYP8B1) in human and mouse: Eukaryotic cytochrome P-450 gene devoid of introns." Genomics **56**(2): 184-196.

Garcia-de Blas, E., Mateo, R. and Alonso-Alvarez, C. (2015). "Accumulation of dietary carotenoids, retinoids and tocopherol in the internal tissues of a bird: a hypothesis for the cost of producing colored ornaments." Oecologia **177**(1): 259-271.

Gauthier, J. (1986). "Saurischian monophyly and the origin of birds." Memoirs of the California Academy of Sciences **8**: 55.

Ge, Z. Y., Johnson, J. D., Cobine, P. A., McGraw, K. J., Garcia, R. and Hill, G. E. (2015). "High concentrations of ketocarotenoids in hepatic mitochondria of *Haemorhous mexicanus*." Physiological and Biochemical Zoology **88**(4): 444-450.

Goldsmith, T. H., Collins, J. S. and Licht, S. (1984). "The cone oil droplets of avian retinas." Vision Research **24**(11): 1661-1671.

Gotoh, O. (1992). "Substrate recognition sites in cytochrome-P450 family-2 (CYP2) proteins inferred from comparative analyses of amino-acid and coding nucleotide sequences." Journal of Biological Chemistry **267**(1): 83-90.

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W. and Gascuel, O. (2010). "New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0." Systematic Biology **59**(3): 307-321.

Hadley, M. E. (1972). "Functional significance of vertebrate integumental pigmentation." American Zoologist **12**(1): 63-76.

Hart, N. S. (2001). "The visual ecology of avian photoreceptors." Progress in Retinal and Eye Research **20**(5): 675-703.

Hart, N. S., Bailes, H. J., Vorobyev, M., Marshall, N. J. and Collin, S. P. (2008). "Visual ecology of the Australian lungfish (*Neoceratodus forsteri*)." BMC Ecology **8**: 21.

Hart, N. S., Lisney, T. J. and Collin, S. P. (2006). "Cone photoreceptor oil droplet pigmentation is affected by ambient light intensity." Journal of Experimental Biology **209**(23): 4776-4787.

Hartley, R. C. and Kennedy, M. W. (2004). "Are carotenoids a red herring in sexual display?" Trends in Ecology & Evolution **19**(7): 353-354.

Hedges, S. B. and Poling, L. L. (1999). "A molecular phylogeny of reptiles." Science **283**(5404): 998-1001.

Hill, G. E. (1990). "Female house finches prefer colorful males - sexual selection for a condition-dependent trait." Animal Behaviour **40**: 563-572.

Hill, G. E. (1999). "Mate choice, male quality, and carotenoid-based plumage coloration." Proceedings of the International Ornithological Congress 22: 1654-1668.

Hill, G. E. (2000). "Energetic constraints on expression of carotenoid-based plumage coloration." Journal of Avian Biology **31**(4): 559-566.

Hill, G. E. and D., J. J. (2012). "The vitamin A–redox hypothesis: A biochemical basis for honest signaling via carotenoid pigmentation." The American Naturalist **180**(5): E127-E150.

Hill, G. E. and McGraw, K. J. (2006). Bird Coloration: vol I, Mechanisms and Measurements. Harvard University Press.

Hoyo, J. d., Elliott, A. and Christie, D. A. (2010). Handbook of the birds of the world. Barcelona, Lynx Edicions.

Hubbard, J. K., Uy, J. A. C., Hauber, M. E., Hoekstra, H. E. and Safran, R. J. (2010). "Vertebrate pigmentation: from underlying genes to adaptive function." Trends in Genetics **26**(5): 231-239.

Huggins, K. A., Navara, K. J., Mendonca, M. T. and Hill, G. E. (2010). "Detrimental effects of carotenoid pigments: the dark side of bright coloration." Naturwissenschaften **97**(7): 637-644.

Hughes, A. L. (1994). "The evolution of functionally novel proteins after gene duplication." Proceedings of the Royal Society B: Biological Sciences **256**(1346): 119-124.

Hughes, A. L. and Nei, M. (1988). "Pattern of nucleotide substitution at major histocompatibility complex class-I loci reveals overdominant selection." Nature **335**(6186): 167-170.

