

# CYP2J19 mediates carotenoid colour introgression across a natural avian hybrid zone

Alexander Kirschel<sup>1</sup>, Emmanuel Nwankwo<sup>1</sup>, Daniel Pierce<sup>2</sup>, Michaela Moysi<sup>1</sup>, Bridget Ogolowa<sup>1</sup>, Ara Monadjem<sup>3</sup>, and Alan Brelsford<sup>2</sup>

<sup>1</sup>University of Cyprus

<sup>2</sup>University of California Riverside

<sup>3</sup>University of Eswatini

May 12, 2020

## Abstract

It has long been of interest to identify the phenotypic traits that mediate reproductive isolation between related species, and more recently, the genes that underpin them. Much work has focused on identifying genes associated with animal colour, with the candidate gene *CYP2J19* identified in laboratory studies as the ketolase converting yellow dietary carotenoids to red ketocarotenoids in birds with red pigments. But evidence that *CYP2J19* explains variation between red and yellow feather coloration in wild populations of birds is lacking. Hybrid zones between related species provide the opportunity to identify genes associated with specific traits. Here we investigate genomic regions associated with forecrown colour in red-fronted and yellow-fronted tinkerbirds across a hybrid zone in southern Africa. We sampled 79 individuals, measuring spectral reflectance of forecrown feathers as well as scoring colours from photographs. We performed a genome-wide association study to identify associations with carotenoid-based coloration, using double-digest RAD sequencing aligned to a short-read whole genome of a *Pogoniulus* tinkerbird that we assembled. Admixture mapping using 104,933 SNPs identified a region of chromosome 8 that includes *CYP2J19* as the only locus with more than two SNPs significantly associated with both crown hue and crown score. The hybrid zone was bimodal with asymmetric backcrossing, consistent with the hypothesis that yellow-fronted females mate more often with red-fronted males than vice versa. Female red-fronted tinkerbirds mating assortatively with red-crowned males may reinforce species divergence and is consistent with the hypothesis that converted carotenoids are an honest signal of quality.



## INTRODUCTION

For centuries, species have been delineated based on their physical appearance. Early naturalists differentiated between species based on functional traits such as jaw and limb shape, and beak size, or otherwise on traits that function in social selection, such as fur pattern and plumage coloration (Birkhead, 2008; Leroi, 2014). Such phenotypic characters, representing clear visual differences to the human eye, have been thought to aid the species themselves in recognising their own kind and finding potential mates. It has long been of interest to identify which traits mediate reproductive isolation between related species (Mayr, 1942), and more recently, which genes underpin those traits; in other words, the genes that mediate speciation (Nosil & Schluter, 2011). Hybrid zones between related species provide an unparalleled opportunity to identify genes associated with phenotypic characters (Delmore et al., 2016; Brelsford et al., 2017), including those that play an important role in mate choice and species recognition (Harrison & Larson, 2016). In some such cases, it is just a small set of genes that maintain phenotypic differences between hybridising taxa, especially those functioning in coloration and pattern differences (Poelstra et al., 2014; Toews et al., 2016). And evidence is starting to mount that the extent of divergence in plumage coloration between populations affects assortative mating (Haas et al., 2009; Scordato et al., 2017; Billerman et al., 2019).

Genes involved in coloration and patterning pathways have received a lot of attention (Mundy, 2005; Hoekstra, 2006), with a number of recent studies identifying genes associated with the melanin pigmentation pathway as a result of research across hybrid zones (e.g., (Delmore et al., 2016; Toews et al., 2016; Walsh et al., 2016; Campagna et al., 2017). Carotenoid-based coloration, incorporating the reds, oranges and yellows, has received far less attention (Toews et al., 2017), because carotenoid uptake is dependent on diet, meaning that identifying the genes involved in carotenoid processing is especially challenging (Campagna et al., 2017). Nevertheless, much progress has been made recently with the scavenger receptor B1 (*SCARB1*) identified as a mediator of carotenoid-colour expression in birds (Toomey et al., 2017), while a region of the Z chromosome has been identified as a potential regulatory region for *follistatin* and is associated with sex-linked red colour polymorphism in Gouldian finch *Erithrura gouldiae* (Kim et al., 2019). Previous work has also identified candidate genes to explain differences between red and yellow feather coloration through lab crosses of yellow canaries with red siskins (Lopes et al., 2016) and red and orange billed individuals of zebra finches (Mundy et al., 2016). Both studies specifically identified the gene *CYP2J19* as the one that encodes the ketolases that catalyze the conversion of yellow dietary carotenoids to red C4-ketocarotenoids. Both studies were focused on laboratory-based crosses involving artificially selected forms of species developed to express certain traits (“red factor” canaries and yellowbeak zebra finches respectively). A field-based study in long-tailed finch *Poephila acuticauda* provided further support for the role of *CYP2J19* in carotenoid-based bill colour across

a natural hybrid zone (Hooper et al., 2019), while *CYP2J19* gene expression has also been associated with red coloration in weaverbirds (Twyman et al., 2018b). But evidence that it explains variation between red and yellow feather coloration in wild populations of birds is still equivocal.

Red-fronted (*Pogoniulus pusillus*, Dumont, 1816) and yellow-fronted (*P. chrysoconus*, Temminck, 1832) tinkerbirds provide such an opportunity to investigate the genes involved in carotenoid processing because they hybridise extensively at a contact zone in southern Africa (Nwankwo et al., 2019). They differ phenotypically most noticeably in their forecrown coloration; hence their vernacular names (Fig. 1). Their other plumage differences are also in carotenoid-based colours, but those are more subtle differences, to the human eye at least, between shades of yellow and white that characterize the colour of their underparts, rump, wing bar, and supercilium (Hockey et al., 2005). The two species have been shown to hybridise extensively with asymmetric introgression of red forecrown plumage into the *P. chrysoconus* genomic background, in spite of phylogenetic analysis on mitochondrial DNA revealing the two species have been diverging for over four million years (Nwankwo et al., 2019). Despite the deep divergence between the species, the extent of introgression indicates that hybrids are viable and fertile, though some reduced hybrid fitness might maintain a tension zone and inhibit the two species collapsing into one (e.g., Kearns et al., 2018). Thus, if at least one species continues to differentiate the two phenotypes and mates assortatively, the plumage trait and the genes underlying it may play an important role in maintaining species divergence.

Here we investigate the genomic regions associated with the red and yellow feathers in red-fronted and yellow-fronted tinkerbirds that could play an important role in reproductive isolation. We sampled 79 individuals across the hybrid zone in southern Africa, collecting forecrown feathers for spectrophotometric analysis (Nwankwo et al., 2019). We performed a genome-wide association study (GWAS) to identify single nucleotide polymorphisms (SNPs) associated with plumage coloration. The study included assembly of a short-read genome of an individual *Pogoniulus pusillus* and double-digest restriction association DNA sequencing (ddRAD). Our aim was to identify the genes associated with carotenoid plumage coloration across a natural contact zone, and specifically which regions of the genome explain differences between red and yellow plumage. We further tested whether those genes, or SNPs at other loci, might be associated with the shade of yellow to white or extent of yellow versus black plumage in other feather patches, while also testing for any differences in plumage attributable to sex. We then compared alleles at loci associated with forecrown colour with the hybrid index across autosomes and the Z chromosome, and interspecific heterozygosity of individuals, to elucidate the patterns of interbreeding across genotypes that explain asymmetric introgression between species that have been diverging for over 4 million years (Nwankwo et al., 2019).

