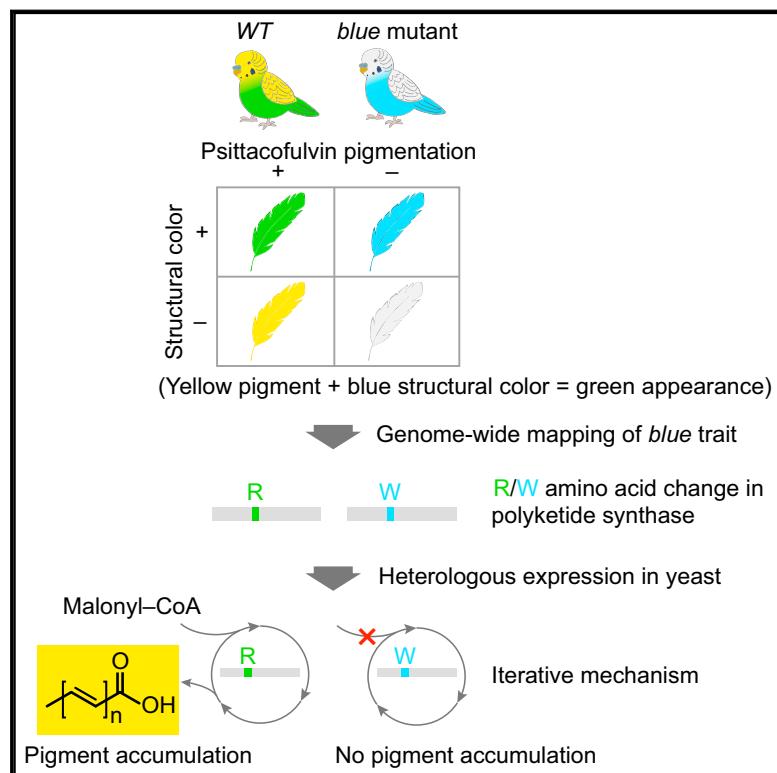


Genetic Mapping and Biochemical Basis of Yellow Feather Pigmentation in Budgerigars

Graphical Abstract



Authors

Thomas F. Cooke, Curt R. Fischer,
Ping Wu, ..., Chaitan Khosla,
Cheng-Ming Chuong,
Carlos D. Bustamante

Correspondence

cdbustam@stanford.edu

In Brief

An enzyme required for yellow pigmentation in a popular variety of pet parrots is discovered through genome-wide mapping and biochemical studies.

Highlights

- Polyene pigment trait in budgerigars maps to uncharacterized polyketide synthase
- Amino acid substitution at a conserved residue is the causative variant
- Co-opting vertebrate polyketide synthase for novel evolutionary use
- Derived expression pattern confers colorful pigmentation in parrot feathers

Genetic Mapping and Biochemical Basis of Yellow Feather Pigmentation in Budgerigars

Thomas F. Cooke,¹ Curt R. Fischer,^{4,5} Ping Wu,⁸ Ting-Xin Jiang,⁸ Kathleen T. Xie,² James Kuo,⁶ Elizabeth Doctorov,¹ Ashley Zehnder,³ Chaitan Khosla,^{4,6,7} Cheng-Ming Chuong,^{8,9,10} and Carlos D. Bustamante^{1,3,11,*}

¹Department of Genetics

²Department of Biochemistry

³Department of Biomedical Data Science

Stanford University School of Medicine, Stanford, CA 94305, USA

⁴ChEM-H

⁵Stanford Genome Technology Center

⁶Department of Chemical Engineering

⁷Departments of Chemistry

Stanford University, Stanford, CA 94305, USA

⁸Department of Pathology, University of Southern California, Los Angeles, CA 90033, USA

⁹Integrative Stem Cell Center, China Medical University, Taichung 404, Taiwan

¹⁰Center for the Integrative and Evolutionary Galliformes Genomics, National Chung Hsing University, Taichung 402, Taiwan

¹¹Lead contact

*Correspondence: cdbustam@stanford.edu

<https://doi.org/10.1016/j.cell.2017.08.016>

SUMMARY

Parrot feathers contain red, orange, and yellow polyene pigments called psittacofulvins. Budgerigars are parrots that have been extensively bred for plumage traits during the last century, but the underlying genes are unknown. Here we use genome-wide association mapping and gene-expression analysis to map the Mendelian *blue* locus, which abolishes yellow pigmentation in the budgerigar. We find that the *blue* trait maps to a single amino acid substitution (R644W) in an uncharacterized polyketide synthase (*MuPKS*). When we expressed *MuPKS* heterologously in yeast, yellow pigments accumulated. Mass spectrometry confirmed that these yellow pigments match those found in feathers. The R644W substitution abolished *MuPKS* activity. Furthermore, gene-expression data from feathers of different bird species suggest that parrots acquired their colors through regulatory changes that drive high expression of *MuPKS* in feather epithelia. Our data also help formulate biochemical models that may explain natural color variation in parrots.

INTRODUCTION

Birds display one of the greatest ranges of coloration mechanisms found in vertebrates. Melanins, carotenoids, porphyrins, pterins, polyenes, and structural colors have all been observed in feathers (McGraw, 2006a). Some of these mechanisms, such as carotenoids and structural colors, appear to have

evolved independently multiple times (Stoddard and Prum, 2011). In most cases, however, the causative genes remain unknown.

Parrots synthesize a unique class of red-to-yellow polyene pigments called psittacofulvins (McGraw, 2006a). Although the existence of psittacofulvins has been known for over a century (Fox, 1976), and their chemical structures have been estimated from mass-spectrometry data (Stradi et al., 2001), little is known of their biosynthetic origin. Unlike carotenoids, which produce similar colors (McGraw, 2006b), psittacofulvins do not come from dietary sources (McGraw and Nogare, 2005), suggesting they are produced by an uncharacterized biochemical pathway. Several adaptive functions for psittacofulvins, including mate choice (Masello and Quillfeldt, 2003) and protection of feathers from bacterial degradation (Burtt et al., 2011), have been proposed. Psittacofulvins may also be involved in the fluorescent plumage exhibited by many parrot species (Hausmann et al., 2003).

In the budgerigar (*Melopsittacus undulatus*), a member of the parrot family, captive breeding since the 19th century has generated a colorful variety of feather phenotypes, including the loss of psittacofulvin pigmentation. These yellow pigments are present throughout the feathers of wild-type budgerigars. The recessive Mendelian *blue* trait abolishes yellow pigmentation, revealing an underlying blue structural color on parts of the body and tail, which otherwise appear green when yellow pigment is present (Auber, 1941; D'Alba et al., 2012).

Although no genes have been identified for budgerigar traits, budgerigars are well suited for genetic mapping experiments. Many budgerigar color phenotypes are simple Mendelian traits (Steiner, 1932), there is a wealth of classical genetic knowledge (Taylor and Warner, 1986), and there is a reference genome assembly (Ganapathy et al., 2014). Furthermore, the species is widespread in captivity, breeders often document their birds with pedigree information, and the strong and recent selection

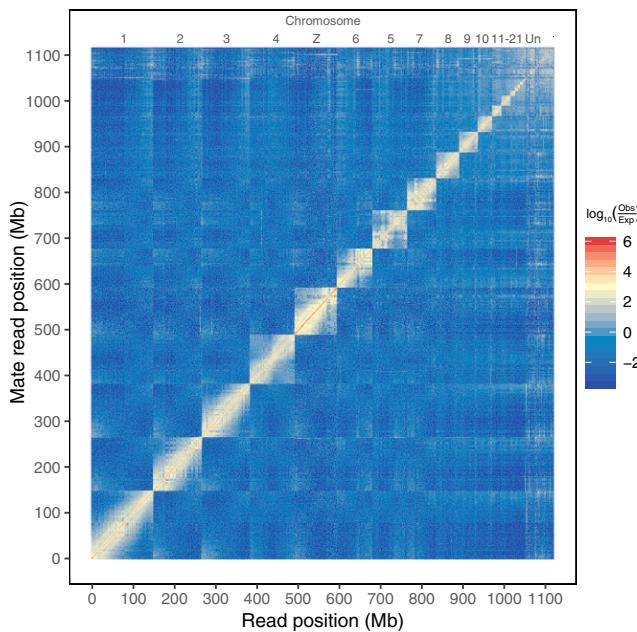


Figure 1. A Hi-C-Based Physical Map of the Budgerigar Genome

The normalized contact frequency matrix for pairs of loci was estimated from chromatin conformation capture (Hi-C) data from budgerigar erythrocytes. For a given pair of loci, the expected number of contacts (Exp) was calculated as the product of the numbers of genome-wide contacts made by each of the two loci, divided by the number of contacts between all loci. The genome was divided into 1 Mb bins, and the ratio of observed (Obs) versus expected contacts within each bin is represented by color. The LACHESIS software was used to cluster the scaffolds according to their contact frequencies by shuffling their order and orientations. The ten largest clusters were matched to the budgerigar karyotype by alignment to probes used in a previous fluorescent *in situ* hybridization (FISH) experiment (Nanda et al., 2007) (see also Figure S2). The remaining clusters were assigned numbers according to size.

on budgerigar color traits makes them amenable to association mapping.

Here we use genome-wide association mapping and gene-expression analysis in regenerating feathers to identify the genetic basis of psittacofulvin pigmentation in budgerigars. We then demonstrate that the gene identified in our analysis, an uncharacterized polyketide synthase, can reconstitute pigment synthesis in a heterologous yeast-based expression system. Finally, we perform a phylogenetic analysis of the understudied family of metazoan homologs to which this gene belongs. Our methodology opens a path for mapping other budgerigar color traits, and our specific results from mapping the *blue* trait help reveal the basis of color differences between budgerigars and other parrot species. These results provide general insights into the evolution of pigments and specialized metabolism.

RESULTS

Physical Map of the Budgerigar Genome

Given the relatively small number of generations since color traits arose in budgerigars, they should exhibit linkage disequilibrium (LD) with nearby genetic markers, making them ideally suited for association mapping. Unlike family-based linkage mapping,

however, association mapping yields no information about the relative positions of the markers. Some positional information is contained in the budgerigar reference genome (Ganapathy et al., 2014), but the budgerigar genomic scaffolds are unordered with respect to chromosomes, as is often the case in the draft genomes of non-model organisms. Unordered scaffolds might lead to mapping difficulties if the causative variant resides on a different scaffold than its association peak, or if the peak is split across multiple scaffolds. To address this concern, we sought to assign each scaffold a chromosomal location.

We used chromosome conformation capture (Hi-C) (Lieberman-Aiden et al., 2009) to measure contact frequencies between pairs of genomic loci and construct a physical map of the budgerigar genome. We chose erythrocytes as the source of nuclei because blood samples can be collected with minimal invasiveness. However, in our experiments, avian erythrocyte chromatin was difficult to isolate and digest with restriction enzymes when we followed established Hi-C protocols, perhaps due to the dense compaction of chromatin that is characteristic of this cell type (Kowalski and Paliya, 2011). We therefore developed a modified Hi-C protocol (Figure S1A). We sequenced Hi-C libraries constructed from blood samples from two female budgerigars, mapped the reads to the budgerigar reference scaffolds, and used the LACHESIS software to cluster, order, and orient the scaffolds based on their contact frequencies (Burton et al., 2013).

After clustering, 22 distinct block domains were apparent in the Hi-C contact matrix (Figure 1). We inferred the identity of ten of the largest blocks with respect to the budgerigar karyotype ($n = 31$) by aligning them to the chicken genome (Figure S2) and comparing this result with data from a previous fluorescent *in situ* hybridization (FISH) experiment in budgerigar with heterologous probes from flow-sorted chicken chromosomes (Nanda et al., 2007). Approximately 94% of all base pairs in the assembly were assigned to a chromosome.

SNP Discovery and Mapping the Mendelian *blue* Locus

To identify the Mendelian locus responsible for the *blue* trait in budgerigars, we discovered single-nucleotide polymorphisms (SNPs) with a modified version of double-digest restriction-site-associated DNA sequencing (ddRAD-seq, Figure S1B) and used these to perform association mapping. We sequenced DNA from 234 domesticated budgerigars, including exhibition and non-exhibition varieties (Bartels et al., 2009), of which 105 showed the *blue* trait (Table S1). In addition, we sequenced DNA from 15 museum specimens from wild Australian populations. Pairwise nucleotide diversity in the domesticated varieties was 0.0037 bp^{-1} , which is similar to levels of diversity in other domesticated birds (Shapiro et al., 2013). After removing low-frequency variants (minor allele frequency < 5%) and sites that failed a likelihood ratio test for restriction-site polymorphism (Cooke et al., 2016), the remaining 69,855 SNPs were used for mapping.

To determine whether the density of SNPs was sufficient for trait mapping, we quantified pairwise LD by the genotype correlation coefficient, r^2 . We found that LD decays rapidly, with $r^2 < 0.3$ (often considered to be the threshold for “useful LD”) (Ardlie et al., 2002) at distances beyond 13 kb (Figure S3A).

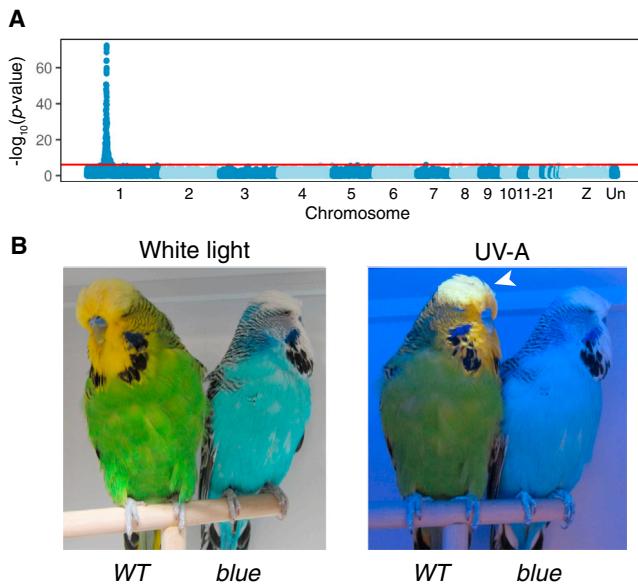


Figure 2. Genome-wide Association Mapping of the *blue* Locus

(A) Fisher's exact test p values (3 genotype classes \times 2 phenotype classes) are shown for 69,855 SNPs segregating in 249 budgerigars. Of these, 105 displayed the recessive *blue* phenotype caused by lack of yellow pigmentation, and 144 displayed WT pigmentation. The red line indicates the Bonferroni-corrected critical value. Unassembled scaffolds are grouped together to the right of the Z chromosome.

(B) WT and *blue* budgerigars photographed under white light (left) or UV-A ("black light") illumination. The crown feathers (arrowhead) of the WT bird, but not the *blue*, exhibit yellow fluorescence under UV-A.

The counts of SNPs in 26 kb bins, genome-wide, suggests that at least half of all trait loci would exhibit LD with one of these markers.

We searched for SNPs associated with yellow pigmentation by testing for significant differences between the distributions of genotypes in wild-type (WT) versus *blue* budgerigars (Figure 2A). A single region on chromosome 1 showed a highly significant association signal ($p = 10^{-72}$, Fisher's exact test). In particular, several SNPs showed genotypes consistent with complete linkage with the causative variant, under a recessive Mendelian model. These data confirm previous classical genetic studies (Steiner, 1932) that a single locus is responsible for the *blue* phenotype in budgerigars and now reveal its precise genomic location.

A Single Haplotype Associated with the *blue* Trait

The simplest genetic model for the loss of yellow pigmentation in budgerigars is that all extant *blue* alleles arose from a single ancestor around 130 years ago, when the trait was first observed (Steiner, 1932). This suggests that the extent of haplotype sharing among *blue* individuals could reveal the precise bounds of the region containing the causative variant. Therefore, we used the software program PHASE (Stephens et al., 2001) to infer the haplotypes of each individual and tested them for association with the *blue* phenotype (Figure 3A). All *blue* individuals shared a single haplotype that is 0.4 Mb long, bounded by SNPs at positions 21,019,187 and 21,445,705 (scaffold coordinates), where ances-

tral recombinations with other haplotypes are evident. Several derived alleles appear to have swept to high frequency on this haplotype, consistent with the history of strong artificial selection for the *blue* trait. Assuming the single-ancestor model, the causative variant must reside within this 0.4 Mb haplotype, which contains 11 predicted genes (Figure 3B).

Given that psittacofulvin pigments are linear-conjugated polyenes (Stradi et al., 2001), we searched the literature for evidence that any of these 11 genes might help synthesize or bind similar compounds. The uncharacterized polyketide synthase gene *LOC101880715*, referred to hereafter as *MuPKS*, was the most promising candidate, as some iterative polyketide synthases in bacteria (Zhang et al., 2008a) and fungi (Zabala et al., 2014) are also known to synthesize yellow polyene pigments.

MuPKS Is Highly Expressed in Regenerating Feathers

Some budgerigars display a rare non-inherited color pattern known as "half-sider," in which one half of the bird is yellow and the other half is non-yellow (Crew and Lamy, 1935), a phenomenon related to gynandromorphism (Agate et al., 2003; Zhao et al., 2010), which can occur by mitotic recombination during early development. This unusual color pattern suggests that the *blue* allele exerts its effect locally within the feather (perhaps in cell-autonomous fashion), rather than systemically, and that the affected gene is expressed during feather development.

To explore this possibility, we performed mRNA-seq on regenerating contour feathers from WT and *blue* budgerigars. Expression of *MuPKS* is in the top 2.7% of genes genome wide (Figure 4A). For comparison, the expression of *KRT75*, a feather keratin responsible for the *frizzle* feather trait in chickens (Ng et al., 2012), is in the top 1.1%. However, we found no evidence of differential expression between WT and *blue* budgerigars for any of the genes within the *blue*-associated haplotype (Figure 4A), indicating that the recessive *blue* phenotype is probably not due to expression changes in one of these genes.

To identify tissue regions expressing *MuPKS*, we performed *in situ* hybridizations in regenerating budgerigar contour feathers (Figures 4B–4E). Longitudinal cross-sections showed that *MuPKS* is expressed highest in the distal tip of the regenerating feather filament. As we saw in the mRNA-seq dataset, the WT and *blue* individuals expressed similar amounts of *MuPKS* (Figures 4D and 4E). Transverse cross-sections revealed strong expression in axial-plate epithelia that are closest to the rachis (or main shaft), which are more mature than those opposite the rachis. Axial-plate epithelia are part of the barb ridges (Figure 4B), and are later lost by apoptosis as the feather matures, allowing the barbules (barb branches) to separate from the ramus (barb shaft) (Chen et al., 2015).

In budgerigar feathers, yellow psittacofulvin pigment is found in the barbules and in the outer cortex of the ramus (Figure 4C) (Auber, 1941; D'Alba et al., 2012). Therefore, if *MuPKS* is the gene responsible for pigment synthesis, pigment produced in the axial plate might be externally deposited on the ramus and barbules, which are in close proximity to the axial plate during development.

To determine whether the level of *MuPKS* expression in regenerating feathers is associated with the presence of these pigments across species, we analyzed recently published RNA-seq

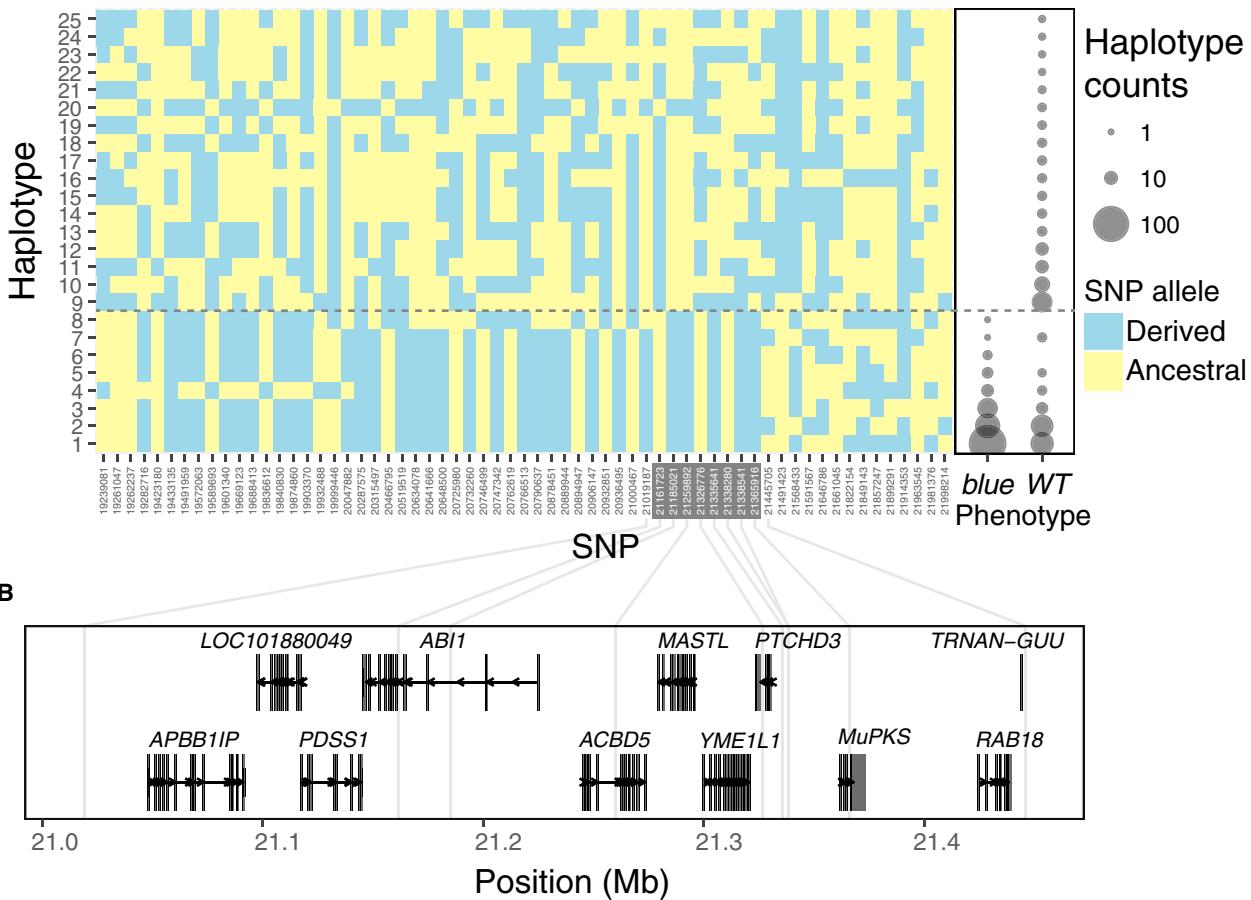
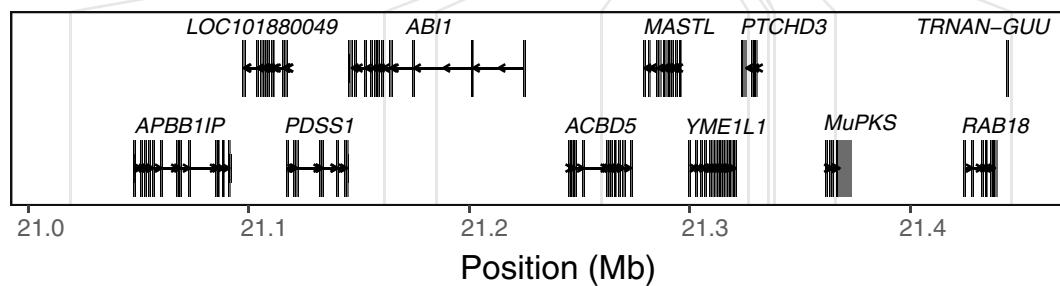
A**B**

Figure 3. Haplotypes Associated with the *blue* Phenotype

(A) Haplotypes for 63 SNPs in a 3 Mb region centered on the *blue* locus association peak were inferred with the software program PHASE. At a subset of these loci, outlined in gray, all *blue* individuals carried the same haplotype. Population-wide haplotype counts were calculated from the most likely pairs of haplotypes carried by each individual and their associated probabilities. Only haplotypes with ≥ 2 counts are shown, except haplotype 8, which was found in only one sample but shows evidence of an ancestral recombination between SNPs at 21,019,187 and 21,161,723. The ancestral alleles were determined by whole-genome sequence alignment to 14 other avian species (Green et al., 2014).