Husak, J. F. and Swallow, J. G. (2011). "Compensatory traits and the evolution of male ornaments." Behaviour **148**(1): 1-29.

Ibanez, A., Polo-Cavia, N., Lopez, P. and Martin, J. (2014). "Honest sexual signaling in turtles: experimental evidence of a trade-off between immune response and coloration in red-eared sliders *Trachemys scripta elegans*." Naturwissenschaften **101**(10): 803-811.

Innan, H. and Kondrashov, F. (2010). "The evolution of gene duplications: classifying and distinguishing between models." Nature Reviews Genetics **11**(2): 97-108.

Jaillon, O., Aury, J. M., Brunet, F., Petit, J. L., Stange-Thomann, N., Mauceli, E., . . . Crollius, H. R. (2004). "Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype." Nature **431**(7011): 946-957.

Jelinek, D. F., Andersson, S., Slaughter, C. A. and Russell, D. W. (1990). "Cloning and regulation of cholesterol 7-alpha-hydroxylase, the rate-limiting enzyme in bile-acid biosynthesis." Journal of Biological Chemistry **265**(14): 8190-8197.

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. and Mooers, A. O. (2012). "The global diversity of birds in space and time." Nature **491**(7424): 444-448.

Jiggins, C. D., Wallbank, R. W. R. and Hanly, J. J. (2017). "Waiting in the wings: what can we learn about gene co-option from the diversification of butterfly wing patterns?" Philosophical Transactions of the Royal Society B: Biological Sciences **372**(1713).

Johnson, J. D. and E., H. G. (2013). "Is carotenoid ornamentation linked to the inner mitochondria membrane potential? A hypothesis for the maintenance of signal honesty." Biochimie **95**(2): 436-444.

Kass, R. E. and Raftery, A. E. (1995). "Bayes factors." Journal of the American Statistical Association **90**(430): 773-795.

Kiefer, C., Sumser, E., Wernet, M. F. and von Lintig, J. (2002). "A class B scavenger receptor mediates the cellular uptake of carotenoids in *Drosophila*." Proceedings of the National Academy of Sciences of the United States of America **99**(16): 10581-10586.

Kinoshita, S. and Yoshioka, S. (2005). "Structural colors in nature: The role of regularity and irregularity in the structure." ChemPhysChem **6**(8): 1442-1459.

Knott, B., Berg, M. L., Morgan, E. R., Buchanan, K. L., Bowmaker, J. K. and Bennett, A. T. D. (2010). "Avian retinal oil droplets: dietary manipulation of colour vision?" Proceedings of the Royal Society B: Biological Sciences **277**(1683): 953-962.

Kondrashov, F. A. (2012). "Gene duplication as a mechanism of genomic adaptation to a changing environment." Proceedings of the Royal Society B: Biological Sciences **279**(1749): 5048-5057.

Koschowitz, M. C., Fischer, C. and Sander, M. (2014). "Beyond the rainbow." Science **346**(6208): 416-418.

Krinsky, N. I. (1993). "Actions of carotenoids in biological systems." Annual Review of Nutrition **13**: 561-587.

Kudla, G., Murray, A. W., Tollervey, D. and Plotkin, J. B. (2009). "Coding-sequence determinants of gene expression in *Escherichia coli*." Science **324**(5924): 255-258.

LaFountain, A. M., Frank, H. A. and Prum, R. O. (2013). "Carotenoids from the crimson and maroon plumages of Old World orioles (Oriolidae)." Archives of Biochemistry and Biophysics **539**(2): 126-132.

LaFountain, A. M., Kaligotla, S., Cawley, S., Riedl, K. M., Schwartz, S. J., Frank, H. A. and Prum, R. O. (2010). "Novel methoxy-carotenoids from the burgundy-colored plumage of the Pompadour Cotinga *Xipholena punicea*." Archives of Biochemistry and Biophysics **504**(1): 142-153.

LaFountain, A. M., Prum, R. O. and Frank, H. A. (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**: 201-212.