## MATERIALS AND METHODS

### *Study species and sample collection*

Tinkerbirds are African barbets (Lybiidae) that while omnivorous, feed primarily on fruit, and have a particular preference for mistletoe (Godschalk, 1985; Dowsett-Lemaire, 1988). Species in the genus differ subtly in plumage characters in spite of millions of years of divergence in mitochondrial DNA (mtDNA) between currently recognised sister species (Kirschel et al., 2009; Kirschel et al., 2018; Nwankwo et al., 2018; Nwankwo et al., 2019). *P. pusillus pusillus* (hereafter *pusillus*) and *P. chrysoconus extoni* (Layard, 1871) (hereafter *extoni*) are the subspecies of red-fronted and yellow-fronted tinkerbird, respectively, that meet at a contact zone in southern Africa (Monadjem et al., 1994). They differ in plumage characters beyond the forecrown colour in the yellower supercilium and dorsal streaking in *pusillus* – whiter in *extoni*, the deeper amber and greater percentage of carotenoid pigment over melanin pigment on the wing coverts in *pusillus* – paler yellow and more black in *extoni* (Fig. 1). The two species might also differ more subtly in underpart coloration, such as in throat and breast colour (Hockey et al., 2005).

Fieldwork was performed at 16 sites across the contact zone where subspecies *extoni* and *pusillus* meet in southern Africa. Fieldwork took place in Tanzania June – September 2013, and in South Africa and Eswatini in March and December 2015, and January – February 2017. Methods used to obtain DNA and feather samples have been previously described (Nwankwo et al., 2019). Briefly, 79 birds were caught in

mist nets and ringed, of which 37 were from within the contact zone. Blood samples were obtained via venepuncture of the brachial vein, and crown feathers were plucked and placed in envelopes for spectral reflectance analysis in the lab. Photographs of 78 birds were taken from several angles using either a DSLR Camera or smartphone for subsequent plumage scoring. Fieldwork was performed with ringing permits obtained from SAFRING, Eswatini's Big Game Parks, Mpumalanga Tourism and Parks Agency, the Game and Fauna Service of Cyprus, and with research and collecting permits from the respective authorities (see Acknowledgements).

### *Plumage scoring & spectral reflectance*

Methods used to obtain reflectance spectra from forecrown feathers using a spectrophotometer have been described in detail (Nwankwo et al., 2019). In brief, we used a JAZ spectrometer (Ocean Optics) with a fibre-optic reflectance probe (Ocean Optics R-200) and PX xenon light source to obtain two opposite orientation measurements per set of 3-5 forecrown feathers from each of 43 individuals for which we collected forecrown feathers. Reflectance spectra were obtained following calibration of the spectrophotometer with a white standard (Ocean Optics WS-1), and dark standard (by covering the fibre-optic connector with its lid). Replicate spectra were averaged in the R package pavo (Maia et al., 2013). Colour was estimated following (Endler, 1990), by computing measures of chroma and hue.

Plumage scoring was performed by two project participants that had no experience from the contact zone and hitherto had no involvement in the study (MM, BOO). They were asked to score forecrown colour of 78 individuals captured and photographed in the field on an integer scale from yellow (0) through orange (1) to red (2). To test whether the same SNPs might otherwise explain other carotenoid-based plumage differences between the species, supercilium colour was also scored from white (0) through cream (1) to yellow (2), and percentage coverage of yellow over black on the wing bar was also estimated (Fig. 1). We then compared the scores of the two observers using spearman rank correlation for each plumage patch, and with  $\rho > 0.85$  in all cases, we averaged the two scores for further analysis. For forecrown colour we designated intermediate scores between red and orange as reddish and between orange and yellow as amber.

### *DNA extraction and genotyping*

We extracted DNA from blood samples collected using a Qiagen QIAmp DNA Blood Mini kit (Qiagen, Valencia, CA, USA) following the manufacturer's protocols. Extracted DNA from samples were quantified using a NanoDrop (Thermo Scientific 2000c) and a Qubit 2.0 fluorometer with the DNA HS assay kit (Life Technologies). We also checked for DNA quality based on migration on agarose gel, which allowed for the selection of samples with appropriate DNA concentration ( $> 20$  ng/ $\mu$ L) and molecular weight ( $> 10000$  bp). We sequenced 79 yellow-fronted and red-fronted tinkerbirds using a double-digest restriction-site associated DNA lab protocol (Brelsford et al., 2016), which incorporates elements of protocols from previous studies (Parchman et al., 2012; Peterson et al., 2012). Briefly, genomic DNA was digested with restriction enzymes SbfI and MseI and double-stranded adapters with inline barcodes of 4 to 8 bp were ligated to the resulting fragments. After magnetic bead purification, the fragments were amplified by PCR in 4 replicate reactions per individual sample and pooled. We implemented a final purification and size selection of the pooled library using magnetic beads in order to remove primer dimers [and fragments less than 150 bp]. Our library was then and sequenced on two lanes on the Illumina HiSeq X platform by Novogene Inc., with 150 bp paired-end reads.

### *Reference genome assembly*

We assembled a low-cost draft genome for *Pogoniulus p. pusillus*. A single male individual (sample no. AR93139) was selected for assembly, and a short-insert library was prepared by Novogene, Inc., using the NEBNext Ultra II DNA kit (New England Biolabs). The library was sequenced to a depth of approximately 50x on the Illumina HiSeq X platform at Novogene Inc., with 150 bp paired-end reads. Overlapping read pairs were collapsed and adapter and low-quality sequence removed prior to assembly using Pear v0.9.10 (Zhang et al., 2014) with minimum overlap size 20, minimum read length 30, quality score threshold 20, and maximum proportion of uncalled bases 0.02. We assembled the resulting reads with SOAPdenovo v2.04

(Luo et al., 2012) for each odd-numbered value of  $k$  between 41 and 111, with default values for all other parameters. The assembly for  $k=93$  was chosen on the basis of higher scaffold N50 and assembly length closer to the expected genome size for birds.

We aligned these scaffolds to the zebra finch (*Taeniopygia guttata*) genome using the Nucmer command in MUMmer 4.0 (Kurtz et al., 2004). Tinkerbird scaffolds aligning to zebra finch chromosomes were ordered and oriented according to these alignments; scaffolds that did not align to the zebra finch genome were ordered by scaffold size.

#### *Single nucleotide polymorphism (SNP) calling*

Adapter sequences were removed and overlapping paired-end reads merged in PEAR v0.9.10 (Zhang et al., 2014) and reads aligned to the *P.p. pusillus* draft genome assembly in BWA MEM v0.7.17 (Li, 2013) using default parameters. Variants were called using bcftools mpileup v 1.8 (Li et al., 2009) with mapping quality  $> 20$  and default values for other parameters. We filtered the resulting variants using VCFtools v 0.1.13 (Danecek et al., 2011), retaining genotypes with depth 4 or greater, and loci with minor allele frequency  $> 0.05$  that were genotyped in at least 80% of individuals.

The SNPs of interest in this study were those most likely to explain differences between red and yellow forecrown colour. Samtools has been shown to perform relatively poorly in calling indels (Hwang et al., 2015), so we manually inspected the genotype for the SNPs most strongly associated with forecrown colour traits for any indels that may have been incorrectly called.

#### *Genome-wide association analysis*

We tested for an association between each genetic marker and each of the five colour traits using a linear mixed model implemented in GEMMA 0.94 (Zhou & Stephens, 2012), which uses a relatedness matrix to control for population structure. Because GEMMA requires loci with no missing data, we imputed missing genotypes using Beagle v5.1 (Browning et al., 2018), based on genotype likelihood scores. Significance was assessed based on a threshold of  $-\log_{10}(a) = 6$ , which results in approximately one expected false positive per trait, taking into account the 104,933 variants tested for each trait. Following Brelford et al. (2017), we set this significance threshold for consistency across all traits rather than using false discovery rates (FDR) (Benjamini & Hochberg, 1995), which would result in different significance thresholds per trait. We note that this threshold was approximately in line with the most conservative threshold obtained using FDR, but FDR would also have resulted in significance thresholds as low as  $-\log_{10}(a) = 4$  for certain traits. Scaffolds significantly associated with forecrown colour were then compared with sequences of candidate genes *CYP2J19* on the zebra finch genome and *BC02* on the downy woodpecker (*Dryobates pubescens*) genome using BLAST.