(B) RefSeq gene models and descriptions for genes located within the *blue*-shared haplotype, with positions of SNPs shown, including the two flanking SNPs.

data from chicken and crow feathers (Ng et al., 2014; Pöhlström et al., 2014). We found that budgerigar *MuPKS* was expressed hundreds to thousands of times higher than its homologs in chicken and crow, which do not exhibit psittacofulvin pigmentation (Figure S4).

Coding SNP in Conserved *MuPKS* Residue Completely Segregates with Pigmentation

Given that the mRNA-seq data suggested that the *blue* allele does not act through gene-expression changes, we searched

for non-synonymous variants in protein-coding regions within the 0.4 Mb *blue*-associated haplotype. The only protein-coding change in an expressed gene was an arginine (R) to tryptophan (W) substitution at residue 644 in *MuPKS* (Figure S5A). We Sanger sequenced this SNP for the full set of samples: all 162 WT birds had genotype R/R or R/W, whereas all 118 *blue* birds had genotype W/W (Figure 5A).

A multiple sequence alignment showed that R644 is conserved from vertebrates to bacteria (Figure 5B). Type I microbial polyketide synthases and vertebrate fatty-acid synthases

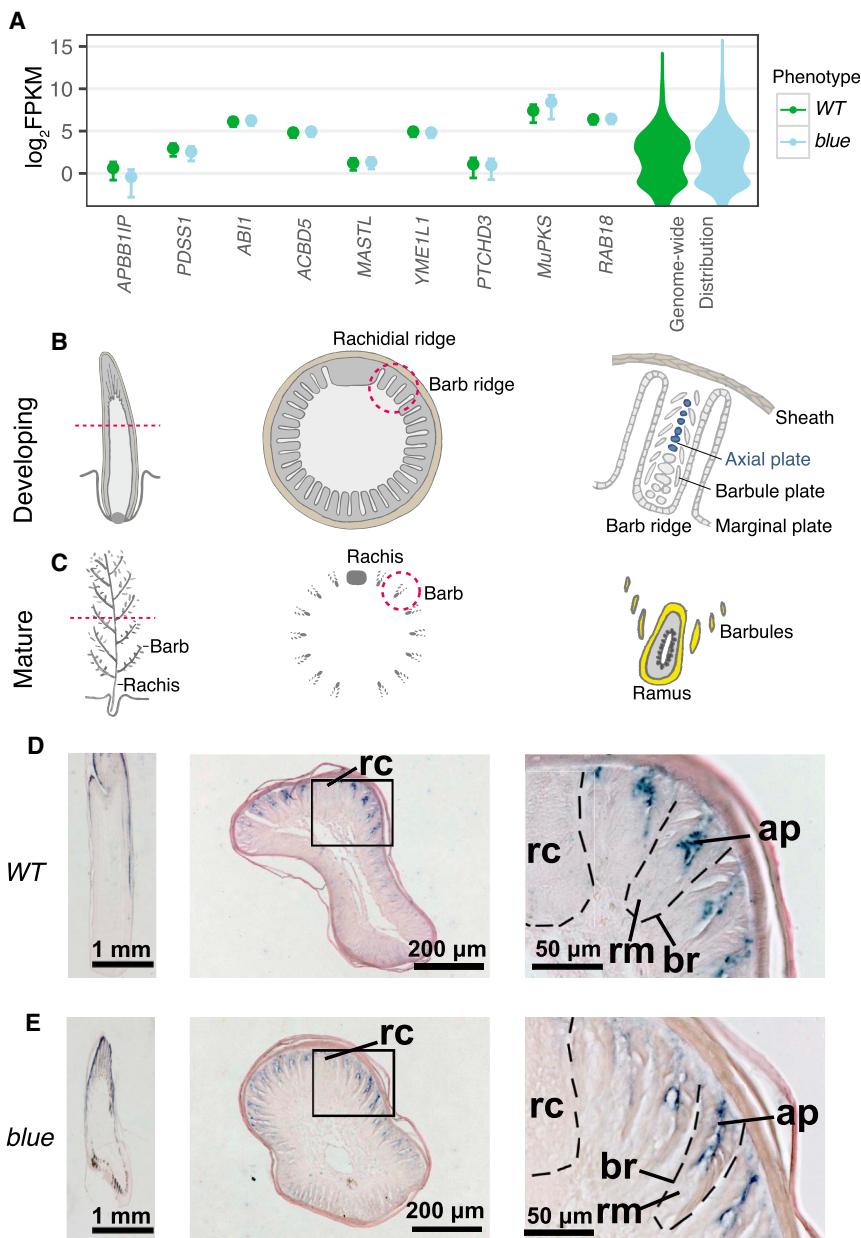


Figure 4. Expression of *blue* Locus Genes in Regenerating Budgerigar Feathers

(A) Transcript levels, measured as fragments per kilobase of transcript per million fragments mapped (FPKM), in regenerating contour feathers from WT ($n = 3$) or *blue* ($n = 4$) budgerigars for genes in the *blue*-associated haplotype (Figure 3B). *LOC101880049* and *TRNAN-GUU* were not expressed (FPKM < 0.04). Error bars represent 95% confidence intervals calculated by cuffdiff (Trapnell et al., 2013). Violin plots to the right show the genome-wide distribution of FPKM values per gene. None of the 9 genes showed significantly different expression levels in WT versus *blue* after applying a Benjamini-Hochberg correction for multiple hypothesis testing.

(B) Illustrated cross-sections of a regenerating feather and its follicle. The barb ridges and rachidial ridge, which give rise to the barbs and rachis, are wrapped in a cylindrical sheath (Chen et al., 2015). As the feather matures, the axial plate and marginal plate are lost by apoptosis, allowing the barbules and ramus to separate. The sheath then sloughs off, allowing the feather to open. The barb ridges closest to the rachis mature earlier than those opposite the rachis.

(C) Mature stage of the same feather parts shown in (B). Yellow psittacofulvin pigment is found in the barbules and in the outer cortex of the ramus in budgerigar feathers (D'Alba et al., 2012).

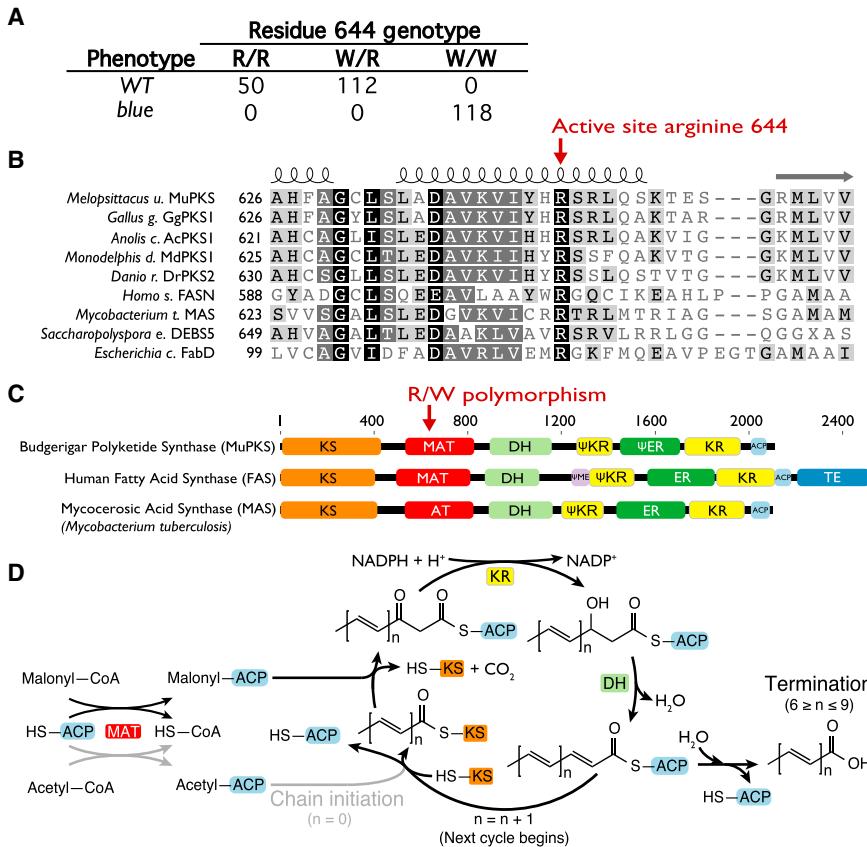
(D and E) *In situ* hybridization of regenerating contour feather follicles from a WT (D) or *blue* (E) budgerigar with probes against the uncharacterized polyketide synthase *MuPKS*. Transcripts are detected in the axial-plate epithelia of more mature barb ridges. Notations: ap, axial plate; br, barb ridge; rc, rachis; rm, ramus.

contain multiple domains with independent catalytic activities (Leibundgut et al., 2008). The R/W polymorphism resides in the malonyl-CoA:ACP transacylase (MAT) domain (Figure 5C). In fatty-acid synthase, the MAT domain primes the acyl-carrier protein (ACP) domain with malonyl-CoA during chain extension (Maier et al., 2008). Structural studies have shown that this conserved arginine forms a salt bridge with the C-3 carboxylate of malonyl-CoA in the active site (Maier et al., 2008; Oefner et al., 2006; Wong et al., 2011). Furthermore, mutating this arginine to alanine resulted in a 100-fold decrease in MAT domain activity of recombinant human fatty-acid synthase (Rangan and Smith, 1997), suggesting that the R644W substitution in MuPKS would likely diminish its activity.

Because the main chemical difference between fatty acids and psittacofulvins is the high degree of conjugation in the latter (Stradi et al., 2001), it is notable that MuPKS appears to lack a functional enoyl-reductase (ER), the domain that reduces double bonds. The canonical GGVGX_A NADPH-binding domain in the MuPKS ER is missing (Figure S5B), as it is in some fungal polyketide synthases known to produce psittacofulvin-like compounds (Ma et al., 2009). Therefore, we suggest that MuPKS synthesizes psittacofulvins by an iterative mechanism that is homologous to that of vertebrate fatty-acid synthase but does not include the ER-catalyzed reduction (Figure 5D).

Reconstitution of Feather Pigment Synthesis in Yeast

Genetic mapping, gene expression, sequence conservation, and structural and biochemical data from homologous enzymes all suggest that the causative variant for the *blue* trait is the R644W polymorphism in MuPKS, and that this enzyme is capable of synthesizing yellow psittacofulvin pigments. To test this hypothesis, we expressed MuPKS in *Saccharomyces*



thioester intermediate to the corresponding unsaturated α,β -unsaturated thioester. The inactive ER domain of MuPKS cannot reduce this double bond, resulting in a conjugated polyene product, such as those observed in the feather pigments of the scarlet macaw (Stradi et al., 2001).

cerevisiae (strain BJ5464-NpgA). This strain expresses a promiscuous phosphopantetheinyl transferase, NpgA, which is able to convert a wide range of polyketide synthase apoenzymes to the corresponding holoenzyme by covalent attachment of phosphopantetheine to a specific serine residue in the ACP domain (Ma et al., 2009). We verified the expression of MuPKS in yeast by western blot analysis (Figure 6A) and verified conversion to holoenzyme by mass spectrometry (Figure S6).

Organic extracts from yeast expressing the MuPKS WT allele are yellow (Figure 6A), whereas extracts from yeast expressing an empty vector control are clear (Figure 6A), indicating that MuPKS does synthesize yellow pigment. The MuPKS WT allele extracts also exhibit a striking fluorescence under UV-A (“black light”) illumination (Figure 6A). This fluorescence is notable because the yellow feathers of many parrot species also fluoresce under UV-A (Arnold et al., 2002; Hausmann et al., 2003), as does octadecaocetaenal purified from red parrot feathers (Adamec et al., 2016). We observed that blue budgerigars lack fluorescence (Figure 2B), suggesting that psittacofulvins are required (though perhaps not sufficient) for plumage fluorescence. Extracts from yeast expressing the MuPKS blue allele were clear (Figure 6A), indicating that MuPKS is the gene responsible for the blue trait in budgerigars. Extracts from yeast expressing the WT allele with an R644W substitution recreated

Figure 5. Candidate Causative Variant

(A) A non-synonymous SNP in budgerigar polyketide synthase segregates completely with the presence (WT) or absence (blue) of yellow feather pigment.

(B) The affected arginine residue is located within the malonyl-CoA:ACP transacylase (MAT) domain and is conserved across distantly related homologs, including human and bacterial fatty-acid synthase and bacterial polyketide synthases. In the crystal structure of *E. coli* FabD (Oefner et al., 2006), the conserved arginine forms a salt bridge with the malonate substrate in the active site.

(C) The domain structure of MuPKS is homologous to mammalian fatty-acid synthase and type I bacterial polyketide synthases. The enoyl-reductase (ER) domain is marked as inactive (Ψ) based on its lack of the canonical NADPH-binding motif. Additional notations: ketoacyl synthase, KS; dehydratase, DH; ketoreductase, KR; acyl-carrier protein, ACP; methyltransferase, ME; thioesterase, TE.

(D) Proposed biochemical mechanism of yellow psittacofulvin pigment synthesis in budgerigars. The ACP is activated by attachment of phosphopantetheine. Chain initiation involves transacylation of an acetyl-primer unit from acetyl-CoA to the active-site cysteine of the ketoacyl synthase (KS) domain. Each cycle of chain elongation begins with a condensation reaction between the KS-bound growing polyketide chain and malonyl-ACP. The MAT domain is responsible for transferring malonyl extender units from malonyl-CoA to the ACP. The ketoreductase (KR) and dehydratase (DH) domains convert each β -keto-

by site-directed mutagenesis were also clear (Figure 6A), indicating that the R644W is the causative change.

To confirm the pigment differences from yeast extracts, and to identify the constituents of the pigment mixture, we performed liquid chromatography electrospray ionization high-resolution mass spectrometry (LC-ESI-HRMS) on the yeast extracts and yellow pigments extracted from budgerigar feathers. Visible wavelength (374 nm) absorbance chromatograms from yeast expressing WT MuPKS displayed three major peaks with retention times that matched pigments extracted from feathers (Figure 6B). These peaks were absent in the extract from yeast expressing the MuPKS blue allele or MuPKS WT allele with the R644W substitution.

Next, we sought to match the spectral peaks in Figure 6B to ions detected by LC-MS. We used the program XCMS (Smith et al., 2006) to perform an untargeted search for ion signals specifically enriched in the pigmented samples (Figures S6C–S6E). This search revealed a family of low-intensity *m/z* values (217.1222, 243.1379, and 269.1535) whose retention times corresponded precisely with the peaks in the absorbance chromatogram (Figure 6C). The predicted molecular formulae for these three peaks ($C_{14}H_{16}O_2$, $C_{16}H_{18}O_2$, and $C_{18}H_{20}O_2$) are consistent with a series of conjugated fatty acids with the general structure shown in Figure 5D, which are presumably the biosynthetic precursors of the conjugated aldehydes previously identified in psittacofulvin

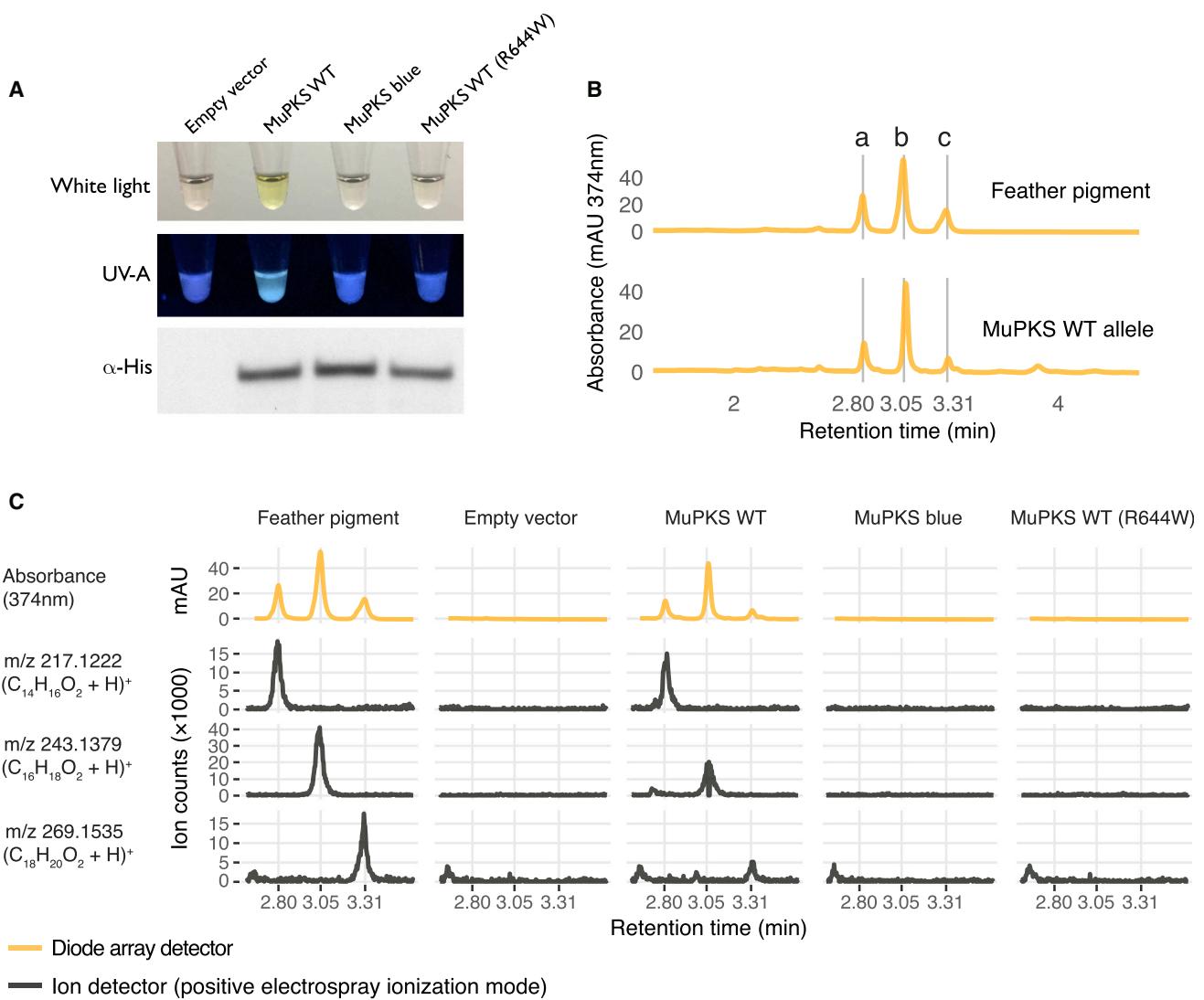


Figure 6. Reconstitution of Feather-Pigment Synthesis in Yeast

(A) Ethyl acetate extracts from yeast strain BJ5464-NpgA expressing 6 × His-tagged MuPKS WT, blue (defined by the amino acid at position 644 but not identical at all other positions), or WT with site-directed mutagenesis to create the R644W substitution. The extracts were illuminated with white light or UV-A (“black light”). MuPKS protein levels were measured in total soluble protein extracts from the same yeast cultures by western blot with α -His antibody.

(B) LC absorbance chromatograms (374 nm) for compounds extracted into methanol from yeast expressing WT MuPKS compared to budgerigar feather pigments extracted into acidified pyridine (McGraw and Nogare, 2005).

(C) Comparison of mass spectrometry data from compounds produced by MuPKS in yeast versus compounds extracted from yellow feathers. The absorbance chromatograms shown in (B) are aligned to extracted ion chromatograms for m/z values identified through an untargeted search for ions enriched in the pigmented samples (see also Figures S6C–S6E). The absorbance chromatogram has been shifted forward by 2.98 s, the delay time between the diode array detector and the ion detector in our setup. The most likely chemical formula for each ion is shown in parenthesis. These formulas are consistent with the family of polyenes shown in Figure 5D.

pigments from red parrot feathers (Adamec et al., 2016; Stradi et al., 2001). We conclude that MuPKS is responsible for synthesizing the highly unsaturated C_{14} , C_{16} , and C_{18} fatty-acyl precursors of the yellow psittacofulvin pigments found in budgerigar feathers, and that R644 is necessary for its biosynthetic activity.