Lau, E. S. W., Zhang, Z. W., Qin, M. M. and Ge, W. (2016). "Knockout of zebrafish ovarian aromatase gene (cyp19a1a) by TALEN and CRISPR/Cas9 leads to all-male offspring due to failed ovarian differentiation." Scientific Reports **6**.

Lee, M. S. Y. (2001). "Molecules, morphology, and the monophyly of diapsid reptiles." Contributions to Zoology **70**(1): 1-22.

Li, Q., Gao, K.-Q., Meng, Q., Clarke, J. A., Shawkey, M. D., D'Alba, L., . . . Vinther, J. (2012). "Reconstruction of microraptor and the evolution of iridescent plumage." Science **335**(6073): 1215-1219.

Li, Q. G., Gao, K. Q., Vinther, J., Shawkey, M. D., Clarke, J. A., D'Alba, L., . . . Prum, R. O. (2010). "Plumage color patterns of an extinct dinosaur." Science **327**(5971): 1369-1372.

Liebman, P. A. and Granda, A. M. (1975). "Super dense carotenoid spectra resolved in single cone oil droplets." Nature **253**(5490): 370-372.

Ligon, R. A., Simpson, R. K., Mason, N. A., Hill, G. E. and McGraw, K. J. (2016). "Evolutionary innovation and diversification of carotenoid-based pigmentation in finches." Evolution **70**(12): 2839-2852.

Lipetz, L. E. (1984). "A new method for determining peak absorbance of dense pigment samples and its application to the cone oil droplets of *Emydoidea blandingii*." Vision Research **24**(6): 597-604.

Liu, H. L., Zheng, H. P., Zhang, H. K., Deng, L. H., Liu, W. H., Wang, S. Q., . . . Zhang, G. F. (2015). "A de novo transcriptome of the noble scallop, *Chlamys nobilis*, focusing on mining transcripts for carotenoid-based coloration." BMC Genomics **16**.

Lopes, R. J., Johnson, J. D., Toomey, M. B., Ferreira, M. S., Araujo, P. M., Melo-Ferreira, J., . . . Carneiro, M. (2016). "Genetic basis for red coloration in birds." Current Biology **26**(11): 1427-1434.

Lozano, G. A. (1994). "Carotenoids, parasites, and sexual selection." Oikos **70**(2): 309-311.

Mahnke, Z., Strotkamp, D., Roos, P. H., Hanstein, W. G., Chabot, G. G. and Nef, P. (1997). "Expression and inducibility of cytochrome P450 3A9 (CYP3A9) and other members of the CYP3A subfamily in rat liver." Archives of Biochemistry and Biophysics **337**(1): 62-68.

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

Matsuno, T. (2001). "Aquatic animal carotenoids." Fisheries Science **67**(5): 771-783.

Mayne, S. T. and Parker, R. S. (1986). "Subcellular-distribution of dietary-beta-carotene in chick liver." Lipids **21**(2): 164-169.

McGraw, K. J. (2004). "Colorful songbirds metabolize carotenoids at the integument." Journal of Avian Biology **35**(6): 471-476.

McGraw, K. J. and Schuetz, J. G. (2004). "The evolution of carotenoid coloration in estrildid finches: a biochemical analysis." Comparative Biochemistry and Physiology B: Biochemistry & Molecular Biology **139**(1): 45-51.

McGraw, K. J. and Toomey, M. B. (2010). "Carotenoid accumulation in the tissues of zebra finches: predictors of integumentary pigmentation and implications for carotenoid allocation strategies." Physiological and Biochemical Zoology **83**(1): 97-109.

McGraw, K. J., Wakamatsu, K., Clark, A. B. and Yasukawa, K. (2004). "Red-winged blackbirds *Agelaius phoeniceus* use carotenoid and melanin pigments to color their epaulets." Journal of Avian Biology **35**(6): 543-550.

Mein, J. R., Dolnikowski, G. G., Ernst, H., Russell, R. M. and Wang, X. D. (2011). "Enzymatic formation of apo-carotenoids from the xanthophyll carotenoids lutein, zeaxanthin and beta-cryptoxanthin by ferret carotene-9 ',10 '-monooxygenase." Archives of Biochemistry and Biophysics **506**(1): 109-121.