#### *Genetic Sex identification*

We used two complementary methods to sex individuals. One procedure involved polymerase chain reaction (PCR) using highly conserved primers 2550F and 2718R (Fridolfsson & Ellegren, 1999) to amplify the differently sized introns of Z- and W-linked chromohelicase-DNA binding protein 1 (CHD1) genes. PCR products were separated on a stained 1% or 2% agarose gel run in TAE buffer, revealing one band for male birds and two for females. The other method was a bioinformatics approach where we compared the ratio of average read depth of SNPs from ddRAD sequencing that were aligned with the zebra finch on the Z chromosome with read depths of SNPs on autosomes (see Toews et al., 2018). Because males have two copies and females one copy of the Z chromosome, we would expect a Z:autosome depth ratio of 1 in males and 0.5 in females. Using the two approaches was confirmatory of the accuracy of each method and allowed us to investigate and correct for any possible error if they produced conflicting results.

#### *Identification of hybrid individuals and direction of backcrossing*

In order to determine whether individuals with intermediate ancestry were progeny of recent interbreeding or historical events, we used Faststructure (Raj et al. 2014) with  $k = 2$  to estimate *pusillus* and *extoni* ancestry

(i.e., hybrid index) based on all loci, and separately for loci on scaffolds that aligned to the zebra finch Z chromosome. We estimated interspecific heterozygosity of each individual based on a genome-wide subset of 1,723 SNPs with allele frequency difference  $> 0.95$  between allopatric *pusillus* and *extoni* populations, and compared this to the hybrid index across the whole genome and the Z chromosome to identify F1 hybrids, first generation backcrosses and the direction of backcrossing for early and later-generation backcrosses.

To illustrate the relationship between hybrid index and interspecific heterozygosity, we produced triangle plots, which would reveal F1 hybrids as individuals with intermediate hybrid index and the highest levels of interspecific heterozygosity at loci diagnostic for each species. F2 hybrids and backcrosses would be expected to have intermediate levels of interspecific heterozygosity and fall in the middle and sides of the triangle respectively, in accordance with simulations (Toews et al., 2018), with non-admixed individuals falling in the lower corners of the triangle.

### *Data analysis*

We tested for an effect of sex on plumage coloration characters using generalized linear models (GLM) in the R package lme4 (Bates et al., 2015). We included the hybrid index score as a fixed factor in the model. In particular, we tested the extent to which differences in plumage scores for forecrown, supercilium, and wing bar, as well as the calculated hue and chroma from spectrophotometry of forecrown feathers, were associated with sex. False discovery rates were calculated to account for the five comparisons (Benjamini & Hochberg, 1995). We also tested for the direction of asymmetric backcrossing by comparing the hybrid index of the whole genome and the Z chromosome for individuals with a hybrid index between 0.2 and 0.8 using a paired t-test in R. All analyses were performed in R version 3.6.2 (R Core R Core Team, 2019).

## **RESULTS**

### *Reference genome assembly*

The assembly for  $k = 93$  resulted in a scaffold N50 of 7.2 Kbp. The assembled genome spanned 1.16 Gbp, and 53% of the assembly was ordered and oriented onto chromosomes on the high-quality zebra finch reference genome (Kirschel & Brelsford, 2020).

### *SNP calling*

We retained 104,933 SNPs, with a mean depth of 19X, a minor allele frequency  $< 0.05$ , which had been genotyped on at least 80% of the 79 individuals. Of these SNPs, 1,532 were on scaffolds anchored to the Z chromosome, 57,913 were on scaffolds anchored to autosomes, and 45,487 were on unplaced scaffolds.

### *Genome-wide association analysis*

Admixture mapping revealed a number of genomic regions that were strongly associated with plumage traits that differ between yellow-fronted and red-fronted tinkerbirds (Table 1). After accounting for the one expected false positive per trait based on our significance threshold of  $-\log_{10}(a = 6)$ , there were eight significant associations of chromosomes with traits. We found that crown colour was significantly associated with three chromosomes. Crown hue was significantly associated with SNPs on chromosome 24 (on scaffolds 70213, 70285 and 70295), crown score with SNPs on chromosome 20 (on scaffolds 67410, 67420, 67433 and 67445), and both were significantly associated with SNPs on chromosome 8 (Fig. 2). Indeed, the same six SNPs on adjacent scaffolds 50344, 50345, and 50346 on chromosome 8 were significantly associated with both crown hue and crown score (as well as 50201 for crown score and 50321 for crown hue). The variant most strongly associated with crown score was an indel on scaffold 50201 on chromosome 8, which aligns to the region 13-21 kbp upstream of *CYP2J19A* in zebra finch, and 16-28 kbp upstream of cytochrome P450 gene XM.009910004 in downy woodpecker (manually annotated as *CYP2J19* by Twyman et al. 2018a). The cluster of SNPs on adjacent scaffolds 50344, 50345, and 50346 on chromosome 8 spans approx. 150 kbp and is located 1.8 Mbp from *CYP2J19A* in zebra finch. It overlaps with several other genes in zebra finch: Leucine-rich repeat-containing protein 42 (*LRRC42*), Intraflagellar transport protein 25 homolog (*HSPB11*), iodothyronine deiodinase 1 (*DIO1*), YIP1 domain family member 1 (*YIPF1*), Nucleoporin

*NDC1* (*NDC1*), and *GLIS* family zinc finger 1 (*GLIS1*). There was no significant association with crown chroma. There was also a significant association with regions of the genome and supercilium score, associated with chromosome 4 (Table 1). There were no chromosomes with >2 significant SNPs associated with wing bar percent of yellow over black.

#### *Sex determination and differences in plumage coloration*

Using PCR and bioinformatics approaches, we identified 44 males and 35 females across the contact zone. The bioinformatics approach involved comparing the mean depth on 1532 SNPs on the Z chromosome with 103,401 other SNPs for each individual. We expected comparatively higher Z chromosome to autosome depth ratios for males compared to females, and we indeed found a clear separation. Male Z:autosome ratios ranged from 0.83 – 1.19, while female ratios ranged from 0.43 – 0.67 (see supporting information). The results of the bioinformatics method corresponded 100% with the PCR and gel electrophoresis method of Fridolfsson and Ellegren (1999).

The GLMs found no differences between the sexes in either forecrown hue or chroma, nor in forecrown, supercilium, or wing bar scores. Hybrid index, on the other hand, was significantly related to all colour measures for crown colour (Table 2) and other plumage patches (Table S1), confirming species differences in those traits.

#### *Identification of hybrid individuals and direction of backcrossing*

Out of 37 individuals sampled in the contact zone, 13 were admixed at the level between first generation backcrosses (with a hybrid index between approx. 0.2 – 0.8, Fig. 3). Among the 13 admixed individuals we identified one F1 hybrid (AR93112), while the remaining hybrids likely had more complex ancestry (Table 3).

Backcrossing was thus asymmetric: autosomes of admixed individuals had a higher percentage of *extoni* compared to *pusillus* genotype than the Z chromosome based on their respective hybrid index (paired t-test:  $df = 12$ ,  $t = 2.55$ ,  $P = 0.026$ , Table 3). This finding suggests that hybrids result from females that have a higher proportion of *extoni* ancestry mating with males with a relatively higher proportion of *pusillus* ancestry, as found in 9 of 13 individuals (compared to 2 with the reverse pattern and 2 with equal indexes). For instance, female AR93167 had autosomal and Z hybrid indexes of 0.559 and 0.000 respectively, with an *extoni* mitochondrial haplotype (Table 3). With over 50% *extoni* autosomal hybrid index (and *extoni* mtDNA haplotype), her mother was likely a ‘pure’ *extoni* and her father mostly *pusillus* with a hybrid index = 0.559 and a Z chromosome hybrid index of 0.000 (Table 3).