Diverse Functions of Metazoan Polyketide Synthases

Although parrots are the only birds known to exhibit psittacofulvin feather pigmentation (McGraw, 2006a), homologs of

MuPKS are widespread among archosaurs (birds and crocodiles) and other vertebrate taxa such as snakes, lizards, and ray-finned fishes (Table S2). To help understand the function of these genes aside from their role in parrot feather pigmentation, and why they are conserved in some clades but largely absent in others, we searched for homologous sequences in metazoans and eukaryotic outgroups such as fungi and constructed a phylogeny based on a multiple sequence alignment of the KS domain (Figure 7).

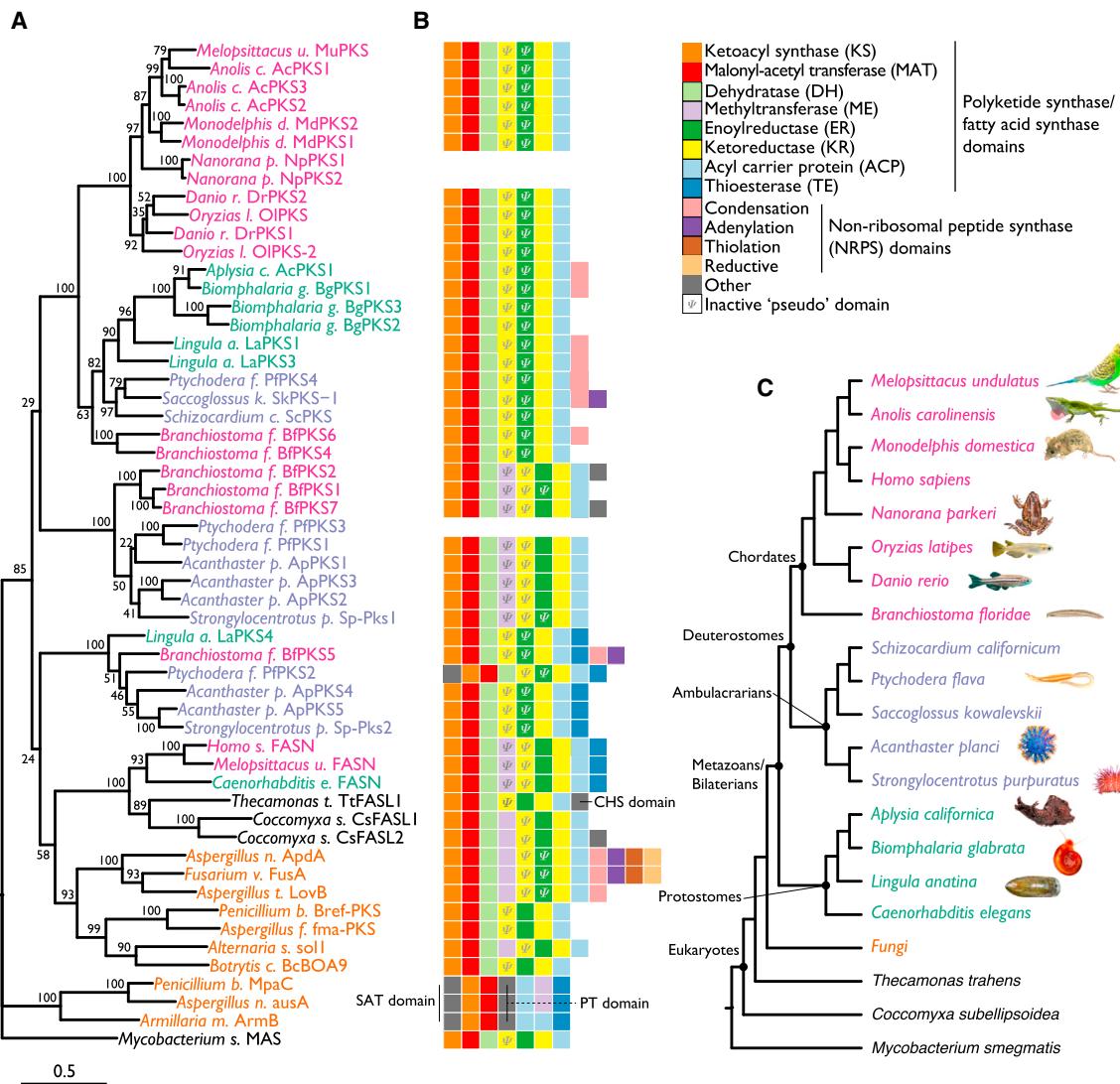


Figure 7. Phylogeny of Metazoan Polyketide Synthases

(A) Maximum likelihood tree based on an alignment of the KS domains of metazoan polyketide synthases, fatty-acid synthases, and their homologs in fungi, eukaryotic outgroups, and bacteria. Bootstrap values (based on 1,000 replicates) are indicated at the tree nodes. The scale bar below the tree denotes substitutions per site. Species are colored according to the clades shown in (C). The tree is rooted by the outgroup mycocerosic acid synthase (MAS) from *Mycobacterium*.

(B) Domain structures for the enzymes shown in (A). Colors denote domains commonly found in polyketide synthases, fatty-acid synthases, and non-ribosomal peptide synthases. Inactive pseudo-domains, or domains likely to be inactive based on sequence features, are denoted by “Ψ” (KS, MAT, and ACP make up the minimal set of domains for a functional polyketide synthase). Partial sequences or those containing probable artifacts from genome assembly errors are left blank.

(C) Cladogram for species shown in (A).

Vertebrate MuPKS homologs are monophyletic and share a common domain structure (Figure 7B) and synteny with neighboring genes (Figure S7). Other than MuPKS, the only gene in this family with a known function is OIPKS in medaka fish (*Oryzias latipes*), which is required for otolith (ear stone) formation (Hojo et al., 2015). The involvement of this MuPKS homolog in controlling a complex developmental phenotype suggests that the biosynthetic products of this family of enzymes have a variety of functions in animals in addition to pigmentation.

As a first step toward determining the degree of consensus in the function of vertebrate polyketide synthases, we cloned the uncharacterized gene *LOC420486* from chicken (referred to here as *GgPKS1*) and expressed it in yeast. Organic extracts from the yeast cells were yellow, and LC-MS revealed a set of pigment components similar to those produced by MuPKS (Figure S6C). This hints that the key evolutionary changes leading to psittacofulvin pigmentation in parrots were not variants in the ancestral PKS coding sequence but variants that affect its expression in feathers.

DISCUSSION

These results represent direct observation of the biochemical activity of a polyketide synthase in vertebrates and the first color trait in parrots mapped to a precise nucleotide difference. Given the chemical similarities between psittacofulvins and carotenoid pigments, and the relatively large number of cellular processes involved in the uptake, modification, transport, and accumulation of carotenoids in birds (McGraw, 2006b), we suspect that natural variation in psittacofulvins is multigenic. Although there are not many known color polymorphisms in wild populations of any single species of parrot (Mundy, 2006), there is a wide range of color variation between species. Previous biochemical work on polyketide synthases can guide the search for additional genes and DNA differences responsible for interspecific color variation in parrots. In addition, the presence of sequences with close homology to MuPKS across vertebrates and other groups raises questions of a broader evolutionary scope concerning the origin of these enzymes and their roles in specialized metabolism.

Genetic Architecture of Parrot Color Variation

Much of the current knowledge of color-trait genetics in animals comes from mutations and Mendelian disorders affecting melanin pigmentation in humans, flies, zebrafish, mice, and various domesticated species (Hubbard et al., 2010). In many cases, these results have been key to understanding the causes of color polymorphisms in wild populations of birds and other animals. A well-known example is melanism in bananaquit birds (*Coereba flaveola*), which is associated with a single amino acid change in the melanocortin-1-receptor (MC1R) (Theron et al., 2001) that was first observed in domesticated mice as a dominant coat-color allele. Partially dominant or recessive MC1R alleles have also been observed in wild populations, for example in light-colored beach mice (Hoekstra et al., 2006). Unlike these melanin-based traits, the psittacofulvin-based *blue* phenotype in parrots is not known to be polymorphic in natural populations, despite its appearance in several captive-bred species (Van den Abeele, 2016). Nevertheless, our identification of *MuPKS* as a necessary pigmentation gene in budgerigars may shed light on other color traits in parrots.

The hue of psittacofulvin-based colors in parrots varies from red to yellow. Red coloration, for example, differs among the overlapping populations of the crimson rosella species complex (*Platycercus elegans*), where color may play an important role in speciation (Joseph et al., 2008). Although the causative genes for this trait are unknown, a number of related traits exist in captive-bred parrots. For instance, the recessive *orange faced* allele in peach-faced lovebirds (*Agapornis roseicollis*) affects red but not yellow or orange psittacofulvin coloration (Van den Abeele, 2016). This could be due to an absorbance shift caused by the pigment binding to a protein, analogous to the interaction between opsin and the visual pigment 11-cis-retinal (Lin et al., 1998). Another possible cause is a chemical change in the pigment itself. Red coloration in canaries and finches is caused by C(4)-oxygenation of carotenoids by the cytochrome P450 enzyme CYP2J19 (Lopes et al., 2016; Mundy et al., 2016). This chemical modification increases the length of the

conjugated part of the carotenoid molecule, causing a red shift in its absorbance spectrum. Previous work demonstrated that psittacofulvins from red macaw feathers are fully conjugated aldehydes that vary in length but not in oxygenation level (Stradi et al., 2001). Carbon chain length was proportional to the wavelength of maximum absorbance (Stradi et al., 2001). Pigments from red feathers of various parrot species were previously shown to resolve into four major chromatographic peaks corresponding to chain lengths C₁₄, C₁₆, C₁₈, and C₂₀ (McGraw and Nogare, 2005). Our results indicate that yellow budgerigar feather pigments lack the C₂₀ component (Figures 6B and 6C) and are not aldehydes. Therefore, the visual difference between red and yellow parrot feathers is most likely due to presence or absence of the C₂₀ pigment component, a different oxidation state at the end of the polyene acyl chains, or both.

The absence of C₂₀ psittacofulvin in budgerigar feathers suggests that the MuPKS product-release step may be an important source of phenotypic variability. For example, the mammalian fatty-acid synthase produces primarily long-chain fatty acids (C₁₄, C₁₆, and C₁₈) that are released by its integrated thioesterase. However, in the mammary gland, a separate *trans*-acting thioesterase shifts the product distribution to C₈, C₁₀, and C₁₂ fatty acids (Libertini and Smith, 1978). Like many iterative polyketide synthases, MuPKS lacks an integrated thioesterase domain (Figure 7B), suggesting that a *trans*-acting partner enzyme is involved in product release. The fact that highly conjugated aldehydes are major components of psittacofulvin pigments from red feathers (Stradi et al., 2001) suggests that the products of MuPKS homologs in these feathers are reductively released (Du and Lou, 2010). Therefore we hypothesize that there may be a reductase expressed in some parrot feathers that converts the fatty-acyl pigment compounds (Figure 6C) to their corresponding aldehydes.

Origins and Evolution of Animal Polyketide Synthases

The similarity between the domain structures of polyketide synthases and the metazoan fatty-acid synthase (FAS) suggests that they are ancient paralogs (Leibundgut et al., 2008). We found high bootstrap support for the clade consisting of metazoan FAS and the FAS-like proteins from the protozoan *Thecamonas trahens* and the green alga *Coccomyxa subellipoidea* (Figure 7A), suggesting that the divergence between FAS and MuPKS predates the last common ancestor of animals. Additional sampling from a diverse range of eukaryotes may be required to resolve deeper branches of the tree.

The existing data on other metazoan polyketide synthases suggest that they have a variety of functions. In nematode worms, the modular polyketide synthase PKS-1 is involved in the production of compounds that extend larval survival under starvation conditions (Shou et al., 2016). In medaka fish, OIPKS is thought to synthesize an otolith nucleation factor, although this compound has not yet been identified (Hojo et al., 2015). In sea urchins, *Sp-Pks1* is required for production of the red pigment echinochrome-A, which also has bacteriocidal properties (Calestani et al., 2003; Service and Wardlaw, 1984). Another sea urchin polyketide synthase gene, *Sp-Pks2*, is required for the formation of calcium carbonate skeletons known as spicules

(Castoe et al., 2007; Hojo et al., 2015). Additional specialized functions are likely to be seen in metazoan polyketide synthases.

Enzymes involved in specialized metabolism typically exhibit low to modest catalytic efficiency relative to their counterparts in primary metabolism (Weng and Noel, 2012). For instance, whereas the apparent k_{cat} of chicken FAS is 23 s⁻¹ (Cox and Hammes, 1983), turnover in many fungal iterative polyketide synthases is orders of magnitude lower (Cacho et al., 2015; Zhang et al., 2008b). Although MuPKS is a specialized enzyme, psittacofulvin pigments accumulate at high levels in feathers (McGraw and Nogare, 2005). Such high accumulation suggests that MuPKS either has unusually high catalytic efficiency or overcomes low efficiency through high expression levels. It would be interesting to determine whether evolution has favored a particular mechanism.

The expression differences between *MuPKS* and its homologs in chicken and crow feathers (Figure S4) suggests that a gene-regulatory change was involved in the evolution of yellow pigmentation in parrots. There is evidence that recent gene duplicates, such as those observed in the polyketide synthase gene tree (Figure 7A), are often quickly lost when they exhibit functional redundancy (Lynch and Conery, 2000). Acquisition of a novel expression pattern is thought to be one way of eliminating functional redundancy (Ohno, 1970). There does not appear to be a consensus polyketide synthase expression pattern in any given tissue in birds, lizards, opossum, or zebrafish (Figure S4). Furthermore, the fact that budgerigar *blue* mutants and medaka *OIPKS* mutants are viable suggests that no such consensus expression pattern is required for survival. Therefore, it may be interesting to see whether additional derived expression patterns exist in other species.

An understanding of the genes and regulatory mechanisms that govern psittacofulvin pigmentation could add an important dimension to ecological and evolutionary studies of parrots. For example, a long-running field study on the reproductive biology of the burrowing parrot (*Cyanoliseus patagonus*) demonstrated assortative mating relative to the size of red psittacofulvin-based belly patches (Masello and Quillfeldt, 2003). The ability to combine such time-series data from behavioral and phenotypic observations in the field with observations of changes at the DNA level would be exciting and could aid studies of adaptation at the molecular level.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
 - Budgerigar sample collections
 - Opossum samples
 - Chicken sample
 - Yeast strain
- METHOD DETAILS
 - Erythrocyte Hi-C library construction
 - Physical map construction from Hi-C data
 - Isolation of genomic DNA

- Modified ddRAD sequencing
- ddRAD-seq SNV calling
- LD estimation
- Genome-wide association mapping
- Haplotype inference
- Feather mRNA-seq
- Feather section in situ hybridizations
- MuPKS expression in yeast
- Pigment extraction from feathers
- Pigment analysis by mass spectrometry
- Opossum *MuPKS* homologs
- Phylogenetic analysis
- Synteny analysis

● QUANTIFICATION AND STATISTICAL ANALYSIS

- Test for genotype-trait association
- Tests for differential gene expression

● DATA AND SOFTWARE AVAILABILITY

- Additional Resources

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and seven tables and can be found with this article online at <https://doi.org/10.1016/j.cell.2017.08.016>.

A video abstract is available at <https://doi.org/10.1016/j.cell.2017.08.016#mmc8>.

AUTHOR CONTRIBUTIONS

Conceptualization, T.F.C., K.T.X., and C.D.B.; Methodology, T.F.C., C.R.F., P.W., T.-X.J., K.T.X., J.K., A.Z., C.K., C.-M.C., and C.D.B.; Software, T.F.C.; Investigation, T.F.C., C.R.F., P.W., T.-X.J., A.Z., K.T.X., J.K., and E.D.; Writing-Original Draft, T.F.C.; Writing-Review & Editing, T.F.C., C.R.F., K.T.X., A.Z., C.K., and C.D.B.; Supervision, C.D.B.; Project Administration, T.F.C.; Funding Acquisition, T.F.C., C.K., and C.D.B.

ACKNOWLEDGMENTS

We thank the American Budgerigar Society (ABS), the Budgerigar Association of America (BAA), and their members for generous contributions of samples and for helpful discussions; Sally Nofs and the Potter Park Zoo of Lansing, Michigan for samples; Leah Tsang and Sandy Ingleby of the Australian Museum and Leo Joseph and Robert Palmer of the Australian National Wildlife Collection for museum specimens; Joanne Paul-Murphy and David Guzman at UC Davis Companion Avian and Pet Exotic Service for assistance with sample collection; Leah Krubitzer and Deepa Ramamurthy of UC Davis for opossum samples; Kotaro Fujii and Bradley French for chicken samples; members of the Tim Wright lab at NMSU for assistance with sampling; Erich Jarvis of Rockefeller University for helpful discussions about the budgerigar genome; Yi Tang of UCLA, Yanran Li, and Yen-Hsiang Wang for plasmids and the BJ5464-NpgA yeast strain; Elizabeth Sattely of Stanford for help in identifying mass spectrometry resources; members of the Vincent Coates Mass Spectrometry Laboratory, and in particular Ryan Leib and Chris Adams, for help in measuring MuPKS phosphopantetheinylation; Emily Crane for advice about Hi-C methods; Chris Lowe of Stanford for providing sequences from the *Schizocardium californicum* transcriptome assembly; members of the Bustamante lab for general discussions; and Shirley Sutton and Alexandra Sockell for assistance with Illumina sequencing. T.F.C. was supported by NIH training grants with award numbers T32HG000044 and T32GM007276. C.-M.C., P.W., T.-X.J. are supported by NIH R01 AR 47364 and AR 60306.

Received: March 27, 2017

Revised: July 14, 2017

Accepted: August 8, 2017

Published: October 5, 2017

REFERENCES

- Adamec, F., Greco, J.A., LaFountain, A.M., Magdaong, N.M., Fuciman, M., Birge, R.R., Polívka, T., and Frank, H.A. (2016). Spectroscopic investigation of a brightly colored psittacofulvin pigment from parrot feathers. *Chem. Phys. Lett.* 648, 195–199.
- Agate, R.J., Grisham, W., Wade, J., Mann, S., Wingfield, J., Schanen, C., Palotie, A., and Arnold, A.P. (2003). Neural, not gonadal, origin of brain sex differences in a gynandromorphic finch. *Proc. Natl. Acad. Sci. USA* 100, 4873–4878.
- Ardlie, K.G., Kruglyak, L., and Seielstad, M. (2002). Patterns of linkage disequilibrium in the human genome. *Nat. Rev. Genet.* 3, 299–309.
- Arnold, K.E., Owens, I.P.F., and Marshall, N.J. (2002). Fluorescent signaling in parrots. *Science* 295, 92.
- Auber, L. (1941). The Colours of Feathers and Their Structural Causes in Varieties of the Budgerigar, *Melopsittacus undulatus* (Shaw) (Edinburgh, Scotland: University of Edinburgh).
- Bartels, T., Cramer, K., Wolf, P., Hässig, M., and Boos, A. (2009). Osteological examinations on the budgerigar (*Melopsittacus undulatus* Shaw 1805) with special reference to skeletal alterations conditioned by breeding. *Anat. Histol. Embryol.* 38, 262–269.
- Burton, J.N., Adey, A., Patwardhan, R.P., Qiu, R., Kitzman, J.O., and Shendure, J. (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* 31, 1119–1125.
- Butt, E.H., Jr., Schroeder, M.R., Smith, L.A., Sroka, J.E., and McGraw, K.J. (2011). Colourful parrot feathers resist bacterial degradation. *Biol. Lett.* 7, 214–216.
- Cacho, R.A., Thuss, J., Xu, W., Sanchar, R., Gao, Z., Nguyen, A., Vederas, J.C., and Tang, Y. (2015). Understanding programming of fungal iterative polyketide synthases: The biochemical basis for regioselectivity by the methyltransferase domain in the lovastatin megasynthase. *J. Am. Chem. Soc.* 137, 15688–15691.
- Calestani, C., Rast, J.P., and Davidson, E.H. (2003). Isolation of pigment cell specific genes in the sea urchin embryo by differential macroarray screening. *Development* 130, 4587–4596.
- Castoe, T.A., Stephens, T., Noonan, B.P., and Calestani, C. (2007). A novel group of type I polyketide synthases (PKS) in animals and the complex phylogenomics of PKSs. *Gene* 392, 47–58.
- Chan, T.R., Hilgraf, R., Sharpless, K.B., and Fokin, V.V. (2004). Polytriazoles as copper(I)-stabilizing ligands in catalysis. *Org. Lett.* 6, 2853–2855.
- Chang, C.C., Chow, C.C., Tellier, L.C.A.M., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7.
- Chen, C.-F., Foley, J., Tang, P.-C., Li, A., Jiang, T.X., Wu, P., Widelitz, R.B., and Chuong, C.M. (2015). Development, regeneration, and evolution of feathers. *Annu. Rev. Anim. Biosci.* 3, 169–195.
- Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Ruden, D.M., and Lu, X. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnPEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6, 80–92.
- Cooke, T.F., Yee, M.-C., Muzzio, M., Sockell, A., Bell, R., Cornejo, O.E., Kelley, J.L., Bailliet, G., Bravi, C.M., Bustamante, C.D., and Kenny, E.E. (2016). GBStools: A statistical method for estimating allelic dropout in reduced representation sequencing data. *PLoS Genet.* 12, e1005631.
- Cox, B.G., and Hammes, G.G. (1983). Steady-state kinetic study of fatty acid synthase from chicken liver. *Proc. Natl. Acad. Sci. USA* 80, 4233–4237.
- Crew, F.A.E., and Lamy, R. (1935). Autosomal colour mosaics in the budgerigar. *J. Genet.* 30, 233–241.
- D'Alba, L., Kieffer, L., and Shawkey, M.D. (2012). Relative contributions of pigments and biophotonic nanostructures to natural color production: a case study in budgerigar (*Melopsittacus undulatus*) feathers. *J. Exp. Biol.* 215, 1272–1277.
- Du, L., and Lou, L. (2010). PKS and NRPS release mechanisms. *Nat. Prod. Rep.* 27, 255–278.
- Elias, J.E., and Gygi, S.P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* 4, 207–214.
- Fox, D.L. (1976). Animal Biochromes and Structural Colors (Berkeley and Los Angeles, California: University of California Press).
- Ganapathy, G., Howard, J.T., Ward, J.M., Li, J., Li, B., Li, Y., Xiong, Y., Zhang, Y., Zhou, S., Schwartz, D.C., et al. (2014). High-coverage sequencing and annotated assemblies of the budgerigar genome. *Gigascience* 3, 11.
- Gonzalez, P., Uhlinger, K.R., and Lowe, C.J. (2017). The adult body plan of indirect developing hemichordates develops by adding a Hox-patterned trunk to an anterior larval territory. *Curr. Biol.* 27, 87–95. Published online December 8, 2016.
- Green, R.E., Braun, E.L., Armstrong, J., Earl, D., Nguyen, N., Hickey, G., Vandewege, M.W., St John, J.A., Capella-Gutiérrez, S., Castoe, T.A., et al. (2014). Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. *Science* 346, 1254449.
- 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. *Syst. Biol.* 59, 307–321.
- Hausmann, F., Arnold, K.E., Marshall, N.J., and Owens, I.P.F. (2003). Ultraviolet signals in birds are special. *Proc. Biol. Sci.* 270, 61–67.
- Hickey, G., Paten, B., Earl, D., Zerbino, D., and Haussler, D. (2013). HAL: a hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics* 29, 1341–1342.
- Hoekstra, H.E., Hirschmann, R.J., Bundey, R.A., Insel, P.A., and Crossland, J.P. (2006). A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* 313, 101–104.
- Hojo, M., Omi, A., Hamanaka, G., Shindo, K., Shimada, A., Kondo, M., Narita, T., Kiyomoto, M., Katsuyama, Y., Ohnishi, Y., et al. (2015). Unexpected link between polyketide synthase and calcium carbonate biomineralization. *Zoological Lett.* 1, 3.
- 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 Genet.* 26, 231–239.
- Joseph, L., Dolman, G., Donnellan, S., Saint, K.M., Berg, M.L., and Bennett, A.T.D. (2008). Where and when does a ring start and end? Testing the ring-species hypothesis in a species complex of Australian parrots. *Proc. Biol. Sci.* 275, 2431–2440.
- Kalhor, R., Tjong, H., Jayathilaka, N., Alber, F., and Chen, L. (2011). Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nat. Biotechnol.* 30, 90–98.
- Keane, T.M., Creevey, C.J., Pentony, M.M., Naughton, T.J., and McInerney, J.O. (2006). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. *BMC Evol. Biol.* 6, 29.
- Keatinge-Clay, A.T. (2012). The structures of type I polyketide synthases. *Nat. Prod. Rep.* 29, 1050–1073.
- Kowalski, A., and Paliya, J. (2011). Chromatin compaction in terminally differentiated avian blood cells: the role of linker histone H5 and non-histone protein MENT. *Chromosome Res.* 19, 579–590.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.
- Leibundgut, M., Maier, T., Jenni, S., and Ban, N. (2008). The multienzyme architecture of eukaryotic fatty acid synthases. *Curr. Opin. Struct. Biol.* 18, 714–725.
- Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760.
- Li, A., Figueiroa, S., Jiang, T.-X., Wu, P., Widelitz, R., Nie, Q., and Chuong, C.-M. (2017). Diverse feather shape evolution enabled by coupling anisotropic signalling modules with self-organizing branching programme. *Nat. Commun.* 8, s14139.