Mendes-Pinto, M. M., LaFountain, A. M., Stoddard, M. C., Prum, R. O., Frank, H. A. and Robert, B. (2012). "Variation in carotenoid-protein interaction in bird feathers produces novel plumage coloration." Journal of the Royal Society Interface **9**(77): 3338-3350.

Messler, W. and Stewart, C. B. (1997). "Episodic adaptive evolution of primate lysozymes." Nature **385**(6612): 151-154.

Miki, W. (1991). "Biological functions and activities of animal carotenoids." Pure and Applied Chemistry **63**(1): 141-146.

Moore, T. (1930). "Vitamin A and carotene. The absence of the liver oil vitamin A from carotene. VI. The conversion of carotene to vitamin A in vivo." Biochemical Journal **24**(3): 692-702.

Mundy, N. I. (2005). "A window on the genetics of evolution: MC1R and plumage colouration in birds." Proceedings of the Royal Society of London Series B: Biological Sciences **272**(1573): 1633-1640.

Mundy, N. I., Stapley, J., Bennison, C., Tucker, R., Twyman, H., Kim, K. W., . . . Slate, J. (2016). "Red carotenoid coloration in the zebra finch is controlled by a cytochrome P450 gene cluster." Current Biology **26**(11): 1435-1440.

Murphy, T. G., Rosenthal, M. F., Montgomerie, R. and Tarvin, K. A. (2009). "Female American goldfinches use carotenoid-based bill coloration to signal status." Behavioral Ecology **20**(6): 1348-1355.

Nadeau, N. J., Burke, T. and Mundy, N. I. (2007). "Evolution of an avian pigmentation gene correlates with a measure of sexual selection." Proceedings of the Royal Society of London Series B: Biological Sciences **274**(1620): 1807-1813.

Naftolin, F., Horvath, T. L., Jakab, R. L., Leranth, C., Harada, N. and Balthazart, J. (1996). "Aromatase immunoreactivity in axon terminals of the vertebrate brain - An immunocytochemical study on quail, rat, monkey and human tissues." Neuroendocrinology **63**(2): 149-155.

Nagaoka, S. M., Martins, A. S., dos Santos, R. G., Tognella, M. M. P., de Oliveira, E. C. and Seminoff, J. A. (2012). "Diet of juvenile green turtles (*Chelonia mydas*) associating with artisanal fishing traps in a subtropical estuary in Brazil." Marine Biology **159**(3): 573-581.

Nakano, M., Kelly, E. J. and Rettie, A. E. (2009). "Expression and characterization of CYP4V2 as a fatty acid omega-hydroxylase." Drug Metabolism and Disposition **37**(11): 2119-2122.

Ninnes, C. E., Webb, S. L. and Andersson, S. (2017). "Are red bishops red enough? On the persistence of a generalized receiver bias in *Euplectes*." Behavioral Ecology **28**(1): 117-122.

Ogishima, T., Deguchi, S. and Okuda, K. (1987). "Purification and characterization of cholesterol 7-alpha-hydroxylase from rat-liver microsomes." Journal of Biological Chemistry **262**(16): 7646-7650.

Ohno, S. (1970). Evolution by gene duplication. New York, Springer-Verlag.

Oliphant, L. W. (1987). "Pteridines and purines as major pigments of the avian iris." Pigment Cell Research **1**(2): 129-131.

Olson, J. A. and Hayaishi, O. (1965). "Enzymatic cleavage of beta-carotene into vitamin a by soluble enzymes of rat liver and intestine." Proceedings of the National Academy of Sciences of the United States of America **54**(5): 1364-1370.

Olson, V. A. and Owens, I. P. F. (1998). "Costly sexual signals: are carotenoids rare, risky or required?" Trends in Ecology & Evolution **13**(12): 510-514.

Orr, H. A. (2005). "The genetic theory of adaptation: A brief history." Nature Reviews Genetics **6**(2): 119-127.

Pagel, M. (1994). "Detecting correlated evolution on phylogenies - a general-method for the comparative-analysis of discrete characters." Proceedings of the Royal Society of London Series B: Biological Sciences **255**(1342): 37-45.

Pagel, M., Meade, A. and Barker, D. (2004). "Bayesian estimation of ancestral character states on phylogenies." Systematic Biology **53**(5): 673-684.