The variant most strongly associated with crown score was an indel on scaffold 50201 on chromosome 8, which aligns to the region 13-21 kbp upstream of *CYP2J19A* in zebra finch, and 16-28 kbp upstream of cytochrome P450 gene XM\_009910004 in downy woodpecker (manually annotated as *CYP2J19* by Twyman et al. 2018a). All individuals homozygous for the *pusillus* allele were scored with red forecrowns, while heterozygotes were scored red, reddish or orange, whatever their hybrid index (Fig. 3A). Forecrown score of homozygotes with the *extoni* allele ranged from yellow and amber for non-admixed individuals, through orange for admixed individuals to reddish for one individual (AR93120) with 96% *pusillus* genotype.

Comparing hybrid index with interspecific heterozygosity in a triangle plot revealed that backcrossing was asymmetric into the *extoni* genotype, with red to orange plumage coloration introgressing into *extoni* carrying the *pusillus* allele at *CYP2J19* (Fig. 3B). There was also a very clear relationship between the genotype at the *CYP2J19* locus and crown hue. All heterozygotes at the locus had redder hues than homozygotes at the locus with comparative hybrid index (Fig. 4). But we note that individual AR93120 homozygous for the *extoni* allele at this locus, with a red forecrown also based on her hue measurement, was homozygous for the *pusillus* allele at two other loci on chromosome 8 strongly associated with crown hue (on scaffolds 50344 and 50345; see supporting information).

## DISCUSSION

We found that *CYP2J19* on chromosome 8 was within one of three regions associated with crown colour across a natural hybrid zone in tinkerbirds. Indeed, it was the only region of the genome that was significantly associated with both hue, determined from spectral reflectance, and plumage colour score. Only *extoni* individuals homozygous for the *CYP2J19* -linked SNP exhibited a yellow forecrown, while all individuals homozygous for the *pusillus* allele had red forecrowns. All heterozygotes at the *CYP2J19* -linked locus had either red, reddish or orange crowns. *CYP2J19* is a cytochrome P450 enzyme, which has been implicated as encoding the ketolase that mediates red coloration in birds (Lopes et al., 2016; Mundy et al., 2016). Ketolases convert dietary yellow carotenoids into red ketocarotenoids in birds with red feathers (Toews et al., 2017; Norden & Price, 2018). *CYP2J19* has also been associated with red retinal oil droplets influencing colour vision in birds (Mundy et al., 2016), proposed to be its ancestral function, with its role in red coloration having evolved in some but not other species (Twyman et al., 2018a). It has also been shown to function in red integumental coloration in turtles (Twyman et al., 2016).

*CYP2J19* is not the only gene shown to function in red coloration, however, with recent work finding that a mutation in the beta-carotenoid oxygenase 2 gene (*BCO2*) functions in red bare part coloration in birds (Gazda et al., 2020). The *CYP2J19* linked SNP on chromosome 8 was also not the only region associated with red forecrown colour in tinkerbirds, with chromosome 20 associated with crown score and chromosome 24 with crown hue. *BCO2* is on chromosome 24, and in addition to its association with crown hue, a single SNP on scaffold 70216 on chromosome 24 was also significantly associated with crown score. However, the scaffolds significantly associated with forecrown colour in tinkerbirds on chromosome 24 aligned to the interval from 5.8 to 6.0 Mbp on zebra finch chromosome 24, while *BCO2* spans from 1.53 to 1.54, suggesting that the SNPs associated with forecrown colour are not linked to *BCO2*. We found that forecrown colour could be intermediate between red and yellow, with some heterozygotes at the *CYP2J19* locus displaying orange forecrowns. This suggests possible additive effects of the gene rather than dominance of one allele over the other, and/or interactions with genes at other regions we found associated with forecrown colour. There was a cluster of SNPs on adjacent scaffolds on chromosome 8 that overlapped with several genes in zebra finch, but to our knowledge there are presently no known functional associations of those genes with carotenoid-based coloration determination.

We found no significant association with crown chroma. Unlike in hue, the two species were shown to overlap in chroma between sympatric forms by Nwankwo et al. (2019), with both species showing much more variation in this measure of colour intensity than in the measures of hue and brightness. We did, however, find SNPs significantly associated with supercilium colour, but not the extent of yellow on the wing bar. The SNPs associated with supercilium colour were on a different chromosome to those associated with forecrown colour, indicating that a different pathway is involved in mediating carotenoid-based coloration in those traits. This is not unexpected. Those plumage patches varied between white and shades of yellow, while *CYP2J19* has been shown to function in red feather colour determination (Lopes et al., 2016; Mundy et al., 2016).

In another pair of sister taxa of *Pogoniulus* tinkerbird, yellow-rumped (*P. bilineatus*) and yellow-throated tinkerbird (*P. subsulphureus*), which meet in the forests of west and central Africa, reproductive isolation is maintained through character differences such as throat colour and song, which have diverged in sympatry (Kirschel et al., 2009; Kirschel et al., 2020). By contrast, yellow-fronted and red-fronted tinkerbird hybridise extensively, and hybridisation is asymmetric (Nwankwo et al., 2019). The pattern of asymmetry in putative hybrids with *extoni* ancestry and predominantly *pusillus* Z chromosome suggests red-fronted males are much more likely to mate with yellow-fronted females than the other way around. Even when the reverse scenario does happen, based on the proportion of genotypically *extoni* individuals in the contact zone with red forecrowns (Nwankwo et al., 2019), a high proportion of genotypically *extoni* males mating with *pusillus* females might indeed sport reddish forecrowns.

Such asymmetry between mitochondrial haplotype and hybrid index of the Z chromosome could result from a genetic incompatibility in the reverse cross, i.e. *pusillus* haplotype females interbreeding with *extoni* males. Yet, such crosses do occur – two individuals in our sample with *pusillus* haplotype had *extoni*



fathers (AR93115 and AR93163), and both individuals are females – the heterogametic sex - theoretically more likely to suffer from inviability and infertility (Haldane, 1922). Instead, we suggest the asymmetry is a consequence of female preference for males with red forecrowns in *pusillus* and potentially in *extoni* females (c.f. Baldassarre et al., 2014). The *pusillus* mitochondrial haplotype is less than one quarter as common as the *extoni* haplotype in the contact zone (8/37 individuals, see Nwankwo et al. (2019)). Yet female *pusillus* appear to put the extra effort into finding *pusillus* males, outnumbered four-to-one by their *extoni* counterparts. By contrast, female *extoni* appear to breed with males of either genotype at rates equivalent to their frequency in the population. Mate choice experiments would be needed to test whether female *extoni* would choose male *pusillus* (or otherwise males with *extoni* ancestry but with introgressed red forecrowns) over yellow-fronted male *extoni*, when given a choice. But it does appear to be the failure of female *extoni* to mate assortatively with males of their own species and phenotype likely drives introgressive hybridisation between these species. Premating isolation between the two species remains weak in spite of 4 million years of divergence, according to their mitochondrial DNA phylogeny (Nwankwo et al., 2019). The choosiness of female *pusillus* to mate assortatively on the other hand might be the main factor maintaining a narrow tension zone that does not progress into panmixia and species collapse (e.g., Kleindorfer et al., 2014; Kearns et al., 2018).