- Li, Y., Chooi, Y.-H., Sheng, Y., Valentine, J.S., and Tang, Y. (2011). Comparative characterization of fungal anthracenone and naphthalenedione biosynthetic pathways reveals an α -hydroxylation-dependent Claisen-like cyclization catalyzed by a dimanganese thioesterase. *J. Am. Chem. Soc.* 133, 15773–15785.
- Libertini, L.J., and Smith, S. (1978). Purification and properties of a thioesterase from lactating rat mammary gland which modifies the product specificity of fatty acid synthetase. *J. Biol. Chem.* 253, 1393–1401.
- Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293.
- Lin, S.W., Kochendoerfer, G.G., Carroll, K.S., Wang, D., Mathies, R.A., and Sakmar, T.P. (1998). Mechanisms of spectral tuning in blue cone visual pigments. Visible and raman spectroscopy of blue-shifted rhodopsin mutants. *J. Biol. Chem.* 273, 24583–24591.
- Lopes, R.J., Johnson, J.D., Toomey, M.B., Ferreira, M.S., Araujo, P.M., Melo-Ferreira, J., Andersson, L., Hill, G.E., Corbo, J.C., and Carneiro, M. (2016). Genetic basis for red coloration in birds. *Curr. Biol.* 26, 1427–1434.
- Lynch, M., and Conery, J.S. (2000). The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155.
- Ma, S.M., Li, J.W.-H., Choi, J.W., Zhou, H., Lee, K.K.M., Moorthie, V.A., Xie, X., Kealey, J.T., Da Silva, N.A., Vederas, J.C., and Tang, Y. (2009). Complete reconstitution of a highly reducing iterative polyketide synthase. *Science* 326, 589–592.
- Maier, T., Leibundgut, M., and Ban, N. (2008). The crystal structure of a mammalian fatty acid synthase. *Science* 321, 1315–1322.
- Masello, J.F., and Quillfeldt, P. (2003). Body size, body condition and ornamental feathers of Burrowing Parrots: variation between years and sexes, assortative mating and influences on breeding success. *Emu* 103, 149–161.
- McGraw, K.J. (2006a). The mechanics of uncommon colors: Pterins, porphyrins, and psittacofulvins. In *Bird Coloration, Volume 1: Mechanisms and Measurements*, G.E. Hill and K.J. McGraw, et al., eds. (Cambridge, MA: Harvard University Press), pp. 354–388.
- McGraw, K.J. (2006b). Mechanics of carotenoid-based coloration. In *Bird Coloration, Volume 1: Mechanisms and Measurements*, G.E. Hill and K.J. McGraw, et al., eds. (Cambridge, MA: Harvard University Press), pp. 177–242.
- McGraw, K.J., and Nogare, M.C. (2005). Distribution of unique red feather pigments in parrots. *Biol. Lett.* 1, 38–43.
- Mundy, N.I. (2006). The genetic basis of color variation in wild birds. In *Bird Coloration, Volume 1: Mechanisms and Measurements*, G.E. Hill and K.J. McGraw, et al., eds. (Cambridge, MA: Harvard University Press), pp. 469–506.
- Mundy, N.I., Stapley, J., Bennison, C., Tucker, R., Twyman, H., Kim, K.-W., Burke, T., Birkhead, T.R., Andersson, S., and Slate, J. (2016). Red carotenoid coloration in the zebra finch is controlled by a cytochrome P450 gene cluster. *Curr. Biol.* 26, 1435–1440.
- Nanda, I., Karl, E., Griffin, D.K., Schartl, M., and Schmid, M. (2007). Chromosome repatterning in three representative parrots (Psittaciformes) inferred from comparative chromosome painting. *Cytogenet. Genome Res.* 117, 43–53.
- Ng, C.S., Wu, P., Foley, J., Foley, A., McDonald, M.-L., Juan, W.-T., Huang, C.-J., Lai, Y.-T., Lo, W.-S., Chen, C.-F., et al. (2012). The chicken frizzle feather is due to an α -keratin (KRT75) mutation that causes a defective rachis. *PLoS Genet.* 8, e1002748.
- Ng, C.S., Wu, P., Fan, W.-L., Yan, J., Chen, C.-K., Lai, Y.-T., Wu, S.-M., Mao, C.-T., Chen, J.-J., Lu, M.-Y.J., et al. (2014). Genomic organization, transcriptomic analysis, and functional characterization of avian α - and β -keratins in diverse feather forms. *Genome Biol. Evol.* 6, 2258–2273.
- Oefner, C., Schulz, H., D'Arcy, A., and Dale, G.E. (2006). Mapping the active site of *Escherichia coli* malonyl-CoA-acyl carrier protein transacylase (FabD) by protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 62, 613–618.
- Ohno, S. (1970). *Evolution by Gene Duplication* (Berlin: Springer).
- Paten, B., Earl, D., Nguyen, N., Diekhans, M., Zerbino, D., and Haussler, D. (2011). Cactus: Algorithms for genome multiple sequence alignment. *Genome Res.* 21, 1512–1528.
- Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., and 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, e37135.
- Poelstra, J.W., Vijay, N., Bossu, C.M., Lantz, H., Ryll, B., Müller, I., Baglione, V., Unneberg, P., Wikelski, M., Grabherr, M.G., and Wolf, J.B.W. (2014). The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* 344, 1410–1414.
- Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842.
- R Core Team (2015). R: A language and environment for statistical computing (Vienna, Austria: R Foundation for Statistical Computing). <http://www.R-project.org/>.
- Rangan, V.S., and Smith, S. (1997). Alteration of the substrate specificity of the malonyl-CoA/acetyl-CoA:acyl carrier protein S-acyltransferase domain of the multifunctional fatty acid synthase by mutation of a single arginine residue. *J. Biol. Chem.* 272, 11975–11978.
- Service, M., and Wardlaw, A.C. (1984). Echinochrome-A as a bactericidal substance in the coelomic fluid of *Echinus esculentus* (L.). *Comp. Biochem. Physiol.* 79, 161–165.
- Shapiro, M.D., Kronenberg, Z., Li, C., Domyan, E.T., Pan, H., Campbell, M., Tan, H., Huff, C.D., Hu, H., Vickrey, A.I., et al. (2013). Genomic diversity and evolution of the head crest in the rock pigeon. *Science* 339, 1063–1067.
- Shou, Q., Feng, L., Long, Y., Han, J., Nunnery, J.K., Powell, D.H., and Butcher, R.A. (2016). A hybrid polyketide-nonribosomal peptide in nematodes that promotes larval survival. *Nat. Chem. Biol.* 12, 770–772.
- Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., and Siuzdak, G. (2006). XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* 78, 779–787.
- Steiner, H. (1932). Vererbungsstudien am Wellensittich *Melopsittacus undulatus* (Shaw) (Zurich, Switzerland: University of Zurich).
- Stephens, M., Smith, N.J., and Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* 68, 978–989.
- Stoddard, M.C., and Prum, R.O. (2011). How colorful are birds? Evolution of the avian plumage color gamut. *Behav. Ecol.* 22, 1042–1052.
- Stradi, R., Pini, E., and Celentano, G. (2001). The chemical structure of the pigments in *Ara macao* plumage. *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* 130, 57–63.
- Taylor, T.G., and Warner, C. (1986). Genetics for Budgerigar Breeders (Northampton, UK: The Budgerigar Society).
- Theron, E., Hawkins, K., Birmingham, E., Ricklefs, R.E., and Mundy, N.I. (2001). The molecular basis of an avian plumage polymorphism in the wild: a melanocortin-1-receptor point mutation is perfectly associated with the melanic plumage morph of the bananaquit, *Coereba flaveola*. *Curr. Biol.* 11, 550–557.
- Trapnell, C., Hendrickson, D.G., Sauvageau, M., Goff, L., Rinn, J.L., and Pachter, L. (2013). Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat. Biotechnol.* 31, 46–53.
- Van den Abeele, D. (2016). Lovebirds Compendium (Oostvoorne, Netherlands: Welzo).
- Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., et al. (2013). From fastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* 43, 11.10.1–11.10.33.
- Weng, J.-K., and Noel, J.P. (2012). The remarkable pliability and promiscuity of specialized metabolism. *Cold Spring Harb. Symp. Quant. Biol.* 77, 309–320.

- Wong, F.T., Jin, X., Mathews, I.I., Cane, D.E., and Khosla, C. (2011). Structure and mechanism of the trans-acting acyltransferase from the disorazole synthase. *Biochemistry* 50, 6539–6548.
- Zabala, A.O., Chooi, Y.-H., Choi, M.S., Lin, H.-C., and Tang, Y. (2014). Fungal polyketide synthase product chain-length control by partnering thiohydrolase. *ACS Chem. Biol.* 9, 1576–1586.
- Zhang, J., Van Lanen, S.G., Ju, J., Liu, W., Dorrestein, P.C., Li, W., Kelleher, N.L., and Shen, B. (2008a). A phosphopantetheinylating polyketide synthase producing a linear polyene to initiate enediyne antitumor antibiotic biosynthesis. *Proc. Natl. Acad. Sci. USA* 105, 1460–1465.
- Zhang, W., Li, Y., and Tang, Y. (2008b). Engineered biosynthesis of bacterial aromatic polyketides in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 105, 20683–20688.
- Zhao, D., McBride, D., Nandi, S., McQueen, H.A., McGrew, M.J., Hocking, P.M., Lewis, P.D., Sang, H.M., and Clinton, M. (2010). Somatic sex identity is cell autonomous in the chicken. *Nature* 464, 237–242.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Goat anti-Mouse IgG (H+L) Secondary Antibody, HRP conjugate	Pierce	Cat# 31430; RRID: AB_228307
6x-His Epitope Tag Antibody (HIS.H8)	Pierce	Cat# MA121315; RRID: AB_557403
Chemicals, Peptides, and Recombinant Proteins		
2',3'-Dideoxythymidine-5'-Triphosphate (ddTTP), 10 mM Sodium Salt Solution	Affymetrix	Cat# 77116 1 UM
Agencourt AMPure XP	Beckman Coulter	Cat# A63882
20% Paraformaldehyde aqueous solution (methanol-free)	Electron Microscopy Sciences	Cat# 50980492
Peptone	Fisher Scientific	Cat# DF0118170
Yeast Nitrogen Base without amino acids	Fisher Scientific	Cat# DF0919153
Yeast Extract	Fisher Scientific	Cat# DF0127179
Agar	Fisher Scientific	Cat# DF0140010
5-Ethynyl-dUTP (5-EdUTP)	Jena Bioscience	Cat# CLK-T07-S
Dynabeads MyOne Streptavidin T1	Life Technologies	Cat# 65601
NuPAGE Transfer Buffer (20X)	Life Technologies	Cat# NP00061
NuPAGE Antioxidant	Life Technologies	Cat# NP0005
Dynabeads MyOne Streptavidin C1	Life Technologies	Cat# 65002
PEG4 carboxamide-6-azidohexanyl biotin	Life Technologies	Cat# B10184
TRIzol reagent	Life Technologies	Cat# 15596026
SuperScript II Reverse Transcriptase	Life Technologies	Cat# 18064014
NuPAGE 3-8% Tris-Acetate Protein Gels, 1.0 mm	Life Technologies	Cat# EA0375BOX
MP FastPrep Lysing Matrix D, 2 ml Tube	MP Biomedicals	Cat# 116913050
BsrGI-HF	NEB	Cat# R3575L
T4 DNA Ligase (2,000,000 units/ml)	NEB	Cat# M0202M
BSA-Molecular Biology Grade (20 mg/ml)	NEB	Cat# B9000S
T4 DNA Polymerase	NEB	Cat# M0203S
T4 DNA Ligase	NEB	Cat# M0202M
BgIII	NEB	Cat# R0144S
Ddel	NEB	Cat# R0175L
Klenow Fragment (3' → 5' exo-)	NEB	Cat# M0212L
NEBNEXT High-Fidelity 2X PCR Master Mix	NEB	Cat# M0541L
Q5 High Fidelity 2X Master Mix	NEB	Cat# M0492S
Glu-C, Sequencing Grade	Promega	Cat# V1651
Asp-N, Sequencing Grade	Promega	Cat# V1621
ProteaseMAX Surfactant, Trypsin Enhancer	Promega	Cat# V2071
Ni-NTA agarose	QIAGEN	Cat# 30210
pDrive cloning vector	QIAGEN	N/A
BluePippin Gel Cassette, 2% Agarose, dye-free, with internal standards	Sage Science	Cat# BDF2010
TBTA	Sigma	Cat# 678937-50MG
Tert-butanol	Sigma	Cat# 471712-100ML

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Phenol:Chloroform:Isoamyl Alcohol 25:24:1, Saturated with 10 mM Tris, pH 8.0, 1mM EDTA	Sigma	Cat# P3803-100ML
Copper(I) Bromide, 99.999% trace metals basis	Sigma-Aldrich	Cat# 254185-10G
RNAlater	Sigma-Aldrich	Cat# R0901-500ML
Zymolyase-100T	Sunrise Science Products	Cat# N0766555
SC-Ura Powder	Sunrise Science Products	Cat# 1306-030
CSM-Ura Powder	Sunrise Science Products	Cat# 1004-010
PVDF Transfer Membrane, 0.45 m, 26.5 cm x 3.75 m	Thermo Fisher	Cat# 88518
Slide-A-Lyzer Dialysis Cassettes, 20K MWCO, 3 mL	Thermo Fisher	Cat# 66003
Glycogen	Thermo Fisher	Cat# 10814010
Halt Protease Inhibitor Cocktail (100X)	Thermo Scientific	Cat# 78437
EZ-Link Iodoacetyl-PEG2-Biotin	Thermo Scientific	Cat# 21334
Proteinase K (recombinant), PCR grade	Thermo Scientific	Cat# FEREO0492
PageRuler Unstained High Range Protein Ladder	Thermo Scientific	Cat# 26637
Critical Commercial Assays		
Kapa Hyper Prep kit	Kapa Biosystems	Cat# KK8505
NextSeq 500 High Output v2 kit (150 cycles)	Illumina	Cat# FC-404-2002
NextSeq 500 Mid Output v2 kit (150 cycles)	Illumina	Cat# FC-404-2001
Qubit dsDNA HS Assay kit	Thermo Fisher	Cat# Q32854
LabChip XT DNA 750 Assay Kit	PerkinElmer	Cat# 760541
ECL Western Blotting Substrate	Pierce	Cat# 32209
GeneRacer Kit with SuperScript III RT and TOPO TA Cloning Kit for Sequencing	Life Technologies	Cat# L150201
DIG RNA Labeling Kit (SP6/T7)	Roche	Cat# 11175025910
TruSeq Stranded mRNA LT-set A kit	Illumina	Cat# RS-122-2101
Turbo DNA-free kit	Life Technologies	Cat# AM1907
Deposited Data		
Whole-genome alignment of 15 avian species from Green et al., 2014	ftp://crocgengomes.org/pub/ICGWG/Supplementary_Materials/main_page/Ancestral_genome_reconstruction_files/crocPaper_reestimated.hal	N/A
Previously published RNA-seq data	See Table S5	See Table S5
Mendeley Dataset	https://doi.org/10.17632/4nxpx8s xv5.1	https://doi.org/10.17632/4nxpx8s xv5.1
Budgerigar ddRAD-seq and RNA-seq	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA378643/	SRA Project# PRJNA378643
Budgerigar Hi-C	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA378785/	SRA Project# PRJNA378785
Experimental Models: Organisms/Strains		
<i>Saccharomyces cerevisiae</i> : BJ5464-NpgA	Ma et al., 2009	N/A
Recombinant DNA		
pDrive	QIAGEN	N/A
pTFC24b (empty vector)	This paper	N/A
pTFC26b (MuPKS WT)	This paper	N/A
pTFC27b (MuPKS blue)	This paper	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
pTFC28b (MuPKS R644W)	This paper	N/A
pTFC30b (GgPKS1)	This paper	N/A
Software and Algorithms		
PhyML v3.2.0	Guindon et al., 2010	https://github.com/stephaneguindon/phym/commit/423ab928e1d7ac9864b81ee164dbd6c793051ef4
ModelGenerator v0.85	Keane et al., 2006	http://mcinerneylab.com/software/modelgenerator
Bowtie2 v2.2.9	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Picard v2.6.0	Broad Institute	https://broadinstitute.github.io/picard/
Lachesis (no version number available; downloaded Oct 2016)	Burton et al., 2013	https://github.com/shendurelab/LACHESIS
BWA v0.7.9	Li and Durbin, 2009	http://bio-bwa.sourceforge.net
GATK v3.3.0	Van der Auwera et al., 2013	https://software.broadinstitute.org/gatk/
PLINK v1.9.0	Chang et al., 2015	https://www.cog-genomics.org/plink2
PHASE v2.1.1	Stephens et al., 2001	http://stephenslab.uchicago.edu/phase/download.html
progressiveCactus	Paten et al., 2011	https://github.com/glennhickey/progressiveCactus/commit/91d6344a6b60fb170d59229f7881146c089536b1
JCVI v0.6.9	https://doi.org/10.5281/zenodo.31631	https://github.com/tanghaibao/jcvi
gbstools v0.1.0	Cooke et al., 2016	https://github.com/cooketho/gbstools
R v3.2.3	R Core Team, 2015	https://www.r-project.org
HAL tools v2.1	Hickey et al., 2013	https://github.com/ComparativeGenomicsToolkit/hal/blob/master/README.md
bedtools v2.25.0	Quinlan and Hall, 2010	http://bedtools.readthedocs.io/en/latest/
FASTX-Toolkit	Gregory Hannon	http://hannonlab.cshl.edu/fastx_toolkit/
snpEff v4.3i	Cingolani et al., 2012	http://snpeff.sourceforge.net
ProteinMetrics Preview v2.6.49	Protein Metrics Inc.	http://www.proteinmetrics.com/products/preview/
ProteinMetrics Byonic v2.6.49	Protein Metrics Inc.	http://www.proteinmetrics.com/products/byonic/
XCMS	Smith et al., 2006	https://bioconductor.org/packages/release/bioc/html/xcms.html
Custom python and R scripts	This paper	https://github.com/cooketho/budgies
HALplot	This paper	https://github.com/cooketho/halplot

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Carlos D. Bustamante (cdbustam@stanford.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS**Budgerigar sample collections**

Whole blood from a clipped toe nail and/or up to three regenerating contour feathers (i.e., non-flight feathers) were collected from healthy budgerigars owned by breeders, zoos, or universities (n = 286, see also **Table S1**). Of these individuals, 183 were male, 102 were female, and 1 was of unknown sex. The median age was 1 year. A signed written consent was obtained from each owner, and the collection methods were approved and regulated by the Stanford Institutional Animal Care and Use Committee (IACUC). Individuals were phenotyped for presence or absence of yellow pigment (i.e., for the *blue* trait) by visual inspection at the time of sample collection. Specimens from the Australian Museum (muscle, liver, or lung tissues) were shipped in 20% DMSO, 250 mM EDTA, and

saturated NaCl. Specimens from the Australian National Wildlife Collection were shipped in 70% ethanol. Permits for the transfer of the samples were obtained from the Australian Government, and from the United States Department of Agriculture.