Pérez-Rodríguez, L., de Blas, E. G.-., Martínez-Padilla, J., Mougeot, F. and Mateo, R. (2016). "Carotenoid profile and vitamins in the combs of the red grouse (*Lagopus*

lagopus scoticus): implications for the honesty of a sexual signal." Journal of Ornithology **157**(1): 145-153.

Peterson, R. S., Yarram, L., Schlinger, B. A. and Saldanha, C. J. (2005). "Aromatase is pre-synaptic and sexually dimorphic in the adult zebra finch brain." Proceedings of the Royal Society B: Biological Sciences **272**(1576): 2089-2096.

Pfaffl, M. W. (2001). "A new mathematical model for relative quantification in real-time RT-PCR." Nucleic Acids Research **29**(9): e45.

Prager, M. and Andersson, S. (2010). "Convergent evolution of red carotenoid coloration in widowbirds and bishops (*Euplectes* spp.)." Evolution **64**(12): 3609-3619.

Prager, M., Johansson, E. I. A. and Andersson, S. (2009). "Differential ability of carotenoid C4-oxygenation in yellow and red bishop species (*Euplectes* spp.)." Comparative Biochemistry and Physiology B: Biochemistry & Molecular Biology **154**(4): 373-380.

Prum, R. O., LaFountain, A. M., Berg, C. J., Tauber, M. J. and Frank, H. A. (2014). "Mechanism of carotenoid coloration in the brightly colored plumages of broadbills (Eurylaimidae)." Journal of Comparative Physiology B: Biochemical Systemic and Environmental Physiology **184**(5): 651-672.

Pryke, S. R. and Andersson, S. (2003). "Carotenoid-based status signalling in red-shouldered widowbirds (*Euplectes axillaris*): epaulet size and redness affect captive and territorial competition." Behavioral Ecology and Sociobiology **53**(6): 393-401.

Pryke, S. R., Andersson, S. and Lawes, M. J. (2001). "Sexual selection of multiple handicaps in the red-collared widowbird: Female choice of tail length but not carotenoid display." Evolution **55**(7): 1452-1463.

Pryke, S. R., Andersson, S., Lawes, M. J. and Piper, S. E. (2002). "Carotenoid status signaling in captive and wild red-collared widowbirds: independent effects of badge size and color." Behavioral Ecology **13**(5): 622-631.

Putnam, N. H., Butts, T., Ferrier, D. E. K., Furlong, R. F., Hellsten, U., Kawashima, T., . . . Rokhsar, D. S. (2008). "The amphioxus genome and the evolution of the chordate karyotype." Nature **453**(7198): 1064-1063.

R Core Team (2009). R: A language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing.

R Core Team (2016). R: A language and environment for statistical computing. Vienna, Austria, R Foundation for Statistical Computing.

Reed, R. D., Papa, R., Martin, A., Hines, H. M., Counterman, B. A., Pardo-Diaz, C., . . . McMillan, W. O. (2011). "*Optix* drives the repeated convergent evolution of butterfly wing pattern mimicry." Science **333**(6046): 1137-1141.

Robertson, G. R., Farrell, G. C. and Liddle, C. (1998). "Sexually dimorphic expression of rat *CYP3A9* and *CYP3A18* genes is regulated by growth hormone." Biochemical and Biophysical Research Communications **242**(1): 57-60.

Rose, K. A., Stapleton, G., Dott, K., Kieny, M. P., Best, R., Schwarz, M., . . . Lathe, R. (1997). "CYP7B, a novel brain cytochrome P450, catalyzes the synthesis of neurosteroids 7 alpha-hydroxy dehydroepiandrosterone and 7 alpha-hydroxy pregnenolone."

Proceedings of the National Academy of Sciences of the United States of America **94**(10): 4925-4930.

Rosenblum, E. B., Römpler, H., Schöneberg, T. and Hoekstra, H. E. (2010). "Molecular and functional basis of phenotypic convergence in white lizards at White Sands." Proceedings of the National Academy of Sciences of the United States of America **107**(5): 2113-2117.