Red feathers result from the conversion of dietary carotenoids to red ketocarotenoids and such converted carotenoids have been shown to be an honest indicator of fitness that could function in mate choice (Weaver et al., 2018). Here, female red-fronted tinkerbirds might only mate with males that display red forecrowns because they consider yellow-forecrowned males of lower quality. Greater extents of difference in plumage between related species have been proposed to play a role in premating isolation (Scordato et al., 2017). Here, though yellow-fronted females instigate asymmetric introgression by not mating assortatively with yellow-fronted males. If there was a sensory bias for red plumage coloration the two species would surely collapse into one if there is no loss of fitness in hybrids, and with heterozygotes at the *CYP2J19* locus displaying more reddish forecrowns, we would also expect red plumage to rapidly introgress across the population. Further work is needed to determine the extent to which species boundaries are maintained and the rate of introgression of red plumage across the contact zone.

Both quantitative (hue) and qualitative (crown score) metrics recovered the same major-effect locus. This might not be surprising if they are both measures of the same trait, but each method has its own inherent biases. Scoring from photographs is subjective and affected by each observer’s perspective of the effects of ambient light. While correlated, there was a resultant variation in their respective scores, hence the number of individuals with intermediate scores such as reddish, which resulted from one observer scoring the forecrown red and the other scoring it orange. Likewise, reflectance spectrometry is also susceptible to variation in feather and probe placement, though we did attempt to control for this with measurements in two orientations. Despite these potential biases in qualitative and quantitative metrics, the same major-effect locus was recovered. Furthermore, we showed that admixture mapping is feasible using RADseq with a sample size < 50, as was shown for forecrown hue, but only for major-effect loci. This was possible in spite of no prior genomic resources available for *Pogoniulus* tinkerbirds or any closely-related genera.

## CONCLUSION

We performed a genome wide association study to investigate which genes may underlie differences between red and yellow feathers in tinkerbirds. We found an association with *CYP2J19* on chromosome 8, to our knowledge this is the first time such an association has been found with feather colour across a natural avian hybrid zone, indeed the first time *CYP2J19* has been associated with red feather colour in a nonpasserine bird. Asymmetric hybridisation suggests a preference for red plumage in females that share the red trait, consistent with converted carotenoids reflecting male fitness.

## ACKNOWLEDGEMENTS

We thank Louis Hadjioannou, Kim Mortega, Mnqobi Mamba, Maneno Mbilinyi, and Phumlile Simelane for assistance in the field, Alan Howland, Colleen Downs, David Ellers-Smith, Matt and Stephanie McGinn,

John and Retha Harding, and Neil Baker and the late Liz Baker for assistance with logistics and Andrea Fulgione for helpful comments on the manuscript. We thank The Kingdom of Eswatini's Big Game Parks, Ezemvelo KZN Wildlife, Tanzania Wildlife Research Institute (TAWIRI), Tanzania Commission for Science and Technology (COSTECH), Tanzania National Parks (TANAPA), and Mpumalanga Tourism and Parks Agency for research permits. Funding was provided by FP7 Marie Curie Reintegration Grant No. 268316 and a University of Cyprus Research Grant (ANGK), an AG Leventis Foundation grant (MM) and by the AP Leventis Ornithological Research Institute, Jos, Nigeria (ECN and BOO).

## DATA ACCESSIBILITY

The *Pogoniulus p. pusillus* genome assembly has been deposited to NCBI's Sequence Read Archive with BioSample accession SAMN14819987 in BioProject PRJNA630018. The master VCF file and input and output files for all analyses will be deposited in the Dryad Digital Repository. Phenotypic data are available in the Supporting Information.

## AUTHOR CONTRIBUTIONS

ANGK conceived the study and designed it with AM and AB; ANGK and ECN performed fieldwork with the help of AM; ECN, DP, MM and BOO performed molecular labwork; ECN performed the spectrophotometry; MM and BOO scored plumage traits; AB assembled the reference genome and performed all bioinformatics; AB, ECN and ANGK analysed data; and ANGK wrote the manuscript with contributions from all authors.

## REFERENCES

- Baldassarre, D. T., White, T. A., Karubian, J., & Webster, M. S. (2014). Genomic and Morphological Analysis of a Semipermeable Avian Hybrid Zone Suggests Asymmetrical Introgression of a Sexual Signal. *Evolution*, *68* (9), 2644-2657. doi:10.1111/evo.12457
- Bates, D., Machler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67* (1), 1-48. doi:DOI 10.18637/jss.v067.i01
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Statistical Methodology*, *57* (1), 289-300.
- Billerman, S. M., Cicero, C., Bowie, R. C. K., & Carling, M. D. (2019). Phenotypic and genetic introgression across a moving woodpecker hybrid zone. *Molecular Ecology*, *28* (7), 1692-1708. doi:10.1111/mec.15043
- Birkhead, T. (2008). *The wisdom of birds : an illustrated history of ornithology* . London: Bloomsbury.
- Brelsford, A., Dufresnes, C., & Perrin, N. (2016). High-density sex-specific linkage maps of a European tree frog (*Hyla arborea*) identify the sex chromosome without information on offspring sex. *Heredity*, *116* (2), 177-181. doi:10.1038/hdy.2015.83
- Brelsford, A., Toews, D. P. L., & Irwin, D. E. (2017). Admixture mapping in a hybrid zone reveals loci associated with avian feather coloration. *Proceedings of the Royal Society B-Biological Sciences*, *284* (1866). doi:10.1098/rspb.2017.1106
- Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. *The American Journal of Human Genetics*, *103* (3), 338-348. doi:10.1016/j.ajhg.2018.07.015
- Campagna, L., Repenning, M., Silveira, L. F., Fontana, C. S., Tubaro, P. L., & Lovette, I. J. (2017). Repeated divergent selection on pigmentation genes in a rapid finch radiation. *Science Advances*, *3* (5). doi:10.1126/sciadv.1602404
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., et al. (2011). The variant call format and VCFtools. *Bioinformatics*, *27* (15), 2156-2158. doi:10.1093/bioinformatics/btr330