Opossum samples

Monodelphis domestica testis samples were a gift of Leah Krubitzer of the University of California, Davis. The samples were collected after routine culling (approved by the UC Davis IACUC) that was unrelated to this study. They are therefore exempt from Stanford IACUC regulation. The samples were stored in 70% ethanol.

Chicken sample

A *Gallus gallus* embryo (day two stage) was a gift of Kotaro Fujii and Bradley French of Stanford University. The sample was stored in TriPure reagent.

Yeast strain

Saccharomyces cerevisiae strain BJ5464-NpgA was a gift of Yi Tang of the University of California, Los Angeles. Its construction from strain BJ5464 (*MAT α* , *ura352*, *trp1*, *leu2Δ1*, *his3Δ200*, *pep4::HIS3*, *prb1Δ1.6R*, *can1*, *GAL*) by chromosomal integration of an ADH2p-npgA-ADH2t cassette has been described previously (Ma et al., 2009). The strain was grown on YPD plates or liquid media at 30°C, or on CSM dropout media when selecting for transformants.

METHOD DETAILS

Erythrocyte Hi-C library construction

Sample collection, DNA crosslinking, and digestion

Approximately 30 µl of whole blood was collected and stored in 0.6 ml wash buffer (5 mM HEPES pH 7.5, 150 mM NaCl, with 30 U/ml heparin) and stored at 4°C overnight. To remove the heparin, the cells were centrifuged (5 min at 500 × g at 4°C), the supernate was discarded, and the cells resuspending in 0.5 ml wash buffer. This process was repeated for a total of 3 washes. The cells were then resuspended in 500 µl lysis buffer (10 mM HEPES pH 7.5, 10 mM NaCl, 0.2% IGEPAL CA-630, 1x protease inhibitor cocktail (HALT, Thermo Scientific)), and incubated 5 min at room temperature to lyse the cells. The DNA was then crosslinked by adding 13.9 µl 37% formaldehyde (stabilized with 10%–15% methanol) for a final concentration of 1%, and incubating 60 s at room temperature before quenching by adding 57.1 µl 2.5 M glycine (final concentration 250 mM). The nuclei were washed by spinning down for 5 min at 200 × g at 4°C, discarding the supernate, and gently resuspending in 0.5 ml of ice-cold 10 mM Tris-Cl pH 8.0. This was repeated twice, and the nuclei were resuspended in 100 µl of the same buffer after the third spin. The nuclei density was measured by hemocytometer, and the sample diluted to 2 × 10⁸ nuclei/ml with 10 mM Tris-Cl pH 8.0. The nuclei were disrupted by mixing 100 µl of the nuclei suspension with 11 µl 10% SDS (final concentration of 1%) and incubating 10 min at 65°C on a mixer at 1000 RPM. The following components, each at room temperature (except the restriction enzyme), were then added in order, with gentle mixing after each addition: 659 µl water, 100 µl 10x NEB CutSmart, 100 µl 10% Triton X-100, 30 µl BsrGI-HF (20 U/µl, NEB), for a final volume of 1000 µl. The sample was split into two 0.5 ml aliquots and incubated overnight at 37°C on a mixer (1000 RPM).

Chromatin tethering, labeling, and ligation

To biotinylate cysteine residues of the crosslinked chromatin, 125 µl of 25 mM iodoacetyl-PEG2-biotin (IPB) (Thermo Scientific) was added to each 0.5 ml aliquot and mixed by rotation for 1 hr at room temperature. To remove excess IPB, the sample was dialyzed in a 20 kDa dialysis cassette (Slide-A-Lyzer, Thermo Scientific) at 4°C for 2 hr in 1 l of 10 mM Tris pH 8.0 with 1 mM EDTA, after which the dialysis buffer was replaced and the sample dialyzed overnight. MyOne Streptavidin T1 magnetic beads (500 µl of a 10 mg/ml suspension, Thermo Scientific) were washed in phosphate-buffered saline pH 7.4 with 0.01% Tween-20 (PBST) and resuspended in 1 ml of the same buffer. The dialyzed chromatin (approximately 1.3 ml) was split into two 650 µl aliquots and each was mixed with 500 µl of the washed bead suspension and rocked 30 min at room temperature. During this incubation 150 µl of 25 mM IPB was quenched with 1 µl of 2-mercaptoethanol. Following the incubation, 5 µl of quenched IPB was added to each chromatin aliquot and rocked 15 min at room temperature to bind any remaining free streptavidin on the beads. The beads were separated from the liquid with a magnet and half of the supernate was removed. The beads were then resuspended and combined into a single tube. The beads were separated again, the supernate removed, and the beads washed with 1x NEBuffer 2, 100 µg/ml BSA, and 1% Triton X-100 (the latter for sequestering residual SDS). This wash step was repeated once, followed by two washes with 1x NEBuffer 2 with 100 µg/ml BSA. The beads were then resuspended in 60 µl of the latter buffer. The 3' T of the BsrGI-digested DNA was replaced by 5-ethynyl uridine (to be conjugated to biotin in the click chemistry step later) by adding 29 µl water, 4 µl 10x NEBuffer 2, 4 µl 100 mg/ml BSA, 2.5 µl 10 mM 5-ethynyl-dUTP (5-EdUTP, Jena Bioscience), and 0.5 µl T4 DNA polymerase (3 U/µl, NEB) and incubating for 5 min at 12°C. The reaction was stopped by adding 40 µl 60 mM EDTA and cooling on ice. The beads were separated with a magnet and washed twice with 10 mM Tris pH 8.0 with 20 mM EDTA and 0.01% Tween-20, followed by two washes with 1x T4 DNA ligase buffer (NEB) with 100 µg/ml BSA. The beads were then resuspended in 1 ml of the latter buffer and mixed with 43.7 ml water, 5 ml 10x T4 DNA ligase buffer (NEB), 250 µl 20 mg/ml BSA (NEB), and 25 µl T4 DNA Ligase (2000 U/µl, NEB) in a 50 ml tube and mixed briefly by inversion. To limit contact between the beads (and hence unwanted intermolecular ligations), mixing was kept to a minimum and the beads were poured into 6 sterile square petri dishes (9 × 9 cm), covered, and allowed to settle. The dishes were incubated at room

temperature for 2 hr, and the ligation reaction was stopped by adding EDTA to a final concentration of 20 mM. The beads were re-suspended briefly at low speed on an orbital shaker. Addition of Tween-20 to a final concentration of 0.01% helped reduce adhesion of the beads to the plates during resuspension. The beads were transferred to a 50 ml tube and separated with a magnet, the supernate removed, and the beads resuspended in 1 ml of 1x NEBuffer 2 with 100 mg/ml BSA and 10 mM EDTA. The beads were then transferred to a 1.5 ml tube and washed once with the same buffer, and twice with 1x NEBuffer 2 with 100 mg/ml BSA. To remove 5-EdU from unligated 3' ends, the beads were resuspended in 60 μ l of the same buffer and mixed with 26.5 μ l water, 4 μ l of 10x NEBuffer 2, 4 μ l of 100 mg/ml BSA, 2.5 μ l of 10 mM dATP, 2.5 μ l of 10 mM dCTP, and 0.5 μ l of T4 DNA polymerase (3 U/ μ l, NEB), and incubated 15 min at 12°C. The reaction was stopped by adding 4 μ l of 0.5 M EDTA and cooling on ice. The beads were separated on a magnet and washed two times with 10 mM Tris pH 8.0 with 10 mM EDTA and 0.01% Tween-20, and twice with the same buffer without EDTA. They were then resuspended in 10 μ l of 10 mM Tris pH 8.0.

Biotinylation of ligation junctions

Ligation junctions containing 5-EdU were labeled with biotin azide by Cu(I)-catalyzed azide-alkyne cycloaddition (CuAAC) click chemistry (Chan et al., 2004). A solution of tris(benzyltriazoylmethyl)amine (TBTA) complexed with Cu(I) was prepared as follows. Approximately 10 mg of CuBr was weighed out and mixed gently into tert-butyl alcohol/DMSO (1:3) to make a 100 mM solution. Then, working quickly to minimize oxidation of Cu(I), 10 μ l of 100 mM CuBr solution was combined with 20 μ l of 100 mM TBTA in tert-butyl alcohol/DMSO (1:3) and 70 μ l tert-butyl alcohol/DMSO (1:3) for a final volume of 100 μ l. The reaction was carried out by mixing the bead-immobilized chromatin suspension (in 10 μ l Tris pH 8.0) with 10 μ l of 1 mM PEG4 carboxamide-6-azidohexanyl biotin (Life Technologies) in DMSO, and 10 μ l of Cu-TBTA solution, and incubating 2 hr at 37°C on a mixer at 1000 RPM.

DNA extraction

The biotinylated chromatin was mixed with 600 μ l extraction buffer (50 mM Tris pH 8.0, 10 mM EDTA, 100 mM NaCl, 0.2% SDS) and 20 μ l 20 mg/ml proteinase K and incubated overnight at 65°C on a mixer at 1000 RPM. An equal volume of phenol:chloroform:isoamyl alcohol (25:24:1, saturated with 10 mM Tris pH 8.0, 1 mM EDTA) was added and mixed by brief shaking. The sample was centrifuged for 10 min at maximum speed, and the aqueous phase transferred to a new tube. An equal volume of chloroform was added and mixed by brief shaking, followed by 10 min centrifugation. The aqueous layer was transferred to a new tube and the DNA precipitated by adding 30 μ l 3 M sodium acetate, 800 μ l ethanol, and 1 μ l glycogen (20 mg/ml) and incubating at least two hours at -20°C. The sample was centrifuged 10 min at maximum speed at 4°C and the supernate discarded. The DNA pellet was then washed with 800 μ l ice-cold 75% ethanol and centrifuged again for 10 min at 4°C. The supernate was discarded and the DNA pellet air-dried for 10 min. The DNA was then resuspended in 50 μ l water.

Shearing, end-repair, A-tailing, and adaptor ligation

The DNA was diluted to 100 μ l with water and sheared on a Covaris S2 instrument (in 6 × 16 mm snap-cap tubes with AFA fiber) with the following settings: intensity 5; duty cycle 5%; cycles/burst 200; time 3 cycles of 60 s each for total of 180 s (after this step the sample from bird B0042 was split into two aliquots, which were processed separately into libraries Mu7.1 and Mu7.2). End-repair and A-tailing (ERAT) was performed with a KAPA Hyper Prep kit by combining 50 μ l of sheared DNA with 7 μ l ERAT buffer (KAPA) and 3 μ l ERAT enzyme mix (KAPA), incubating 30 min at 20°C followed by 30 min at 65°C. Adapters were ligated by adding 1.5 μ l of 50 mM sequencing adaptor (see Table S3), 8.5 μ l water, 30 μ l ligation buffer (KAPA) and 10 μ l DNA ligase (KAPA), and incubating 1 hr at 20°C.

Streptavidin pulldown and PCR

MyOne streptavidin C1 beads (10 μ l of a 10 mg/ml suspension, Thermo Scientific) were washed three times in binding and washing (BW) buffer (5 mM Tris pH 8.0, 0.5 mM EDTA, 1 M NaCl) and resuspended in 220 μ l BW. The washed beads were added to the DNA sample and rocked for 30 min at room temperature, then separated by magnet and washed three times with BW (Libraries Mu7.1 and Mu7.2 were instead washed three times in BW with 0.1% Tween20, followed by one wash in 10 mM Tris pH 8.0 with 0.01% Tween20). The beads were resuspended in 20 μ l of 10 mM Tris pH 8.0. The library was amplified by 16 cycles of PCR after adding 5 μ l of 10x Library Amplification Primer Mix (20 μ M each, KAPA), and 25 μ l 2x KAPA HiFi Hotstart ReadyMix polymerase. The following cycling conditions were used: initial denaturation 98°C for 45 s; denaturation 98°C for 15 s; annealing 60°C for 30 s; extension 72°C for 30 s; final extension 72°C for 1 min.

Size selection and sequencing

PCR products were purified by solid phase reversible immobilization (SPRI) with Agencourt AMPure XP beads according to the manufacturer's instructions, at a beads:sample ratio of 1:1. The sample was resuspended in 33 μ l H2O. Library insert size selection (300-600 bp) was then performed on a BluePippin 2% gel, using the 'V1' internal size standard (Library Mu7.2 was instead size-selected on a Caliper LabChip with a DNA 750 kit with a size range setting of 320-480 bp). SPRI (1:1) was performed again and the DNA resuspended in 10 mM Tris pH 8.0. The libraries were then sequenced on an Illumina NextSeq in 2 × 76 bp mode.

Physical map construction from Hi-C data

Read mapping and filtering

Reads were mapped to the budgerigar genome (version Mu6.3) with bowtie2 (v2.2.9) with the –very-sensitive-local flag. The custom python script `add_tags.py` was used to identify the closest BsrGI restriction site downstream of the start of each read, and to append a custom SAM-formatted tag to the read denoting this site. Furthermore, read pairs that met the following criteria were tagged as proper Hi-C pairs: estimated insert size \leq 600 bp (based on the distances to the closest BsrGI site in a read and its mate pair);

phred-scale mapping quality ≥ 20 ; and intra-scaffold distance between read and mate pair ≥ 500 bp (for pairs mapped to the same scaffold). Read pairs that met these criteria were kept, and PCR duplicates were removed with Picard (v.2.6.0). The libraries sequenced are shown below:

Library	Bird ID	Total pairs (M)	Mapped (% of total)	Passing filter (% of total)	Passing filter and unique (% of total)	Intra-chromosomal contacts (%)
Mu5.2	B0142	51.3	69.8	0.9	0.5	58.7
Mu7.1	B0042	186.5	93.4	11.3	2.2	59.0
Mu7.2	B0042	286.5	96.4	3.7	0.9	58.1

The majority of read pairs were from intra-chromosomal contacts, indicating a low number of spurious contacts caused by random inter-chromosomal ligations (Kalhor et al., 2011). Fewer read pairs passed filter than in previous Hi-C experiments. For comparison, 45% of reads passed filter in a previous tethered Hi-C experiment with human lymphoblastoid cells (Kalhor et al., 2011).

Scaffold clustering

Lachesis was used to cluster the Mu6.3 scaffolds based on pairwise contact frequencies between loci in the Hi-C dataset. Multiple iterations of clustering were performed using different heuristic parameter settings, the ranges of which are summarized below:

LACHESIS INI file parameter	Range
CLUSTER_N	12-30
CLUSTER_MIN_RE_SITES	6-20
CLUSTER_MAX_LINK_DENSITY	5-12
CLUSTER_NONINFORMATIVE_RATIO	2-3
ORDER_MIN_N_RES_IN_TRUNK	15-30
ORDER_MIN_N_RES_IN_SHREDS	10-15

Scaffolds that exhibited intra-scaffold contact frequencies that appeared as a block diagonal matrix, indicative of incorrect joining between two or more of their component contigs (i.e., chimeric scaffolds), were manually split into separate sequences. The splits were positioned at existing gaps between contigs (filled with poly-N), such that the sequence of each contig remained undisturbed. The coordinate mapping between this new split assembly, Mu6.4, and its parent Mu6.3 was encoded in AGP format (mu6.4.agp, generated in R with the script *agp.r*). A new Mu6.4 genome fasta file was generated using the JCVI python utility (v0.6.9). The Hi-C reads were mapped to Mu6.4, and the Lachesis clustering performed again. The results from these several iterations of clustering were merged by hand, and in some cases manually adjusted, to give a consensus set of clusters.

Next, the clustering was refined by extracting reads that did not map to the ten largest clusters of scaffolds (Figure 1), and clustering them separately with different parameters:

LACHESIS INI file parameter	Range
CLUSTER_N	5-10
CLUSTER_MIN_RE_SITES	8-10
CLUSTER_MAX_LINK_DENSITY	5-10
CLUSTER_NONINFORMATIVE_RATIO	3
ORDER_MIN_N_RES_IN_TRUNK	15-25
ORDER_MIN_N_RES_IN_SHREDS	5-10

The results of the second round of clustering were then manually merged with the initial results to yield the final assembly, Mu6.5 (coordinate mapping between Mu6.3 and Mu6.5 described in mu6.5.agp).

Identification of clusters with respect to karyotype

The previously-published whole-genome alignment between budgerigar and chicken (Green et al., 2014) (see Key Resources Table), was extracted from HAL format to PSL format with haltools, and converted to BED format with the custom python script *psl2bed.py*. The intervals in this file were lifted over to Mu6.5 coordinates, and dot-plots summarizing the alignment were generated with ggplot2 in R (Figure S2). Clusters were identified with respect to the budgerigar karyotype by comparing these dot-plots to previously published fluorescent in situ hybridization (FISH) data from budgerigars with heterologous probes from flow-sorted chicken

chromosomes (Nanda et al., 2007). Chromosomes 1–9 were unambiguously identified using this method. The Z-chromosome was identified by its reduced sequencing coverage, which is expected to be half that of the autosomes in a female. The remaining chromosomes were numbered 10–21 according to size (bp).

Contact map normalization

The genome was divided into 1 Mb bins ($n = 1118$) and the observed number of pairwise contacts between bins i and j ($c_{i,j}$) was counted for every combination of i and j . The expected number of contacts between bins i and j ($e_{i,j}$) was calculated as follows:

$$e_{i,j} = \frac{\sum_{k=1}^n c_{i,k} \sum_{l=1}^n c_{k,l}}{\sum_{k=1}^n \sum_{l=1}^n c_{k,l}}$$

Isolation of genomic DNA

Blood from a toe-nail clipping or regenerating contour feathers (one or two ‘pin feathers’ from the top of the head) were collected in Longmire Buffer (100 mM Tris HCl pH 8.0, 100 mM EDTA, 10 mM NaCl, 0.5% SDS), and stored on ice for up to a day. Samples were digested with 500 µg proteinase K overnight at 55°C with mixing. Phenol:chloroform:IAA extraction, followed by ethanol precipitation was used to isolate genomic DNA.

DNA was extracted from the museum specimens with a similar procedure. Approximately 10 mg of tissue was removed from the storage solution and digested with 600 µg proteinase K in 10 mM Tris pH 8.0, 100 mM NaCl, 10 mM EDTA, and 0.5% SDS overnight at 55°C with mixing. DNA was then extracted, as described above.

Opossum (*Monodelphis domestica*) DNA from approximately 20 mg of testis tissue, preserved in 70% ethanol, was isolated with the same procedure as that described above for the museum samples.

Modified ddRAD sequencing

Custom sequencing adapters with BgIII or Ddel 5' overhangs were prepared by annealing single-stranded oligos (45 µM each) in 10 mM Tris-Cl pH 8.0 with 50 mM NaCl. The oligo sequences are listed in Table S3. Annealing was carried out in a thermal cycler by a stepwise reduction in temperature (30 s per cycle with –0.5°C steps between cycles) from 97°C to 12°C. The adaptor concentration was then measured by a fluorometric assay (Qubit). Genomic DNA (200 ng) was digested with 10 U BgIII and 10 U Ddel in 40 µl 1x NEBuffer 3.1 at 37°C for 4 hr, or overnight. The digested DNA was cleaned up by SPRI with Agencourt AMPure XP beads (sample:beads ratio of 1:1.8). An 11-fold molar excess of BgIII sequencing adaptor (2 pmol), relative to the estimated amount of BgIII 5' overhangs per 200 ng digested budgerigar DNA (assuming a genome molecular weight of 6.78×10^{11} g/mol and 3.33×10^5 BgIII sites per genome) and an 11-fold excess of Ddel sequencing adaptor (30 pmol) were added to the samples, and ligated by 400 U T4 DNA ligase in 40 µl 1x NEB T4 DNA ligase buffer for 2 hr at room temperature. The ligation reaction was stopped by a 10 min incubation at 65°C, followed by cleanup by SPRI (sample:beads ratio of 1:1.8). To prevent PCR amplification of inserts flanked on both ends by Ddel adapters, the nick-containing ligation product was extended with ddTTP (100 µM) in 1x NEBuffer 2 with 5 U Klenow 3'→5' exo- for 30 min at 37°C. The reactions were cleaned up by SPRI (sample:beads ratio of 1:1.8), and amplified by 8–12 cycles of PCR with Q5 polymerase (2x NEBNext high fidelity master mix), and a combination of index-containing primers (0.5 µM each, see also Table S3) with the following cycling conditions: initial denaturation 98°C for 30 s; denaturation 98°C for 10 s; annealing 62°C for 30 s; extension 72°C for 30 s; final extension 72°C for 5 min. Up to 96 samples were pooled in equal amounts according to the concentration of PCR product between 280–480 bp as measured by bioanalyzer. The library pools were cleaned up by SPRI (sample:beads ratio of 1:1.8), and resuspended in 100 µl water. Inserts were size-selected on a Caliper LabChip with the DNA 750 kit (10 µl sample per lane with the setting ‘extract and stop’ with size range 350 bp ± 14%). Variability in the resulting size distributions across DNA 750 chips and lanes was dealt with by performing replicate size selections across 3–10 lanes for each pool, assaying the DNA from each lane by bioanalyzer, and using only those that had the desired size distribution. DNA from the chosen lane(s) was cleaned up by SPRI (sample:beads ratio of 1:1.8), and sequenced on the Illumina NextSeq in single-end 151 bp mode with 8 bp dual indexes.

ddRAD-seq SNV calling

Reads were mapped to the budgerigar genome (Mu6.3) with BWA (v0.7.9). Tables of predicted BgIII and Ddel restriction digest fragments were generated with the custom python scripts *digest.py* and *fragments.py*, and sequencing coverage was measured at these sites with the script *restriction_site_coverage.py*. The “target region” for SNV calling was defined as the set of fragments between 100–350 bp long that had non-zero coverage in at least one individual. The mapping data are summarized in Table S4. GATK (v.3.3.0) UnifiedGenotyper (with -stand_call_conf 20) was used to call SNVs within this set of intervals (37.6 Mb in total).