Sakudoh, T., Kuwazaki, S., Iizuka, T., Narukawa, J., Yamamoto, K., Uchino, K., . . . Tsuchida, K. (2013). "CD36 homolog divergence is responsible for the selectivity of carotenoid species migration to the silk gland of the silkworm *Bombyx mori*." Journal of Lipid Research **54**(2): 482-495.

Schaefer, H. M., Schmidt, V. and Bairlein, F. (2003). "Discrimination abilities for nutrients: which difference matters for choosy birds and why?" Animal Behaviour **65**: 531-541.

Schott, R. K., Muller, J., Yang, C. G. Y., Bhattacharyya, N., Chan, N., Xu, M. S., . . . Chang, B. S. W. (2016). "Evolutionary transformation of rod photoreceptors in the all-cone retina of a diurnal garter snake." Proceedings of the National Academy of Sciences of the United States of America **113**(2): 356-361.

Schwarz, M., Lund, E. G., Lathe, R., Björkhem, I. and Russell, D. W. (1997). "Identification and characterization of a mouse oxysterol 7α-hydroxylase cDNA." Journal of Biological Chemistry **272**(38): 23995-24001.

Seimiya, M. and Gehring, W. J. (2000). "The *Drosophila* homeobox gene *optix* is capable of inducing ectopic eyes by an eyeless-independent mechanism." Development **127**(9): 1879-1886.

Senar, J. C., Moller, A. P., Ruiz, I., Negro, J. J., Broggi, J. and Hohtola, E. (2010). "Specific appetite for carotenoids in a colorful bird." PLoS ONE **5**(5).

Shabalina, S. A., Spiridonov, N. A. and Kashina, A. (2013). "Sounds of silence: synonymous nucleotides as a key to biological regulation and complexity." Nucleic Acids Research **41**(4): 2073-2094.

Shaffer, H. B., Minx, P., Warren, D. E., Shedlock, A. M., Thomson, R. C., Valenzuela, N., . . . Wilson, R. K. (2013). "The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage." Genome Biology **14**(3).

Shen, X. X., Liang, D., Wen, J. Z. and Zhang, P. (2011). "Multiple genome alignments facilitate development of NPCL markers: A case study of tetrapod phylogeny focusing on the position of turtles." Molecular Biology and Evolution **28**(12): 3237-3252.

Sillman, A. J., Ronan, S. J. and Loew, E. R. (1991). "Histology and microspectrophotometry of the photoreceptors of a crocodilian, alligator-mississippiensis." Proceedings of the Royal Society B: Biological Sciences **243**(1306): 93-98.

Simons, M. J. P., Maia, R., Leenknegt, B. and Verhulst, S. (2014). "Carotenoid-dependent signals and the evolution of plasma carotenoid levels in birds." American Naturalist **184**(6): 741-751.

Stavenga, D. G. and Wilts, B. D. (2014). "Oil droplets of bird eyes: microlenses acting as spectral filters." Philosophical Transactions of the Royal Society B: Biological Sciences **369**(1636).

Steffen, J. E., Learn, K. M., Drumheller, J. S., Boback, S. M. and McGraw, K. J. (2015). "Carotenoid composition of colorful body stripes and patches in the painted turtle (*Chrysemys picta*) and red-eared slider (*Trachemys scripta*)." Chelonian Conservation and Biology **14**(1): 56-63.

Stern, D. L. (2013). "The genetic causes of convergent evolution." Nature Reviews Genetics **14**(11): 751-764.

Stern, D. L. and Orgogozo, V. (2008). "The loci of evolution: How predictable is genetic evolution ?" Evolution **62**(9): 2155-2177.

Storz, J. F. (2016). "Causes of molecular convergence and parallelism in protein evolution." Nature Reviews Genetics **17**(4): 239-250.

Stradi, R., Celentano, G., Rossi, E., Rovati, G. and Pastore, M. (1995). "Carotenoids in bird plumage .1. The carotenoid pattern in a series of palearctic carduelinae." Comparative Biochemistry and Physiology B: Biochemistry & Molecular Biology **110**(1): 131-143.

Sundvold, H., Helgeland, H., Baranski, M., Omholt, S. W. and Vage, D. I. (2011). "Characterisation of a novel paralog of scavenger receptor class B member I (*SCARB1*) in Atlantic salmon (*Salmo salar*)." BMC Genetics **12**.