- Delmore, K. E., Toews, D. P. L., Germain, R. R., Owens, G. L., & Irwin, D. E. (2016). The Genetics of Seasonal Migration and Plumage Color. *Current Biology*, *26* (16), 2167-2173. doi:10.1016/j.cub.2016.06.015
- Dowsett-Lemaire, F. (1988). Fruit choice and seed dissemination by birds and mammals in the evergreen forests of upland Malawi. *Revue d'écologie*, *43*, 251-285.
- Endler, J. A. (1990). On the measurement and classification of colour in studies of animal colour patterns. *Biological Journal of the Linnean Society*, *41* (4), 315-352. doi:10.1111/j.1095-8312.1990.tb00839.x
- Fridolfsson, A. K., & Ellegren, H. (1999). A simple and universal method for molecular sexing of non-ratite birds. *Journal of Avian Biology*, *30* (1), 116-121. doi:10.2307/3677252
- Gazda, M. A., Toomey, M. B., Araújo, P. M., Lopes, R. J., Afonso, S., Myers, C. A., Serres, K., Kiser, P. D., et al. (2020). Genetic Basis of De Novo Appearance of Carotenoid Ornamentation in Bare Parts of Canaries. *Molecular Biology and Evolution*. doi:10.1093/molbev/msaa006
- Godschalk, S. K. B. (1985). Feeding-Behavior of Avian Dispersers of Mistletoe Fruit in the Loskop-Dam-Nature-Reserve, South-Africa. *South African Journal of Zoology*, *20* (3), 136-146.
- Haas, F., Pointer, M. A., Saino, N., Brodin, A., Mundy, N. I., & Hansson, B. (2009). An analysis of population genetic differentiation and genotype-phenotype association across the hybrid zone of carrion and hooded crows using microsatellites and MC1R. *Molecular Ecology*, *18* (2), 294-305. doi:10.1111/j.1365-294X.2008.04017.x
- Haldane, J. B. S. (1922). Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics*, *7*, 101-109.
- Harrison, R. G., & Larson, E. L. (2016). Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology*, *25* (11), 2454-2466. doi:10.1111/mec.13582
- Hockey, P. A. R., Dean, W. R. J., & Ryan, P. (2005). *Roberts Birds of Southern Africa* (7th ed.). Cape Town: Trustees of the John Voelcker Bird Book Fund.
- Hoekstra, H. E. (2006). Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity*, *97* (3), 222-234. doi:10.1038/sj.hdy.6800861
- Hooper, D. M., Griffith, S. C., & Price, T. D. (2019). Sex chromosome inversions enforce reproductive isolation across an avian hybrid zone. *Molecular Ecology*, *28* (6), 1246-1262. doi:10.1111/mec.14874
- Hwang, S., Kim, E., Lee, I., & Marcotte, E. M. (2015). Systematic comparison of variant calling pipelines using gold standard personal exome variants. *Scientific Reports*, *5* (1). doi:10.1038/srep17875
- Kearns, A. M., Restani, M., Szabo, I., Schröder-Nielsen, A., Kim, J. A., Richardson, H. M., Marzluff, J. M., Fleischer, R. C., et al. (2018). Genomic evidence of speciation reversal in ravens. *Nature Communications*, *9* (1). doi:10.1038/s41467-018-03294-w
- Kim, K.-W., Jackson, B. C., Zhang, H., Toews, D. P. L., Taylor, S. A., Greig, E. I., Lovette, I. J., Liu, M. M., et al. (2019). Genetics and evidence for balancing selection of a sex-linked colour polymorphism in a songbird. *Nature Communications*, *10* (1). doi:10.1038/s41467-019-09806-6
- Kirschel, A. N. G., Blumstein, D. T., & Smith, T. B. (2009). Character displacement of song and morphology in African tinkerbirds. *Proceedings of the National Academy of Sciences of the United States of America*, *106* (20), 8256-8261. doi:10.1073/pnas.0810124106
- Kirschel, A. N. G., & Brelsford, A. (2020). *BioProject PRJNA630018*. *NCBI Sequence Read Archive*.
- Kirschel, A. N. G., Nwankwo, E. C., & Gonzalez, J. C. T. (2018). Investigation of the status of the enigmatic White-chested Tinkerbird *Pogoniulus makawai* using molecular analysis of the type specimen. *Ibis*, *160* (3), 673-680. doi:10.1111/ibi.12597

- Kirschel, A. N. G., Nwankwo, E. C., Seal, N., & Grether, G. F. (2020). Time spent together and time spent apart affect song, feather colour and range overlap in tinkerbirds. *Biological Journal of the Linnean Society*, *129* (2), 439-458. doi:10.1093/biolinnean/blz191
- Kleindorfer, S., O'Connor, J. A., Dudaniec, R. Y., Myers, S. A., Robertson, J., & Sulloway, F. J. (2014). Species Collapse via Hybridization in Darwin's Tree Finches. *American Naturalist*, *183* (3), 325-341. doi:10.1086/674899
- Kurtz, S., Phillippy, A., Delcher, A. L., Smoot, M., Shumway, M., Antonescu, C., & Salzberg, S. L. (2004). Versatile and open software for comparing large genomes. *Genome Biology*, *5* (2). doi:doi:10.1186/gb-2004-5-2-r12
- Layard, E. L. (1871). Notes on South-African Ornithology. *Ibis*, *13* (2), 225-234. doi:10.1111/j.1474-919X.1871.tb05837.x
- Leroi, A. M. (2014). *The lagoon : how Aristotle invented science* . In (pp. 1 online resource).
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv e-prints* , arXiv:1303.3997. Retrieved from <https://ui.adsabs.harvard.edu/abs/2013arXiv1303.3997L>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., et al. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25* (16), 2078-2079. doi:10.1093/bioinformatics/btp352
- Lopes, R. J., Johnson, J. D., Toomey, M. B., Ferreira, M. S., Araujo, P. M., Melo-Ferreira, J., Andersson, L., Hill, G. E., et al. (2016). Genetic Basis for Red Coloration in Birds. *Current Biology*, *26* (11), 1427-1434. doi:10.1016/j.cub.2016.03.076
- Luo, R. B., Liu, B. H., Xie, Y. L., Li, Z. Y., Huang, W. H., Yuan, J. Y., He, G. Z., Chen, Y. X., et al. (2012). SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience*, *1* . doi:10.1186/2047-217x-1-18
- Maia, R., Eliason, C. M., Bitton, P. P., Doucet, S. M., & Shawkey, M. D. (2013). pavo: an R package for the analysis, visualization and organization of spectral data. *Methods in Ecology and Evolution*, *4* (10), 906-913. doi:10.1111/2041-210x.12069
- Mayr, E. (1942). *Systematics and the origin of species : from the viewpoint of a zoologist* . New York: Columbia University Press.
- Monadjem, A., Passmore, N. I., & Kemp, A. C. (1994). Territorial Calls of Allopatric and Sympatric Populations of 2 Species of Pogoniulus Tinkerbarbet in Southern Africa. *Ostrich*, *65* (3-4), 339-341.
- Mundy, N. I. (2005). A window on the genetics of evolution: MC1R and plumage colouration in birds. *Proceedings of the Royal Society B: Biological Sciences*, *272* (1573), 1633-1640. doi:10.1098/rspb.2005.3107
- Mundy, N. I., Stapley, J., Bennison, C., Tucker, R., Twyman, H., Kim, K.-W., Burke, T., Birkhead, T. R., et al. (2016). Red Carotenoid Coloration in the Zebra Finch Is Controlled by a Cytochrome P450 Gene Cluster. *Current Biology*, *26* (11), 1435-1440. doi:10.1016/j.cub.2016.04.047
- Norden, K. K., & Price, T. D. (2018). Historical Contingency and Developmental Constraints in Avian Coloration. *Trends in Ecology & Evolution*, *33* (8), 574-576. doi:10.1016/j.tree.2018.05.003
- Nosil, P., & Schluter, D. (2011). The genes underlying the process of speciation. *Trends in Ecology & Evolution*, *26* (4), 160-167. doi:10.1016/j.tree.2011.01.001
- Nwankwo, E. C., Mortega, K. G., Karageorgos, A., Ogolowa, B. O., Papagregoriou, G., Grether, G. F., Monadjem, A., & Kirschel, A. N. G. (2019). Rampant introgressive hybridization in Pogoniulus tinkerbirds (Piciformes: Lybiidae) despite millions of years of divergence. *Biological Journal of the Linnean Society*, *127* (1), 125-142.