Sites with coverage $< 8x$ in $\geq 10\%$ of samples, or coverage $> 250x$ in any sample, were removed. Within the remaining set of sites, pairwise nucleotide diversity ($\hat{\pi}$), defined below, was 0.0037 bp^{-1} , where $N = 5,676,200 \text{ bp}$, $S_n = 240,004 \text{ segregating sites}$, $n = 234 \text{ samples}$ (domesticated varieties), and \hat{p}_i is the allele frequency estimate for site i .

$$\hat{\pi} = \frac{1}{N} \frac{n}{(n-1)} \sum_{i=1}^{S_n} 2\hat{p}_i(1 - \hat{p}_i)$$

As a first filtering step, low frequency SNVs ($\text{MAF} < 0.05$) were removed, leaving 102,633 segregating sites.

Polymorphic restriction sites, which can cause genotyping errors in ddRAD datasets, were estimated from the genotypes and sequencing coverage with gbstools (v0.1.0). Because gbstools calculates genotype likelihoods based on the assumption of a randomly-mating population, only individuals without a pedigree-listed parent in the dataset were used in this analysis ($n = 148$). To account for variation in insert size distributions across samples, a table of normalization factors, $r_{i,j}$ were calculated for every insert size i between 1-1000 bp and every sample j ($n = 148$), where $c_{i,j}$ is the genome-wide number of reads of insert size i in sample j :

$$r_{i,j} = \frac{c_{i,j}}{\sum_{k=1}^n c_{i,k}}$$

Since the reads were single-ended, the insert sizes were estimated from the mapped locations using the results of the simulated restriction digest described above. Likelihood ratio tests for restriction site polymorphism were then performed with gbstools using the $-n$ option to specify the coverage normalization factor table, and with the heuristic coverage dispersion settings $-dispersion_slope 0.12$ and $-dispersion_intercept 2.0$. SNVs that exhibited a likelihood ratio > 2.71 were removed, leaving 69,855 segregating sites, which were used in downstream analyses.

LD estimation

Before estimating pairwise LD, cryptically related samples were removed from the dataset by estimating identity-by-descent (IBD) with PLINK (v1.9), using the $-genome$ option. Since this calculation itself is LD-sensitive, the SNPs were first pruned for LD using the $-indep-pairwise 50 5 0.2$ option, leaving 17,698 SNPs. After estimating pairwise IBD (PI_HAT), the set of individuals was sampled without replacement many times to obtain the largest subset ($n = 72$) not connected by $\text{PI_HAT} > 0.125$ (equivalent to third degree relatives or less). Pairwise LD was then calculated for common SNPs ($\text{MAF} \geq 0.3$) in the 69,855-SNP set, but with just this set of 72 individuals, using the options $-r2-ld-window 99999-ld-window-r2 0$. SNP pairs were binned by distance and the mean pairwise LD per bin was plotted in R (Figure S3A). Long runs of homozygosity (ROHs), which are indicative of more recent events such as inbreeding, were estimated using the $-homozyg$ option and plotted in R (Figure S3B).

Genome-wide association mapping

SNPs were lifted over to Mu6.5 coordinates (the Hi-C assembly described above). At each SNP locus, a 3×2 contingency table was made, summarizing the counts of genotypes (homozygous reference, heterozygous, and homozygous non-reference) versus phenotypes (WT, and blue). Fisher's exact test (fisher.test R function) was used to calculate p -values for each observed distribution of counts. The Bonferroni-corrected significance threshold was calculated as $0.05 / n$, where $n = 69,855$ SNPs.

Haplotype inference

To simplify the analysis, a subset of the SNPs on scaffold NW_004848279.1 was used. SNPs in the interval 19,000,000-22,000,000 bp were grouped according to the BglII restriction site they mapped to in the ddRAD dataset. The SNPs within each group were ordered by association mapping p -value and minor allele frequency, and all were removed except the top SNP per group. Low-frequency SNPs ($\text{MAF} < 0.1$) were removed, leaving 65 sites. Next, individuals with missing data at more than 4 of 65 sites were removed, leaving 232 out of 249 samples. Loci with missing data in more than 9 of 232 samples were then removed, leaving 63 sites. Haplotypes were inferred with PHASE (v2.1.1) using the default settings. The population-wide count of each haplotype was calculated as shown below, where $n = 232$ individuals, and where the N haplotypes inferred by PHASE are denoted by $\{h_1, h_2, \dots, h_N\}$. The probability that (h_j, h_k) is the true haplotype pair for individual i (denoted by H_i) is expressed as $\text{Pr}\{H_i = (h_j, h_k)\}$, which is output by PHASE for the most likely pairs for each individual (i.e., those with a value of at least 0.01).

$$h_m \text{ count} = \sum_{i=1}^n \sum_{j=1}^N \sum_{k=1}^N \text{Pr}\{H_i = (h_j, h_k)\} \{1_{h_m}(h_j) + 1_{h_m}(h_k)\}$$

$$1_{h_m}(h) = \begin{cases} 1 & \text{if } h = h_m \\ 0 & \text{otherwise} \end{cases}$$

Haplotypes with a count of ≥ 2 (i.e., a frequency of $\geq 0.4\%$) were plotted (Figure 3A), along with columns of marginal counts in WT and blue birds. Haplotype number 8 (Figure 3A) was found in only one sample, but was also plotted because it shows evidence of an ancestral recombination between markers at 21,019,187 and 21,161,723 bp. Alleles were called as derived or ancestral based on a previously-published whole-genome alignment of 15 bird species (Green et al., 2014) (See Key Resources Table).

The halLiftOver utility (haltools v2.1) was used to extract SNP-aligned regions in each of the 15 target genomes to BED and PSL files. The getfasta utility (bedtools v2.25.0) was used to obtain the nucleotides located within these intervals in each target genome and the data were merged in R. In cases where multiple matches existed for a single SNP in a given target genome, one was chosen at random. Nucleotides that did not match either the budgerigar reference or non-reference allele were discarded and treated as

missing data. The ancestral alleles were then called manually by comparison to outgroup species, and by minimizing the number of independent substitutions required to explain the data, given the species tree.

Feather mRNA-seq

Isolation of RNA from regenerating feathers

Regenerating contour feathers (one or two “pin feathers” from the top of the head) were collected with tweezers into 1 ml RNAlater and stored on ice. Feathers were then sliced longitudinally one time with a sterile scalpel and transferred into 1 ml TRIzol in an MP FastPrep “matrix D” tube. The tissue was homogenized for 40 s at 6.0 m/s at 4°C on an MP FastPrep. The homogenate was cleared by centrifugation (12,000 × g for 10 min at 4°C). The supernate was transferred to a new tube and mixed with 0.2 ml chloroform and incubated 3 min at room temp. The aqueous phase was separated by centrifugation, washed with 0.5 ml chloroform, and the RNA precipitated with isopropanol with 20 µg glycogen. The RNA pellet was dried and redissolved in water, and residual DNA was removed with a “turbo DNA-free” kit (Ambion) according to the manufacturer’s instructions. RNA integrity numbers (RINs) were estimated by Agilent Bioanalyzer (RNA 6000 nano assay) or Tapestation (high sensitivity RNA ScreenTape assay).

Library construction and sequencing

RNA-seq libraries were constructed with Illumina TruSeq stranded mRNA-seq kits, following the manufacturer’s instructions. The input consisted of 4–5 µg of whole RNA from samples with RIN ≥ 8.0. The recommended RNA fragmentation time of 8 min at 94°C was used. The reverse transcriptase was SuperScript II. At the PCR step, 15 cycles were used. The libraries were sequenced on an Illumina NextSeq in either 2 × 76 bp or 2 × 151 bp mode.

Read mapping and transcript quantification

Reads were trimmed to 75 bp with the fastx toolkit (v0.0.13), and mapped to the budgerigar genome (version Mu6.3) with tophat, using the –G option to specify the RefSeq gene annotations as known transcripts. The libraries sequenced are shown below:

Library	Bird ID	Sample ID	RIN	Total Read Pairs (M)	Mapped (% of Total)
Mu4.1	B0214	TCR3	9.4	27.3	64.5
Mu4.1	B0217	TCR9	9.0	26.6	62.5
Mu9.1	B0216	TCR10	8.5	34.9	77.6
Mu9.1	B0225	TCR11	8.2	30.1	74.3
Mu9.1	B0297	TCR12	8.7	25.5	75.8
Mu9.1	B0224	TCR16	9.2	34.0	74.5
Mu9.1	B0213	TCR4	8.8	21.3	73.7

Transcripts were assembled separately for each sample with cufflinks, and merged with the cuffmerge utility. Tests for differential expression in the merged set of transcripts between WT ($n = 3$) and blue ($n = 4$) samples were performed with cuffdiff.

Multi-species mRNA-seq comparison

RNA-seq data from chicken, crow, lizard, opossum, and zebrafish was downloaded from the Sequence Read Archive (Table S5) and mapped with tophat to the RefSeq genome assemblies shown below, using the –G option to specify the splice junctions contained in the RefSeq genome annotations.

Species	RefSeq Assembly
<i>Anolis carolinensis</i>	AnoCar2.0
<i>Corvus cornix cornix</i>	Hooded_Crow_genome
<i>Danio rerio</i>	GRCz10
<i>Gallus gallus</i>	Gallus_gallus-4.0
<i>Monodelphis domestica</i>	MonDom5

Transcripts were assembled separately for each sample with cufflinks and merged by species with cuffmerge. FPKM expression levels for these merged sets of transcripts, normalized by library size, were calculated with cuffquant and cuffnorm. For the purpose of cross-species comparisons, FPKM levels were also calculated in this way for the budgerigar data.

Budgerigar mRNA-seq variant calling

Budgerigar RNA-seq reads containing a splice junction were split using the GATK SplitNCigarReads utility and variants called with GATK haplotypecaller. The effect of each variant on genes in the region NW_004848279.1:21,019,187-21,445,705 was predicted with snpEff (v4.3i) using the budgerigar RefSeq gene annotations (version Mu6.3). Variants that displayed the expected segregation pattern under a recessive Mendelian model, and had predicted effects of MODERATE (e.g., missense) or HIGH (e.g., nonsense) were

then tabulated (Figure S5A). For a full description of annotations and putative impacts, see http://snpeff.sourceforge.net/VCFannotationformat_v1.0.pdf.

MuPKS genotyping

The *MuPKS* coding region near the R644W SNP was amplified by PCR (with Q5 polymerase) from genomic DNA with primers BL_36_F and BL_36_R, and Sanger sequenced with primer BL_39_F (Table S3). The genotypes for each bird were determined by manual inspection of the sequencing chromatograms (Table S6).

Feather section *in situ* hybridizations

Regenerating feathers (2 to 3 “pin feathers” from the top of the head) were sampled with tweezers and fixed in 4% paraformaldehyde in PBS (pH 7.4) overnight at 4°C. The samples then serially dehydrated with alcohol, embedded in paraffin and sectioned at 7 µM. The paraffin section *in situ* hybridization was performed as previously described (Li et al., 2017). For MuPKS probe preparation, we PCR-amplified a 522 base pair fragment of *MuPKS* by using sense primer (5'-tgctttggatttggaggaaac-3') and antisense primer (5'-ctggaaaagctctggattcgc-3'). The cDNA from a budgerigar contour feather was used as template. The PCR product was inserted into the pDrive plasmid (QIAGEN). Digoxigenin-labeled probes against *MuPKS* was synthesized by use of a digoxigenin RNA labeling kit according to the instructions from the manufacturer (DIG RNA Labeling kit, Roche). Diluted eosin was used for faint counter-staining.

MuPKS expression in yeast

Verification of *MuPKS* gene model

To verify the *MuPKS* gene model generated by cufflinks from the feather mRNA-seq data (and the automated RefSeq gene model) the 5' and 3' ends of the *MuPKS* coding sequence were determined by 5' RACE and 3' RACE with a GeneRacer kit (Invitrogen). The RNA sample used was TCR3 from bird B0214. The reaction was primed with oligo dT for 3' RACE, and by the internal primer BL_49 for 5' RACE. Nested PCR was performed with primers BL_52 and BL_53 (5' RACE), shown in Table S3, or BL_54 and BL_55 (3' RACE) and the nested primers provided in the kit. The PCR products were cloned into a TOPO-TA vector and several clones were Sanger sequenced.

Generating cDNA from chicken embryo

The chicken embryo stored in TriPure reagent was homogenized for 40 s at 6.0 m/s at 4°C on an MP FastPrep in a “lysing matrix D” tube. RNA was then isolated according to the manufacturer’s instructions. Reverse transcription was performed with Protoscript II (NEB) primed with oligo dT (dT₂₃VN).

Expression plasmid construction

The 2µ-based yeast-*E. coli* shuttle vector pYR291, which carries the URA3 marker, was a gift of Yi Tang (Li et al., 2011). The gene carried between the ADH2 promoter and ADH2 terminator on pYR291 was excised as an NdeI-RsrlI digest fragment and replaced with the annealed oligonucleotides pTFC24_MCS_top and pTFC24_MCS_bot, resulting in the empty vector pTFC24. The cloning site was later modified again by excising the NaeI-RsrlI fragment, and replacing it with the annealed oligonucleotides pTFC24b_MCS_top and pTFC24b_MCS_bot, resulting in the empty vector pTFC24b. A similar process was used to generate pTFC25b. The full-length *MuPKS* coding sequence (CDS) was Gibson-assembled from two half-CDSs that were generated from cDNA from bird B0214 (heterozygous for *blue*). The PCR primers used were PKS_10, BL_25_R, PKS_19, and PKS_69. The two half-CDSs were Gibson-assembled into pTFC24b digested with BstEII. The correct sequence for the assembly products (two of the four possible products are chimeras of the WT and *blue* alleles) was determined by sequencing genomic DNA from the same bird. Briefly, genomic DNA was PCR-amplified with primers PKS_11 and PKS_12. These products were cloned into a TOPO-TA vector and sequenced with primers BL_25_F and BL_25_R. The Gibson assembly products were then screened for the correct sequences of the *blue* and WT alleles by sequencing with primers shown in Table S3, resulting in pTFC26b (*MuPKS* WT) and pTFC27b (*MuPKS* blue). Site-directed mutagenesis was performed by PCR-amplifying pTFC26b with primers PKS_10, PKS_58, PKS_59, and PKS_69, and cloning them into pTFC24b by Gibson assembly, resulting in pTFC28b (*MuPKS* WT R644W). Chicken GgPKS1 was amplified from cDNA with the primer pairs GgPKS_1 and GgPKS_2, and GgPKS_3 and GgPKS_4, and cloned by Gibson assembly into pTFC25b digested with BstEII, resulting in pTFC30b.

Transformation and expression

For transformation, the yeast strain BJ5464-NpgA was cultured to mid-log phase in 50 ml YPD medium and harvested by centrifugation for 5 min at 800 × g. The supernate was removed and the cell pellet was resuspended in 5 ml water and centrifuged again. The supernate was again removed and the cell pellet was resuspended in 1 ml of 100 mM lithium acetate and briefly centrifuged. This step was repeated, but the pellet was resuspended in a final volume of 0.5 ml of 100 mM lithium acetate. The following components were added to a 1.5 ml tube, in the order they appear, and with vortexing between each addition: 240 µl 50% (w/v) polyethylene glycol (average M_n 3350), 36 µl 1 M lithium acetate, 50 µl yeast suspension, 5 µl salmon sperm DNA (10 mg/ml), 10 µl plasmid. The yeast were then incubated 45 min at 30°C, followed by addition of 35 µl DMSO. The yeast were then heat-shocked at 42°C for 15 min. Transformants were selected on CSM-Ura plates. For protein expression, overnight cultures (3 ml) were grown in SD-Ura at 30°C and transferred to 1 l of YPD in fernbach flasks. The cultures were grown for 72 hr at 28°C. The cells were then harvested by centrifugation and stored at –80°C.

Yeast western blot

Approximately 100 mg (wet weight) of frozen yeast pellet was resuspended in 150 µl of zymolyase buffer (50 mM Tris-HCl pH 7.5, 10 mM MgCl₂, 1M sorbitol) with 30 mM DTT and incubated 15 min at room temperature. The cells were centrifuged 5 min at 1500 × g and the supernate discarded. The cell pellet was then resuspended in 100 µl zymolyase buffer with 1 mM DTT and approximately 1 mg/ml zymolyase and incubated 1 hr at 30°C with mixing at 1000 RPM. The spheroplasts were pelleted by centrifugation (5 min at 1500 × g), the supernate discarded, and the pellet washed with 500 µl zymolyase buffer. The spheroplasts were centrifuged again (5 min at 1500 × g), the supernate discarded, and the pellet resuspended in 500 µl cold PBS with 1x HALT protease inhibitor (Thermo Scientific) and lysed by sonication (30 s. at 3W). The homogenate was cleared by centrifugation (10 min at 16,000 × g at 4°C), and the A260 of the cleared homogenate was measured. Approximately 0.43 AU of homogenate was run on a 3%–8% tris-acetate gel, and the proteins were transferred to a PVDF membrane (wet transfer). MuPKS-6 × His was detected by western blot. Briefly, the membrane was blocked for 1 hr at room temperature with 5% nonfat dry milk in PBS with 0.1% Tween-20 (PBST). The membrane was rinsed briefly in PBST and then incubated overnight at 4°C with α-His antibody (Pierce) diluted 1:2000 in 0.5% nonfat dry milk in PBST. The membrane was then washed 3 times (15 min each) with PBST. Next the membrane was incubated for 1 hr at room temperature with goat anti-mouse IgG, HRP-conjugated secondary antibody diluted 1:40,000 in 0.5% nonfat dry milk in PBST. The membrane was then washed 3 times (15 min each) and developed by incubating 1 min with ECL western blotting substrate (Pierce).

Verification of MuPKS phosphopantetheinylation

Protein purification and digestion. 6 × His-tagged MuPKS was purified by Ni-NTA chromatography as follows. Approximately 10 g of frozen yeast pellet (see *Transformation and Expression* section above) was thawed and resuspended in 10 ml zymolyase buffer with 30 mM DTT and incubated 15 min at room temperature. The cells were pelleted by centrifugation (5 min at 1500 × g at 4°C). The supernate was discarded and the cells resuspended in 30 ml zymolyase buffer with 1 mM DTT. The cells were converted to spheroplasts by incubating 2 hr at 30°C with approximately 20 mg zymolyase (Sunrise Science Products), with gentle mixing. The spheroplasts were pelleted by centrifugation (5 min at 1500 × g at 4°C), and resuspended in 20 ml cold zymolyase buffer with 1 mM DTT. This was repeated for a total of 3 washes. The final wash was with cold lysis buffer (50 mM sodium phosphate pH 7.8, 500 mM NaCl, 10 mM imidazole, 0.1 mM PMSF, 1 mM TCEP). The spheroplasts were then resuspended in 20 ml cold lysis buffer and lysed by sonication (3 cycles of 1 min each at 4W, with cooling on ice for 1 min between cycles). The lysate was incubated with 0.5 ml Ni-NTA resin (QIAGEN) for 30 min at 4°C with rocking, and then loaded onto a gravity column. The resin was washed on the column with wash buffer (same composition as lysis buffer but with 20 mM imidazole) until the A280 stabilized. Proteins were eluted from the column with 50 mM sodium phosphate pH 7.8, 500 mM NaCl, 250 mM imidazole, 1 mM TCEP into 1 ml fractions. The protein concentration in the fractions was estimated by A280 and approximately 5 µg was run on a NuPAGE 3%–8% tris-acetate gel with tris-acetate SDS running buffer. The gel was stained with Coomassie blue and the 235 kDa band was excised. This sample was in-gel digested using proteases GluC and AspN (Promega). In brief, samples were washed with 50 mM ammonium bicarbonate, followed by reduction with DTT (5 mM) and alkylation using propionamide (10 mM). Gels were further washed with a acetonitrile/ammonium bicarbonate buffer until all stain was removed. Next, 500 ng of GluC reconstituted to 25 ng/µl in 50 mM ammonium bicarbonate with 0.1% protease max (Promega) was added to each gel band and incubated 30 min. Next, 500 ng of AspN in 50 mM ammonium bicarbonate with 0.1% protease max was added to the gel and digested overnight. Peptides were extracted from the gels in duplicate and dried completely by speedvac.

LC/MS data collection

Peptide pools were reconstituted and injected onto a C18 reversed phase analytical column, ~10.5 cm in length (Picochip, New Objective). The UPLC was a Waters NanoAcuity, operated at 450 nL/min using a linear gradient from 4% mobile phase B to 45% B. Mobile phase A consisted of 0.585% acetic acid, water. Mobile phase B was 0.585% acetic acid, water. The mass spectrometer was an Orbitrap Elite set to acquire data in a data dependent fashion selecting and fragmenting the 15 most intense precursor ions in the ion-trap where the exclusion window was set at 45 s and multiple charge states of the same ion were allowed.

LC-MS data analysis

MS/MS data were analyzed using both Preview and Byonic v2.6.49 (ProteinMetrics). All data were first analyzed in Preview to provide recalibration criteria if necessary and then reformatted to .MGF before full analysis with Byonic. Data were searched at 12 ppm mass tolerances for precursors, with 0.4 Da fragment mass tolerances assuming up to two missed cleavages and allowing for fully specific and ragged GluC and AspN peptides. These data were validated at a 1% false discovery rate using typical reverse-decoy techniques (Elias and Gygi, 2007). The resulting identified peptide spectral matches and assigned proteins were then exported for further analysis using custom tools developed in MATLAB (MathWorks) to provide visualization and statistical characterization.