Suryamohan, K. and Halfon, M. S. (2015). "Identifying transcriptional cis-regulatory modules in animal genomes." Wiley Interdiscipinary Reviews. Developemental Biology **4**(2): 59-84.

Svensson, P. A. and Wong, B. B. M. (2011). "Carotenoid-based signals in behavioural ecology: a review." Behaviour **148**(2): 131-189.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013). "MEGA6: Molecular evolutionary genetics analysis version 6.0." Molecular Biology and Evolution **30**(12): 2725-2729.

Thomas, D. B., McGraw, K. J., Butler, M. W., Carrano, M. T., Madden, O. and James, H. F. (2014). "Ancient origins and multiple appearances of carotenoid-pigmented feathers in birds." Proceedings of the Royal Society B: Biological Sciences **281**(1788).

Tian, R., Pitchford, W. S., Morris, C. A., Cullen, N. G. and Bottema, C. D. K. (2010). "Genetic variation in the beta, beta-carotene-9 ', 10 '-dioxygenase gene and association with fat colour in bovine adipose tissue and milk." Animal Genetics **41**(3): 253-259.

Tillotson, H. M. and Oliphant, L. W. (1990). "Iris stromal pigment-cells of the ringed turtle dove." Pigment Cell Research **3**(6): 319-323.

Toews, D. P., Hofmeister, N. R. and Taylor, S. A. (2017). "The evolution and genetics of carotenoid processing in animals." Trends in Genetics **33**(3): 171-182.

Toomey, M. B., Collins, A. M., Frederiksen, R., Cornwall, M. C., Timlin, J. A. and Corbo, J. C. (2015). "A complex carotenoid palette tunes avian colour vision." Journal of the Royal Society Interface **12**(111).

Toomey, M. B., Lind, O., Frederiksen, R., Curley, R. W., Riedl, K. M., Wilby, D., . . . Corbo, J. C. (2016). "Complementary shifts in photoreceptor spectral tuning unlock the full adaptive potential of ultraviolet vision in birds." Elife **5**.

Toomey, M. B., Lopes, R. J., Araújo, P. M., Johnson, J. D., Gazda, M. A., Afonso, S., . . . Carneiro, M. (2017). "High-density lipoprotein receptor SCARB1 is required for carotenoid coloration in birds." Proceedings of the National Academy of Sciences **114**(20): 5219-5224.

Toomey, M. B. and McGraw, K. J. (2010). "The effects of dietary carotenoid intake on carotenoid accumulation in the retina of a wild bird, the house finch (*Carpodacus mexicanus*)." Archives of Biochemistry and Biophysics **504**(1): 161-168.

Twyman, H., Valenzuela, N., Literman, R., Andersson, S. and Mundy, N. I. (2016). "Seeing red to being red: conserved genetic mechanism for red cone oil droplets and co-option for red coloration in birds and turtles." Proceedings of the Royal Society B: Biological Sciences **283**(1836).

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M. and Rozen, S. G. (2012). "Primer3-new capabilities and interfaces." Nucleic Acids Research **40**(15).

Uy, J. A. C., Cooper, E. A., Cutie, S., Concannon, M. R., Poelstra, J. W., Moyle, R. G. and Filardi, C. E. (2016). "Mutations in different pigmentation genes are associated with parallel melanism in island flycatchers." Proceedings of the Royal Society B: Biological Sciences **283**(1834).

Vage, D. I. and Boman, I. A. (2010). "A nonsense mutation in the beta-carotene oxygenase 2 (*BCO2*) gene is tightly associated with accumulation of carotenoids in adipose tissue in sheep (Ovis aries)." BMC Genetics **11**.

Valenzuela, N. (2009). "Egg incubation and collection of painted turtle embryos." Cold Spring Harbor Protocols **2009**(7): pdb prot5238.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. and Speleman, F. (2002). "Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes." Genome Biology **3**(7): research0034.0031-research0034.0011.

Vershinin, A. (1999). "Biological functions of carotenoids - diversity and evolution." BioFactors **10**(2-3): 99-104.