- Nwankwo, E. C., Pallari, C. T., Hadjioannou, L., Ioannou, A., Mulwa, R. K., & Kirschel, A. N. G. (2018). Rapid song divergence leads to discordance between genetic distance and phenotypic characters important in reproductive isolation. *Ecology and Evolution*, *8* (1), 716-731. doi:10.1002/ece3.3673
- Parchman, T. L., Gompert, Z., Mudge, J., Schilkey, F. D., Benkman, C. W., & Buerkle, C. A. (2012). Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology*, *21* (12), 2991-3005. doi:10.1111/j.1365-294X.2012.05513.x
- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *Plos One*, *7* (5). doi:10.1371/journal.pone.0037135
- Poelstra, J. W., Vijay, N., Bossu, C. M., Lantz, H., Ryll, B., Muller, I., Baglione, V., Unneberg, P., et al. (2014). The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science*, *344* (6190), 1410-1414. doi:10.1126/science.1253226
- R Core Team. (2019). A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
- Scordato, E. S. C., Wilkins, M. R., Semenov, G., Rubtsov, A. S., Kane, N. C., & Safran, R. J. (2017). Genomic variation across two barn swallow hybrid zones reveals traits associated with divergence in sympatry and allopatry. *Molecular Ecology*, *26* (20), 5676-5691. doi:10.1111/mec.14276
- Toews, D. P. L., Hofmeister, N. R., & Taylor, S. A. (2017). The Evolution and Genetics of Carotenoid Processing in Animals. *Trends in Genetics*, *33* (3), 171-182. doi:10.1016/j.tig.2017.01.002
- Toews, D. P. L., Lovette, I. J., Irwin, D. E., & Brelsford, A. (2018). Similar hybrid composition among different age and sex classes in the Myrtle–Audubon’s warbler hybrid zone. *The Auk*, *135* (4), 1133-1145. doi:10.1642/auk-18-45.1
- Toews, D. P. L., Taylor, S. A., Vallender, R., Brelsford, A., Butcher, B. G., Messer, P. W., & Lovette, I. J. (2016). Plumage Genes and Little Else Distinguish the Genomes of Hybridizing Warblers. *Current Biology*, *26* (17), 2313-2318. doi:10.1016/j.cub.2016.06.034
- Toomey, M. B., Lopes, R. J., Araújo, P. M., Johnson, J. D., Gazda, M. A., Afonso, S., Mota, P. G., Koch, R. E., et al. (2017). High-density lipoprotein receptor SCARB1 is required for carotenoid coloration in birds. *Proceedings of the National Academy of Sciences*, *114* (20), 5219-5224. doi:10.1073/pnas.1700751114
- Twyman, H., Andersson, S., & Mundy, N. I. (2018a). Evolution of CYP2J19, a gene involved in colour vision and red coloration in birds: positive selection in the face of conservation and pleiotropy. *BMC Evolutionary Biology*, *18* (1). doi:10.1186/s12862-018-1136-y
- Twyman, H., Prager, M., Mundy, N. I., & Andersson, S. (2018b). Expression of a carotenoid-modifying gene and evolution of red coloration in weaverbirds (Ploceidae). *Molecular Ecology*, *27* (2), 449-458. doi:10.1111/mec.14451
- Twyman, H., Valenzuela, N., Literman, R., Andersson, S., & 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). doi:10.1098/rspb.2016.1208
- Walsh, J., Shriver, W. G., Olsen, B. J., & Kovach, A. I. (2016). Differential introgression and the maintenance of species boundaries in an advanced generation avian hybrid zone. *BMC Evolutionary Biology*, *16* (1). doi:10.1186/s12862-016-0635-y
- Weaver, R. J., Santos, E. S. A., Tucker, A. M., Wilson, A. E., & Hill, G. E. (2018). Carotenoid metabolism strengthens the link between feather coloration and individual quality. *Nature Communications*, *9* . doi:10.1038/s41467-017-02649-z

Zhang, J. J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, 30 (5), 614-620. doi:10.1093/bioinformatics/btt593

Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, 44 (7), 821-U136. doi:10.1038/ng.2310

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section for this article.

Table 1: Number of SNPs associated with each of eight feather patch traits. Only chromosomes with >2 significant SNPs are shown with associations to each trait because based on the permutation test, we expect one false positive at  $-\log_{10}(P) > 6$ .

Trait	Number of SNPs $-\log_{10}(P) > 6$	Chromosomes with >2 significant SNPs
Crown hue	19	8, 24
Crown chroma	3	none
Crown score	19	8, 20
Wing bar score	1	none
Supercilium score	9	4

Table 2: Results of GLMs testing for associations of sex and hybrid index with crown score, hue and chroma.

<b>Crown Score</b>	<b>Estimate</b>	<b>SD</b>	<b>t</b>	<b>P</b>
Intercept	2.08	0.1	20.62	<0.0001
Sex	0.01	0.09	0.12	0.906
Hybrid Index	-1.95	0.11	-18.58	<0.0001
<b>Crown Hue</b>				
Intercept	0.16	0.05	2.93	0.006
Sex	-0.05	0.05	-0.95	0.35
Hybrid Index	0.47	0.06	8.28	<0.0001
<b>Crown Chroma</b>				
Intercept	1451.8	253.7	5.72	<0.0001
Sex	346.7	234.6	1.48	0.15
Hybrid Index	966.6	268.5	3.6	0.0009



—

Table 3: Admixed individuals, their autosomal hybrid index, Z chromosome hybrid index, interspecific heterozygosity, mtDNA haplotype (from Nwankwo et al., 2019), and sex. Banner (above) shows photos of forecrown for each individual except AR93119 in the order presented in the table.

individual	hybrid index autosomes	hybrid index Z chromos.	inter-specific heterozygosity	mtDNA hap
AR93112	0.498	0.495	0.84	<i>extoni</i>
AR93115	0.681	1.000	0.41	<i>pusillus</i>
AR93118	0.493	0.210	0.60	<i>extoni</i>
AR93122	0.689	0.097	0.46	<i>extoni</i>
AR93153	0.696	0.878	0.37	<i>extoni</i>
AR93157	0.632	0.417	0.36	<i>extoni</i>
AR93159	0.629	0.490	0.66	<i>extoni</i>
AR93167	0.559	0.000	0.61	<i>extoni</i>
AR93168	0.695	0.284	0.51	<i>extoni</i>
AR93170	0.695	0.645	0.46	<i>extoni</i>

individual	hybrid index autosomes	hybrid index Z chromos.	inter-specific heterozygosity	mtDNA hap
AR93172	0.647	0.260	0.53	<i>extoni</i>
K69365	0.459	0.000	0.65	<i>extoni</i>
AR93119	0.249	0.250	0.41	<i>extoni</i>

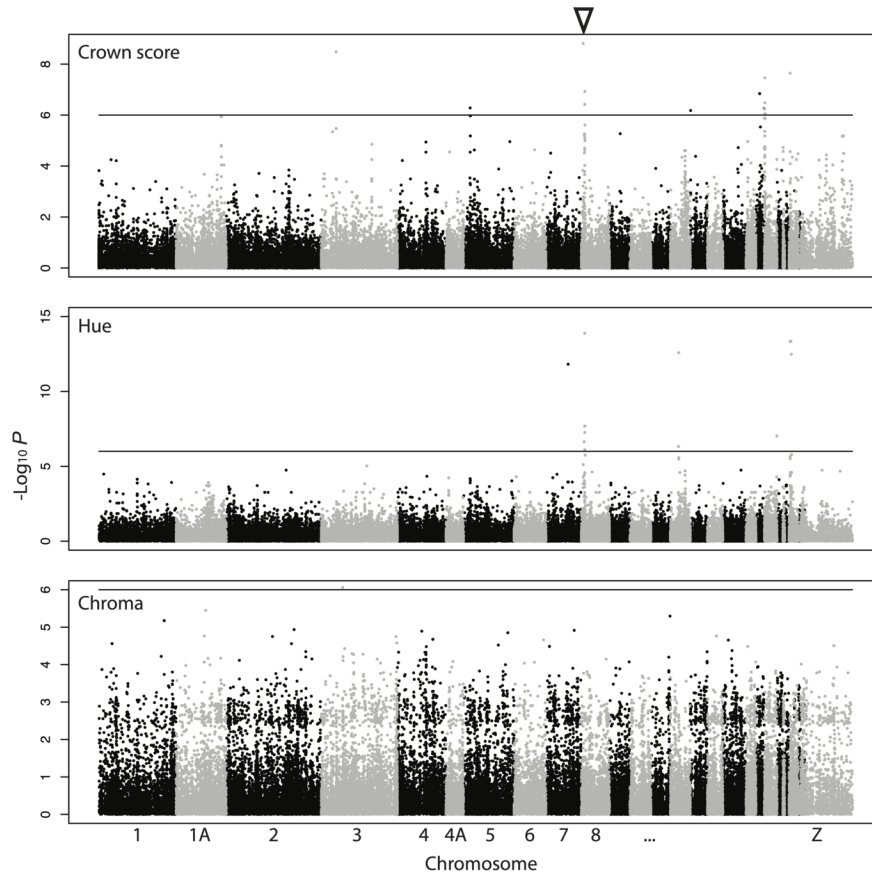
+F = Female, M = Male.