Pigment extraction

Approximately 150–200 mg of yeast cells (wet weight) were homogenized in a 2 ml tube with roughly 0.5 g of glass beads (212–300 µm) in 0.5 ml methanol by shaking 4 × 45 s at 6.5 m/s with an MP FastPrep. The homogenate was cleared by centrifugation for 5 min at 16,000 × g. The supernate was then transferred to a new tube and dried under a stream of N₂. The residue was redissolved in 0.5 ml 2% acetic acid in ethyl acetate, and washed by vortexing briefly with 0.5 ml water. The organic phase was transferred to a new tube and dried under a stream of N₂. The residues were then redissolved in 0.5 µl methanol per mg of cells used (wet weight), and analyzed by LC-MS.

Pigments used for photographs (Figure 6A) were extracted by a simplified procedure. Approximately 300 mg of yeast cells (wet weight) were homogenized in the same manner, but in 0.5 ml 2% acetic acid in ethyl acetate. The homogenate was cleared by centrifugation, as above, and an aliquot of the supernate was taken (approximately 0.63 µl per mg of cells used). The

samples were concentrated under a stream of N₂ until 40 µl remained. Samples were then transferred to 0.5 ml qubit tubes for photographs.

Pigment extraction from feathers

The feather pigment extraction procedure was similar to the one used in a previous study of psittacofulvin pigments in parrot feathers (McGraw and Nogare, 2005). Yellow budgerigar flight feathers (molted) were washed with detergent and thoroughly rinsed in warm water, followed by a thorough rinsing in ethanol. They were then briefly rinsed in hexane and dried. Approximately 5 mg of feather barbs were trimmed from the rachis and weighed. Pigment was extracted from the barbs by incubation for 1 hr at 95°C in approximately 1 ml of 2% HCl in pyridine. After cooling to room temperature, the solvent was evaporated under a stream of N₂, and the residue redissolved in 0.5 ml 2% acetic acid in ethyl acetate. The organic phase was washed by vortexing briefly with 0.5 ml water, followed by centrifugation for 5 min at 16,000 × g. The supernate was transferred to a new tube and dried under a stream of N₂. The pigment was then redissolved in 75 µl methanol per mg of barbs and analyzed by LC-MS.

Pigment analysis by mass spectrometry

LC-MS analysis was conducted on an Agilent 6545 quantitative time-of-flight mass spectrometer interfaced to an Agilent 1290 HPLC system. The ion source for Figure 6 and unless otherwise specified was an electrospray ionization source (dual-inlet Agilent Jet Stream or “dual AJS”). The column for Figure 6 was a Thermo Hypersil Gold perfluorophenylpropyl (PFP) column (2.1 mm × 50 mm × 1.9 µm).

Ion source conditions for all analyses are specified in the table below:

Ion source parameter	Value
gas temperature	250°C
drying gas	12 l / min
nebulizer	10 psig
sheath gas temp.	400°C
sheath gas flow	12 l / min
capillary voltage (Vcap)	3500 V
nozzle voltage	1400 V
fragmentor	100 V
skimmer	50 V
octopole 1 RF Vpp	750 V

For Figure 6, mobile phase A was 0.1% v/v formic acid in water, and mobile phase B was 0.1% v/v formic acid in acetonitrile. Gradient conditions were isocratic at 95% A for 0.2 min, gradient from 95% A to 50% A from 0.2 to 2.2 min, gradient from 50% A to 5% A from 2.2 to 8.2 min, gradient from 5% A to 0% A from 8.2 to 9.2 min, followed by isocratic column regeneration at 95% A from 9.21 to 10 min.

For other methods, the column was an Agilent Zorbax 50 mm × 2.1 mm RRHD Eclipse C18 column with 1.8 µm beads, and gradient conditions were isocratic at 95% A from 0 to 0.2 min, with a gradient from 95% A to 5% A from 0.2 to 4.2 min, followed by isocratic conditions at 5% A from 4.2 to 5.2 min, followed by a gradient from 5% A to 95% A from 5.2 to 5.2 min, followed by isocratic re-equilibration at 95% A from 5.2 to 6 min.

For the APCI analysis, the ion source was an APCI ion source (an Agilent multimode ionization source “MMI”), and mobile phase B was 0.1% v/v formic acid in methanol (acetonitrile is not compatible with this ion source). Ion source conditions are shown in the table below:

Ion source parameter	Value
gas temperature	350°C
drying gas	7.5 l / min
nebulizer	20 psig
vaporizer	250°C
capillary voltage (Vcap)	1500 V
corona discharge	4 µA
fragmentor	120 V
skimmer	50 V
octopole 1 RF Vpp	750 V
charging voltage	1000 V

LC-MS peaks associated with the observed UV peaks were identified by xcms analysis (Smith et al., 2006) where yellow feather extracts, chicken MuPKS-expressing yeast, and parakeet MuPKS-expressing yeast were grouped into a “positive” sample class, and yeast negative controls as well as white feather extracts were grouped into a “negative” sample class. XCMS-identified features were filtered to focus only on those in the retention time window of the UV peaks, and arranged in order of the ANOVA statistic. The sixth feature in the resulting list proved to have an extracted ion chromatogram (EIC) that perfectly eluted with an observed pigment peak (Figure 6).

Opossum *MuPKS* homologs

PCR primers were designed to amplify a gap-containing region of the opossum genome that spanned part of the coding sequence of the *MuPKS* homolog *LOC100024872* (NC_008808.1:245,751,915-245,752,413). PCR products were amplified from genomic DNA with primers MonDom_3 and MonDom_9 and Sanger-sequenced using primers MonDom_2 and MonDom_4. The sequences were de novo assembled with Geneious v7.1.9, and the sequence gap filled in manually. A spurious 1 bp deletion at NC_008808.1:245,755,415 was also found and corrected within the same gene by examination of the opossum RNA-seq dataset, which was described above. The gene model *MdPKS2* was then manually constructed from *LOC100024872* using the new sequence. *MdPKS1* was identified upstream, in the interval NC_008808.1:245,663,536-245,692,698, by manually searching for long open reading frames and canonical splice junction sequences.

Phylogenetic analysis

Protein sequences homologous to MuPKS in metazoans were identified by blast queries against the GenBank non-redundant protein database, and against the following databases:

http://marinegenomics.oist.jp/acornworm/blast/search?project_id=33
http://marinegenomics.oist.jp/cots/blast/search?project_id=46

The sequence of the *Schizocardium californicum* polyketide synthase was provided by Chris Lowe of Hopkins Marine Station at Stanford University, and is from a previously described transcriptome assembly (Gonzalez et al., 2017).

A representative set of outgroup sequences, such as metazoan fatty acid synthases and microbial polyketide synthases, were chosen based on previous phylogenetic work, with special emphasis placed on enzymes with known biochemical activity or solved crystal structures. Table S7 lists accession numbers for the sequences used in the KS alignment. The sequences were aligned with mcoffee and the alignments were then refined by hand, making use of known secondary structures when possible.

A maximum likelihood phylogenetic tree for the KS domain was constructed with PhyML v3.2.0 under the LG + Γ model (4 substitution rate categories). The most appropriate model was estimated with ModelGenerator v0.85 (Keane et al., 2006). The tree and the domain structures corresponding to each protein were plotted with the R package ggtree v1.2.17. Domain annotations were downloaded from GenBank, and ME domains were categorized as active/inactive based on the presence or absence of a conserved sequence involved in binding the cofactor S-adenosyl-methionine (Maier et al., 2008). A similar analysis was carried out for the ER domain and the conserved NADPH cofactor binding site (Keatinge-Clay, 2012).

Synteny analysis

Genomic scaffolds that harbor *MuPKS* homologs from opossum, lizard, medaka, or human were identified by blast (human chromosome 10 contains a remnant *MuPKS*-like sequence that aligns to the first four exons). The scaffolds were aligned with progressive-cactus, using the newick-formatted tree:

((human, opossum), (budgerigar, anole)), medaka;

Default branch lengths of 1 were used, and each of the assemblies except budgerigar was treated as reference-quality (* option). The alignment was stored as a HAL graph (Hickey et al., 2013), and a new method developed for this study (Figure S7A, <https://github.com/cooketho/halplot>) was used to traverse the HAL-formatted alignment and return edges connected to the genes near the *MuPKS* homologs (Figure S7B). The custom python script *count_inversions_greedy.py* was used to find the orders and orientations of “parent” genome sequences (e.g., Figure S7A) that minimize the number of edge crossings and inversions between child-parent genome pairs, and between child-parent-child trios. The HAL graph edges were then plotted with ggplot2 in R using the script *halplot.r*.

QUANTIFICATION AND STATISTICAL ANALYSIS

Test for genotype-trait association

Fisher’s exact test was chosen because both the collected phenotype data (wild-type versus blue) and the genotype data (homozygous reference, heterozygous, or homozygous non-reference) are categorical. The observations at each SNP were summarized in 2 × 3 contingency tables. The null hypothesis is that the observed genotype counts are independent of the phenotype counts. The experiment-wide critical value was $\alpha = 0.05$, and the Bonferroni-corrected critical values for each individual test were α / n , where $n = 69,855$.

Tests for differential gene expression

Cuffdiff was used to calculate p values for differential expression of genes in regenerating feathers of WT ($n = 3$) and *blue* ($n = 4$) budgerigars (Trapnell et al., 2013). This program estimates each gene's expression level based on a model that accounts for uncertainty in the assignment of reads to transcript isoforms (genes exhibiting splice variation will have multiple isoforms), and overdispersion in read counts across replicates. As described previously (Trapnell et al., 2013), the log-transformed ratio of expression levels between the two sets of replicates (WT versus *blue*) divided by the variance of the log-ratio constitutes the test statistic. Under the null hypothesis of equal expression levels in both sets, the test statistic follows an approximately standard normal distribution. Two-sided tests for significance were performed for genes in the *blue*-shared haplotype (Figure 3A). *LOC101880715* and *TRNAN-GUU* showed FPKM < 0.04 in both *wild-type* and *blue*, and were thus excluded from the analysis. The experiment-wide critical value was $\alpha = 0.05$. The Benjamini-Hochberg procedure was used to correct the critical values for each individual test. By default, cuffdiff carries out this correction using the full set of genes (24,378 tests), but since this experiment concerns only genes in the *blue*-shared haplotype, we manually performed the correction with just this set of genes (9 tests). None of the 9 genes showed significant differential expression.

DATA AND SOFTWARE AVAILABILITY

Sequencing data from the ddRAD-seq and mRNA-seq experiments were deposited at the Sequence Read Archive (SRA) under project accession PRJNA378643. Sequencing data from the Hi-C experiments were deposited under project accession PRJNA378785.

Custom software used in the analyses described above are available at the following github repositories:

<https://github.com/cooketho/budgies>
<https://github.com/cooketho/halplot>

Additional Resources

The animal icons in Figure 7C are derived from images that have been published online under Creative Commons license, allowing them to be used, shared, or modified. The licenses, and information about the licensors can be found at the following URLs:

Species	Creative Commons License URL
Budgerigar	https://commons.wikimedia.org/wiki/File:Andulkavinkovana.jpg
Lizard	https://commons.wikimedia.org/wiki/File:Anolis_carolinensis_by_Jeff_Heard.jpg
Frog	https://eu.wikipedia.org/w/index.php?title=Fitxategi:Nanorana_conaensis,_Haa,_Bhutan.jpg&oldid=1000000
Medaka	https://commons.wikimedia.org/wiki/File:Kitano_Medaka_Male_by_Hino,_Tokyo.jpg
Sea hare	https://commons.wikimedia.org/wiki/File:Aplysia_californica_NHGRI-79108.jpg
Snail	https://commons.wikimedia.org/wiki/File:Biomphalaria_glabrata.jpg
Brachiopod	https://commons.wikimedia.org/wiki/File:Lingula_anatina_7.JPG
Acorn worm	https://commons.wikimedia.org/wiki/File:Enteropneusta.png
Sea star	https://commons.wikimedia.org/wiki/File:Acanthaster_planci_in_Ko_Tao,_Thailand_1695.jpg
Sea urchin	https://commons.wikimedia.org/wiki/File:Seurchin_300.jpg
Amphioxus	https://en.wikipedia.org/w/index.php?title=Amphioxus&oldid=1000000
Opossum	http://animaldiversity.org/collections/contributors/philmeyers/classic/monodelphis/
Zebrafish	https://commons.wikimedia.org/wiki/File:Zebrafish_(26436913602).jpg

Supplemental Figures

Cell

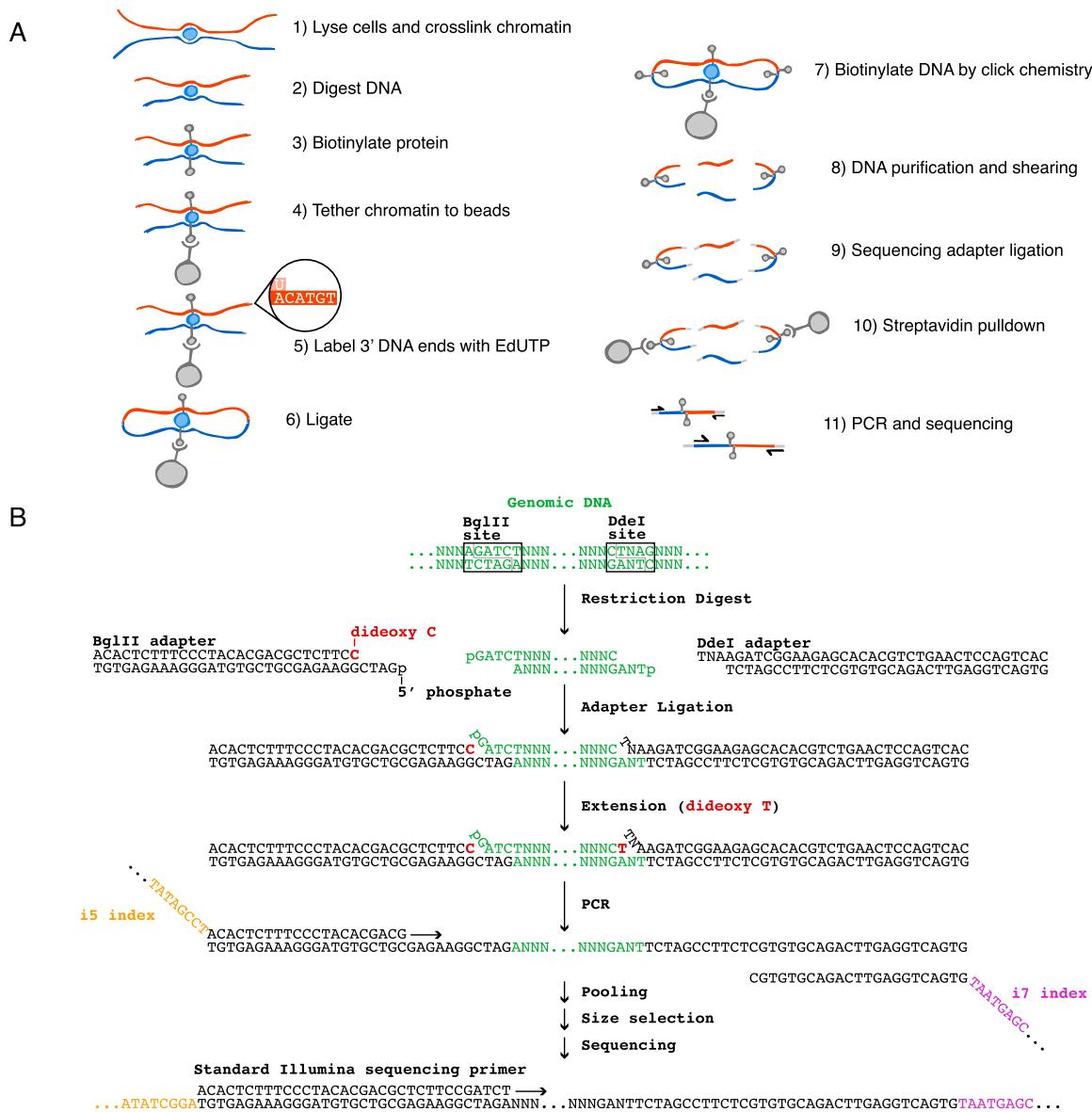


Figure S1. Overview of Modified Hi-C and ddRAD-seq Methods, Related to STAR Methods

(A) Erythrocyte Hi-C protocol. The steps are similar to a previously published method (Kalhor et al., 2011), except at steps 1, 5, and 7. Step 1 involves simultaneous instead of sequential fixation and lysis. Step 5 employs the exonuclease activity of T4 DNA polymerase to exchange the 3' terminal deoxythymidine with 5-ethynyl-dUTP (EdUTP). This avoids the use of a bulky biotin-conjugated nucleotide, and should thus increase ligation efficiency. In step 7, the 5-ethynyl uridine at the ligation junction is conjugated to biotin by Cu(I)-catalyzed cycloaddition click chemistry (Chan et al., 2004).

(B) Modified ddRAD-seq protocol. The protocol is similar to a previously published method (Peterson et al., 2012), but with a number of changes. The 3' end of a standard Illumina adaptor contains a partial BglII site (GATCT). In this protocol, BglII restriction fragments are ligated to a custom adaptor. This enables the fragments to be sequenced with a standard Illumina sequencing primer without sequencing the BglII site itself. Self-ligation of sequencing adaptors is prevented by a dideoxycytidine at the 3' end of the BglII adaptor and the lack of a 5' phosphate on the Ddel adaptor. This results in a doubly-nicked ligation product in which only the bottom strand is fully ligated. To prevent DNA polymerase from extending from the Ddel adaptor nick during PCR (which would enable amplification of unwanted inserts flanked by Ddel adaptors on both ends), an extension step is carried out with the chain terminator, dideoxy thymidine triphosphate, and Klenow fragment (3'→5' exo-) before PCR. The PCR primers contain 8 bp sequencing indexes which can be used in combinatorial fashion.

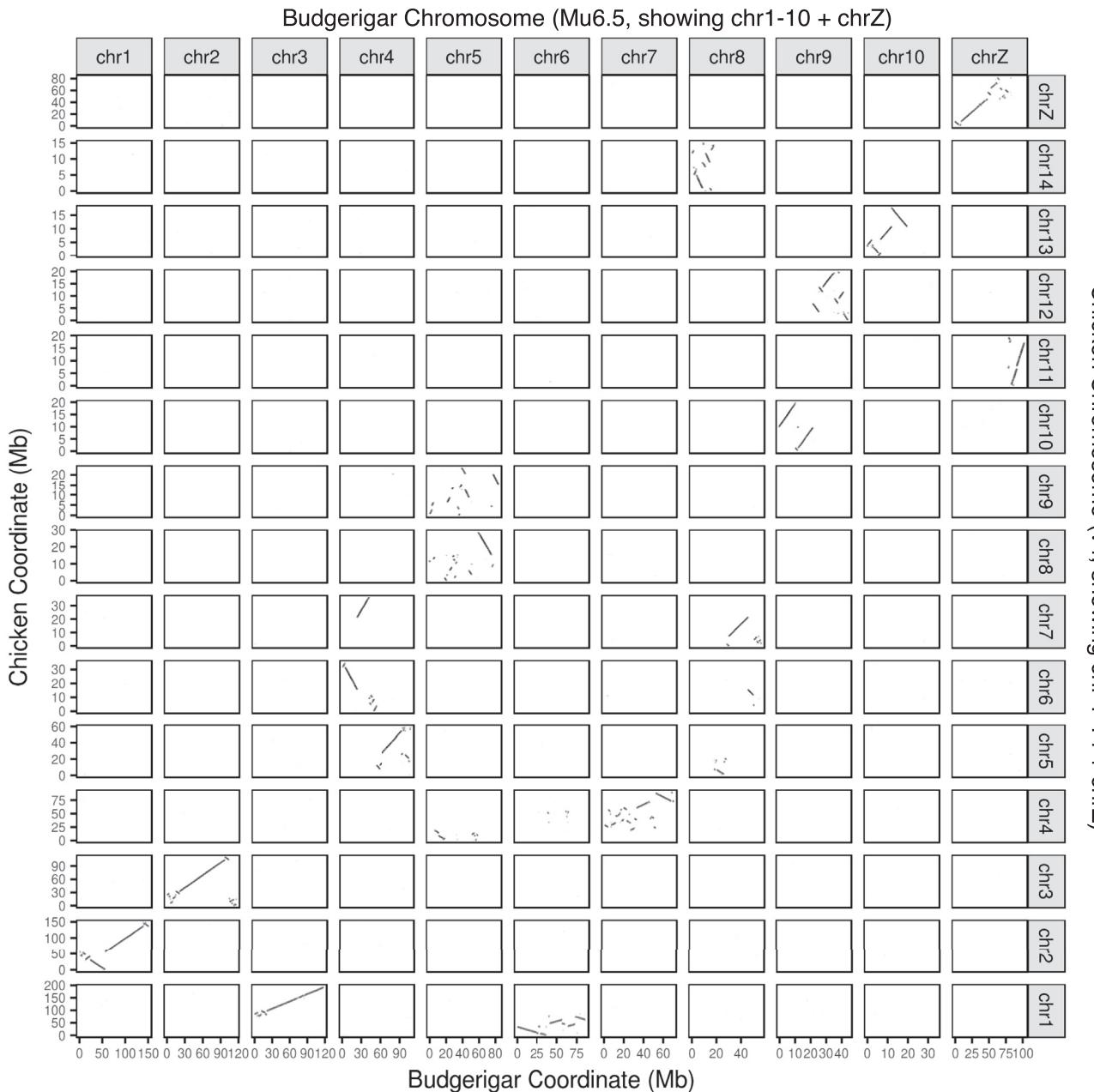


Figure S2. Alignment between Chicken and Budgerigar Genomes, Related to Figure 1

The ten largest clusters of budgerigar scaffolds shown in Figure 1 were identified relative to the budgerigar karyotype by a two-step process. First, a whole-genome alignment between budgerigar and chicken (Green et al., 2014) was lifted over into the new cluster coordinates, and represented as dot plots (shown). Next, the dot plots were compared to fluorescent in situ hybridization (FISH) data from budgerigar, with heterologous probes from flow-sorted chicken chromosomes (Nanda et al., 2007).