Vorobyev, M. (2003). "Coloured oil droplets enhance colour discrimination." Proceedings of the Royal Society B: Biological Sciences **270**(1521): 1255-1261.

Vorobyev, M., Osorio, D., Bennett, A. T. D., Marshall, N. J. and Cuthill, I. C. (1998). "Tetrachromacy, oil droplets and bird plumage colours." Journal of Comparative Physiology A **183**(5): 621-633.

Walls, G. L. (1942). The vertebrate eye and its adaptive radiation, Cranbrook Institute of Science, Bloomfield Hills, Michigan.

Walls, G. L. and Judd, H. D. (1933). "The intra-ocular colour-filters of vertebrates." British Journal of Ophthalmology **17**(11): 641-675.

Walsh, N., Dale, J., McGraw, K. J., Pointer, M. A. and Mundy, N. I. (2012). "Candidate genes for carotenoid coloration in vertebrates and their expression profiles in the carotenoid-containing plumage and bill of a wild bird." Proceedings of the Royal Society B: Biological Sciences **279**(1726): 58-66.

Wang, H. M. and Strobel, H. W. (1997). "Regulation of *CYP3A9* gene expression by estrogen and catalytic studies using cytochrome P450 3A9 expressed in *Escherichia coli*." Archives of Biochemistry and Biophysics **344**(2): 365-372.

Wang, Z., Pascual-Anaya, J., Zadissa, A., Li, W., Niimura, Y., Huang, Z., . . . Irie, N. (2013). "The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan." Nature Genetics **45**(6): 701-706.

Watanabe, K. P., Kawai, Y. K., Ikenaka, Y., Kawata, M., Ikushiro, S.-I., Sakaki, T. and Ishizuka, M. (2013). "Avian cytochrome P450 (CYP) 1-3 family genes: Isoforms, evolutionary relationships, and mRNA expression in chicken liver." PLoS ONE **8**(9): e75689.

Werck-Reichhart, D. and Feyereisen, R. (2000). "Cytochromes P450: a success story." Genome Biology **1**(6): reviews3003.3001.

Wittkopp, P. J. and Kalay, G. (2012). "Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence." Nature Reviews Genetics **13**(1): 59-69.

Wood, T. E., Burke, J. M. and Rieseberg, L. H. (2005). "Parallel genotypic adaptation: when evolution repeats itself." Genetica **123**(1-2): 157-170.

Wray, G. A. (2007). "The evolutionary significance of cis-regulatory mutations." Nature Reviews Genetics **8**(3): 206-216.

Wright, M. W. and Bowmaker, J. K. (2001). "Retinal photoreceptors of paleognathous birds: the ostrich (*Struthio camelus*) and rhea (*Rhea americana*)." Vision Research **41**(1): 1-12.

Yang, Z. (1997). "PAML: a program package for phylogenetic analysis by maximum likelihood." Computer Applications in the Biosciences **13**(5): 555-556.

Yang, Z. (1998). "Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution." Molecular Biology and Evolution **15**(5): 568-573.

Yang, Z. (2007). "PAML 4: Phylogenetic analysis by maximum likelihood." Molecular Biology and Evolution **24**(8): 1586-1591.

Yang, Z. and Bielawski, J. P. (2000). "Statistical methods for detecting molecular adaptation." Trends in Ecology & Evolution **15**(12): 496-503.

Zana, Y., Ventura, D. F., de Souza, J. M. and DeVoe, R. D. (2001). "Tetrachromatic input to turtle horizontal cells." Visual Neuroscience **18**(5): 759-765.

Zhang, G., Hubalewska, M. and Ignatova, Z. (2009). "Transient ribosomal attenuation coordinates protein synthesis and co-translational folding." Nature Structural & Molecular Biology **16**(3): 274-280.

Zhang, J. Z., Nielsen, R. and Yang, Z. H. (2005). "Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level." Molecular Biology and Evolution **22**(12): 2472-2479.

Zhang, L. X., Cooney, R. V. and Bertram, J. S. (1992). "Carotenoids up-regulate connexin 43 gene-expression independent of their provitamin-A or antioxidant properties." Cancer Research **52**(20): 5707-5712.