### Hosted file

image2.emf available at <https://authorea.com/users/317820/articles/447916-cyp2j19-mediates-carotenoid-colour-introgression-across-a-natural-avian-hybrid-zone>

### Figure 1

Forecrown plumage (1) coloration was scored in: A) *Pogoniulus pusillus pusillus* ; and B) *Pogoniulus chrysoconus extoni* . We also scored: (2) supercilium; and (3) wing bar extent of yellow over black, from photographs, as shown of C) *P. p. pusillus* and D) *P. c. extoni* from allopatric populations. Samples, photographs scored and forecrown feathers were obtained from individuals that were phenotypically red-fronted (red circles) and yellow-fronted (yellow circles) across the contact zone (green line) in southern Africa (E).

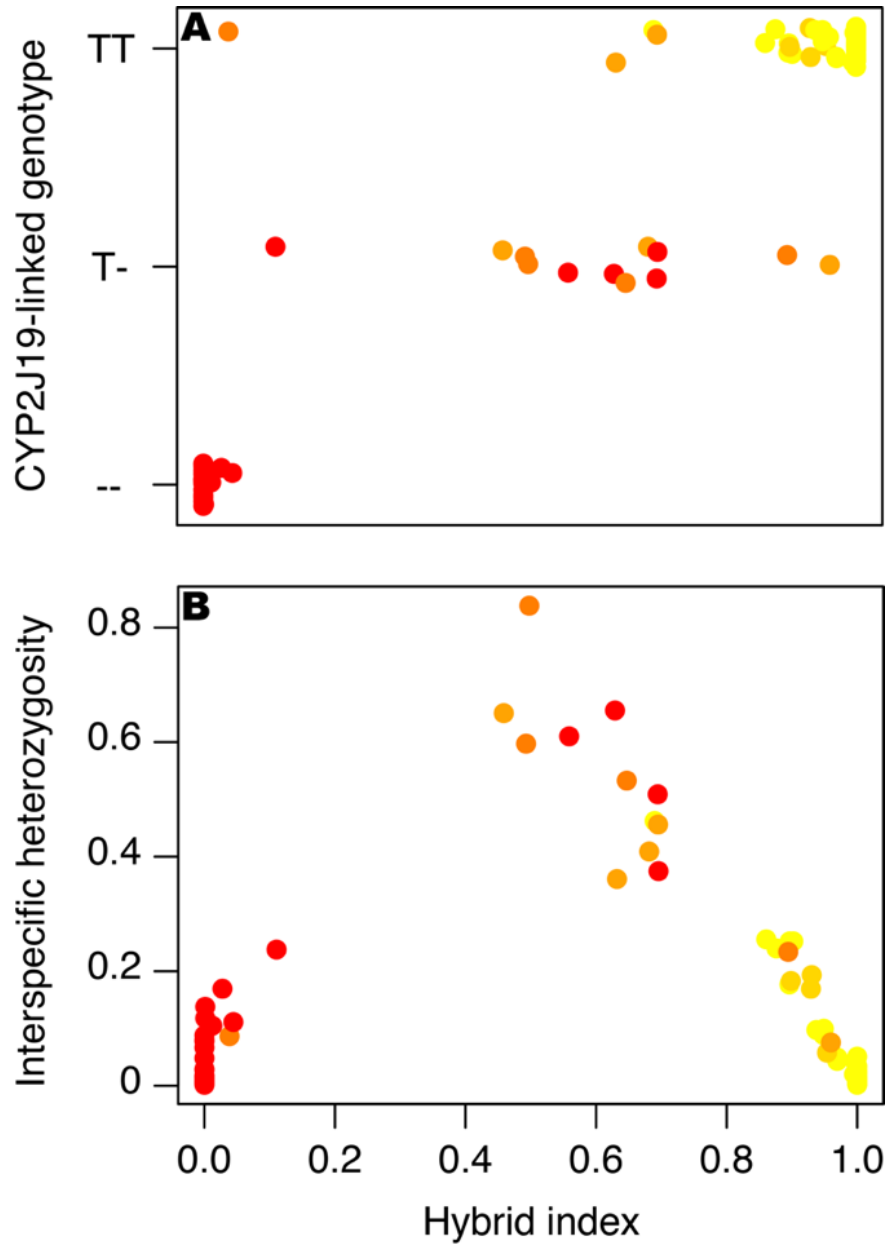


### Figure 2

Manhattan plots for crown score, hue and chroma revealing the region of chromosome 8 significantly associated with red and yellow colour differences (black triangle points to the region). The locus is associated

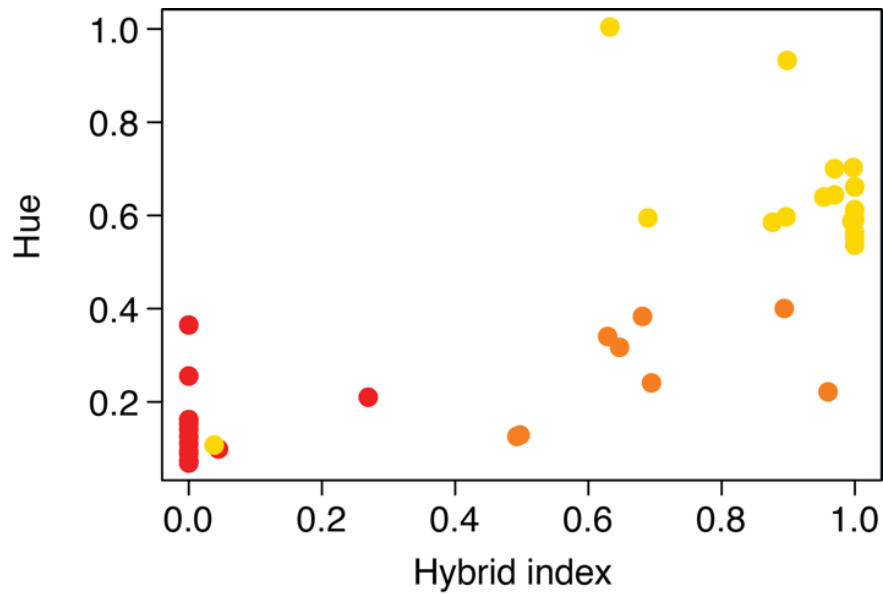


with the *CYP2J19* gene. Black line indicates genome-wide significance at  $-\log_{10}(P) > 6$ .



**Figure 3**

Forecrown colour, represented by dot colour (scale: red-reddish-orange-amber-yellow) in relation to hybrid index on a scale from 0 (*P. p. pusillus* genotype) to 1 (*P. c. extonigena* genotype), and A) genotype of an indel linked to *CYP2J19* on scaffold 50201 on chromosome 8, and B) interspecific heterozygosity. While non hybrids homozygous for their respective allele are characterised by species-specific forecrown colour, individuals heterozygous at the *CYP2J19*-linked SNP have orange to red forecrowns, even in *extoni* genotypes with little admixture evident. Backcrosses are mostly into *extoni*.



**Figure 4**

Forecrown hue in relation to hybrid index. Dot shade represents the genotype at the variant on scaffold 50201 linked to *CYP2J19*, with the (-/-) homozygote in red, (T/T) in yellow, and the heterozygote in orange. Individuals homozygous for the red allele or heterozygous at the locus have non overlapping hue with individuals homozygous for the yellow allele, except for one individual (AR93120) whose genome is 96% *pusillus* based on its hybrid index and was scored reddish from its photo. Notably, all individuals heterozygous at the locus have redder hue (with hue < 0.5) than all the other individuals homozygous for the yellow allele (with hue > 0.5).