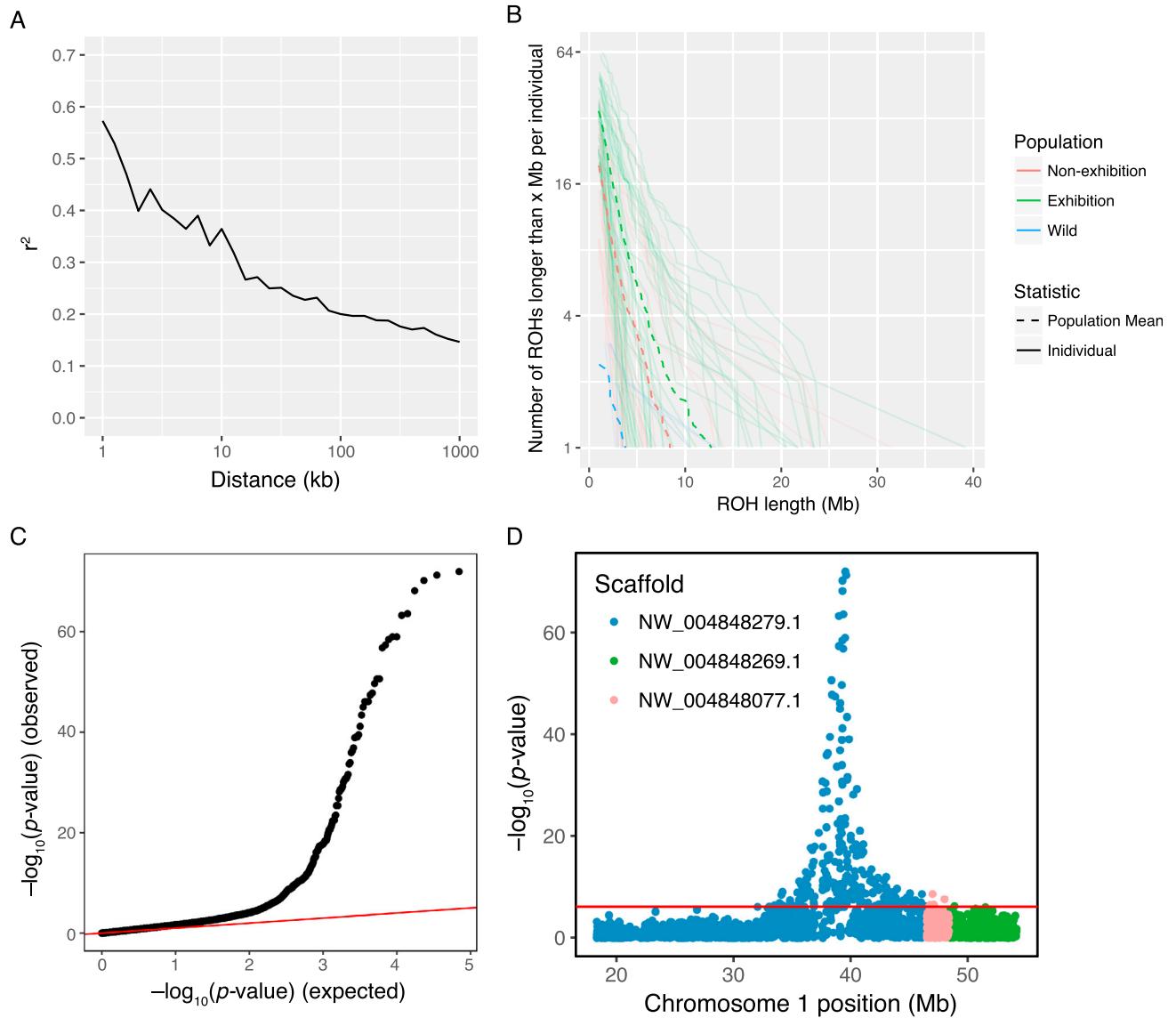


Figure S3. LD Decay in Budgerigars, Related to Figure 2

(A) LD decay curve based on mean r^2 between common SNPs (minor allele frequency ≥ 0.30).

(B) Distributions of long runs of homozygosity (ROHs) in exhibition and non-exhibition budgerigars, and museum specimens collected from the wild in Australia.

(C) Observed versus expected quantiles of the genome-wide association p values shown in Figure 2A.

(D) Association p values for scaffolds that show one or more significantly associated SNPs.

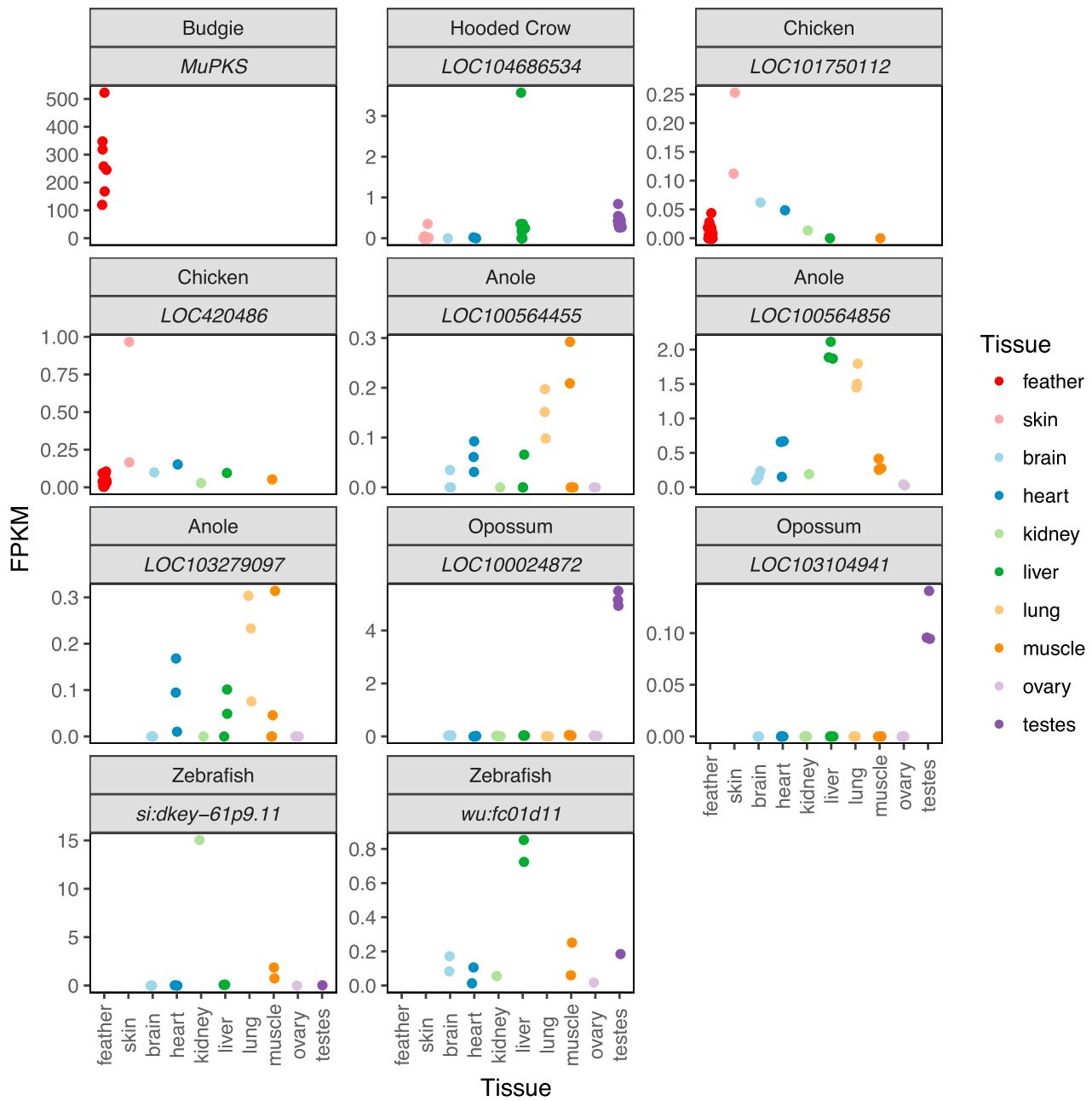
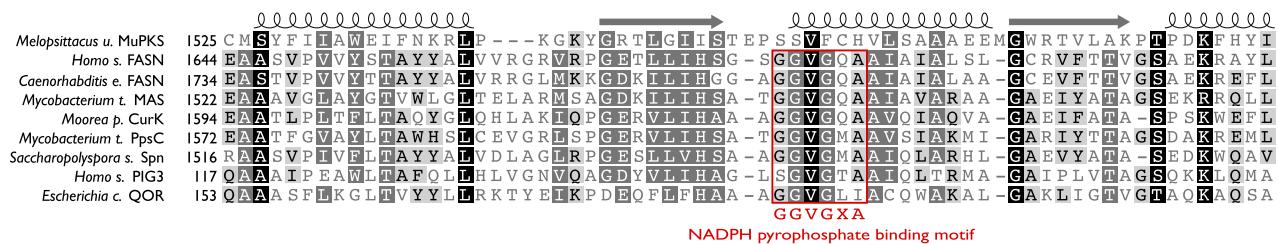


Figure S4. Expression of MuPKS Homologs in Different Species, Related to Figure 4

Published RNA-seq data for hooded crow (*Corvus cornix cornix*), chicken (*Gallus gallus*), lizard (*Anolis carolinensis*), opossum (*Monodelphis domestica*), and zebrafish (*Danio rerio*) were downloaded from the NCBI Sequence Read Archive and expression levels of MuPKS homologs in those species were measured with cuffnorm (Trapnell et al., 2013). The crow samples labeled as “skin” contained both skin and synchronized regenerating feathers, as described previously (Poelstra et al., 2014).

A

Scaffold	Position	Reference Allele	Alternative Allele	RefSeq Gene Model	Gene Model Substitution	RefSeq Protein	Protein Substitution	WT genotypes			blue genotypes			
								B0214	B0213	B0224	B0217	B0216	B0225	B0297
NW_004848279.1	21053242	T	A	XM_005152976.1	262T>A	APBB1P	Ser88Thr	1/1	./.	0/1	0/0	0/0	0/0	0/0
NW_004848279.1	21053296	G	A	XM_005152976.1	316G>A	APBB1P	Asp106Asn	1/1	./.	0/1	./.	0/0	0/0	0/0
NW_004848279.1	21369066	T	C	XM_005153087.1	1930T>C	LOC101880715	Trp644Arg	0/1	0/1	0/1	0/0	0/0	0/0	0/0

B**Figure S5. Search for the Causative Variant for the *blue* Trait, Related to Figure 5**

(A) Genotypes of WT and *blue* birds at non-synonymous SNPs located within the *blue*-shared haplotype (see Figure 3A). The genotype notations are: “0/0,” homozygous reference; “0/1,” heterozygous; “1/1,” homozygous non-reference; “./,” missing data. The genotypes were inferred from RNA-seq data. Only SNPs that exhibited genotypes consistent with a recessive Mendelian model for the *blue* trait are shown. Two of these are in *APBB1P*, which is not appreciably expressed in feathers (see Figure 4A). The third is a C to T transition (C is ancestral) at position 1930 of the CDS of *MuPKS* (*LOC101880715*) that results in an arginine to tryptophan amino acid substitution at residue 644 of MuPKS.

(B) MuPKS enoylreductase (ER) domain lacks cofactor binding site. The MuPKS ER domain shares sequence homology with several known ERs, and the more distantly-related quinone oxidoreductases (QORs) from human and bacteria. Unlike these enzymes, however, it lacks the conserved NADPH cofactor binding site, and is therefore most likely an inactive pseudo-domain.

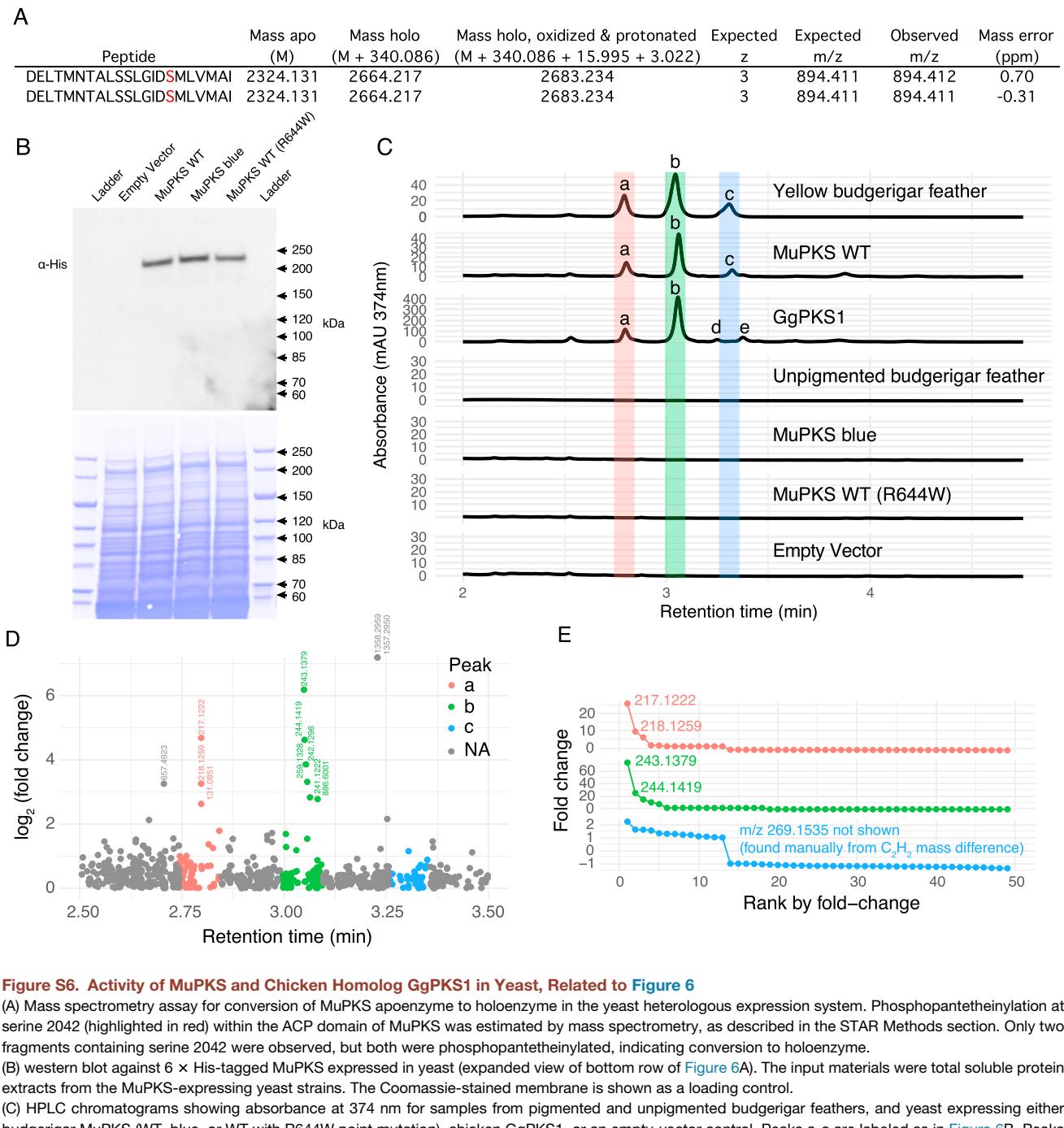


Figure S6. Activity of MuPKS and Chicken Homolog GgPKS1 in Yeast, Related to Figure 6

(A) Mass spectrometry assay for conversion of MuPKS apoenzyme to holoenzyme in the yeast heterologous expression system. Phosphopantetheinylation at serine 2042 (highlighted in red) within the ACP domain of MuPKS was estimated by mass spectrometry, as described in the STAR Methods section. Only two fragments containing serine 2042 were observed, but both were phosphopantetheinylated, indicating conversion to holoenzyme.

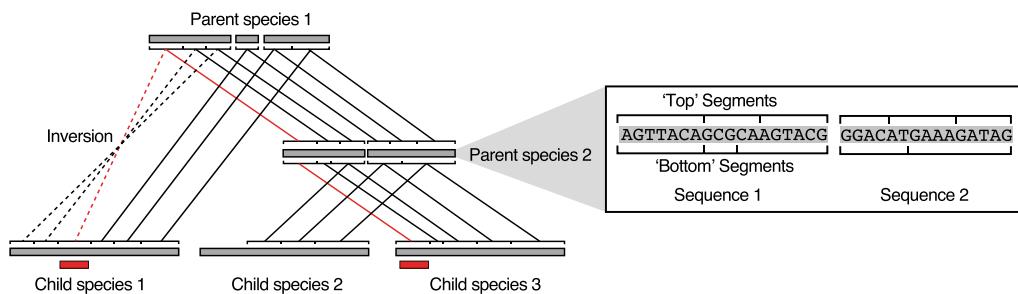
(B) western blot against 6 × His-tagged MuPKS expressed in yeast (expanded view of bottom row of Figure 6A). The input materials were total soluble protein extracts from the MuPKS-expressing yeast strains. The Coomassie-stained membrane is shown as a loading control.

(C) HPLC chromatograms showing absorbance at 374 nm for samples from pigmented and unpigmented budgerigar feathers, and yeast expressing either budgerigar MuPKS (WT, blue, or WT with R644W point mutation), chicken GgPKS1, or an empty-vector control. Peaks a-c are labeled as in Figure 6B. Peaks d and e are unique to the GgPKS1 sample.

(D) Results of untargeted search with XCMS (Smith et al., 2006) for ion signals enriched in the three pigmented samples (yellow feather, MuPKS WT, GgPKS1) versus the four unpigmented samples (the remainder). The fold-change in signal between the two sets for each observed m/z is shown as a function of retention time. Colors correspond to the shaded regions in (C).

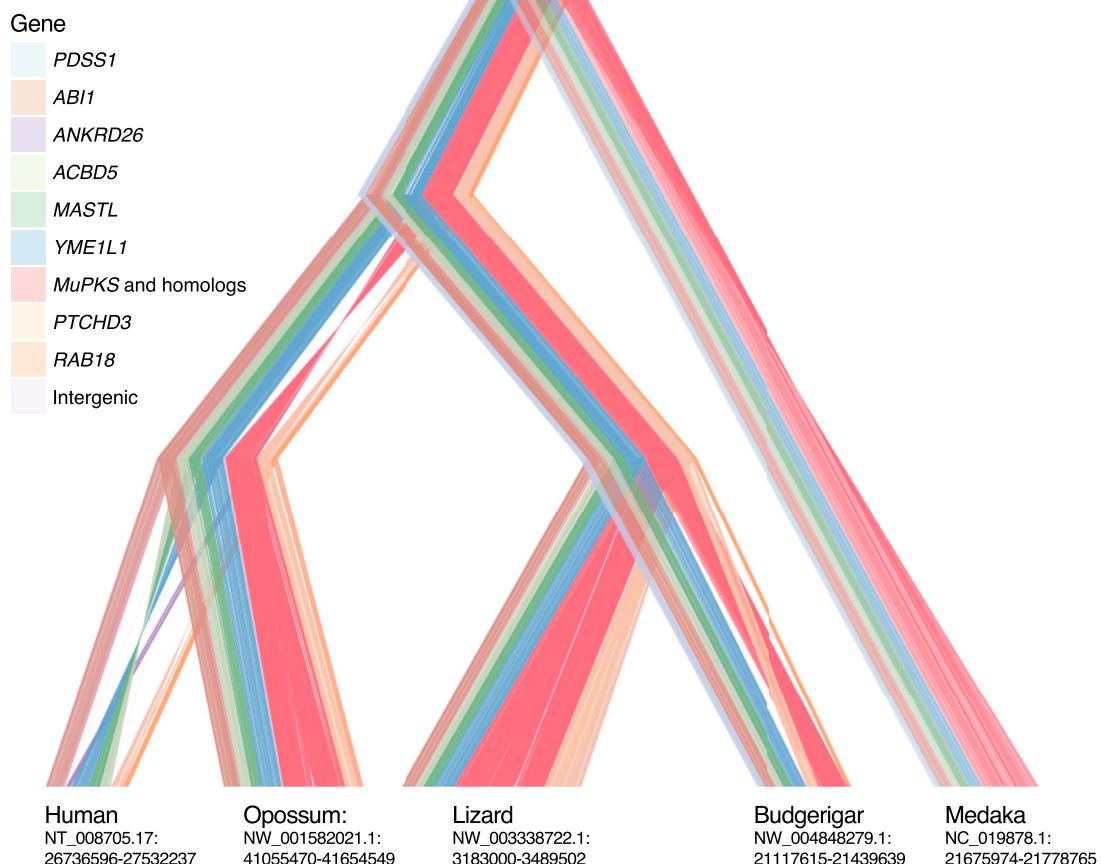
(E) Points corresponding to peaks a, b, and c, ordered by fold-change.

A

**HAL graph plotting algorithm:**

- 1) Traverse HAL graph upwards, starting from segments overlapping target intervals of interest (—).
- 2) Return all edge paths (—) that connect targets in ≥ 2 child genomes through a segment in the root parent genome.
- 3) Order and orient sequences by greedy algorithm to minimize number of edge crossings and inversions within parent-child pairs, and within parent-children trios (without this step the plotted graph can appear quite messy).
- 4) To increase legibility, compress sequences by deleting all segments not connected by one of the plotted edges.
- 5) Plot edges as polygons with widths corresponding to segment length. Plot inverted edges (---) as 'twisted' polygons (X).

B

**Figure S7. Synteny and Rearrangements near MuPKS in Vertebrates, Related to Figure 7**

- (A) Description of a new method for plotting HAL (hierarchical alignment) graphs. The HAL data structure has been described previously (Hickey et al., 2013), and contains information about synteny.
- (B) Alignment of genes near *MuPKS* in vertebrates, displayed as a HAL graph. Alignments were generated by the software package progressive-cactus (github.com/glennhickey/progressiveCactus). The RefSeq IDs for sequences used in the alignment, and the intervals displayed, are noted below the species names.