

1 A nonsense mutation in *TFEC* is the likely cause of the recessive piebald phenotype in ball
2 pythons (*Python regius*)

3 *Alan Garcia-Elfring, Heather L. Roffey, *Andrew P. Hendry, and *Rowan D. H. Barrett

4

5

6

7 *Redpath Museum and Department of Biology, McGill University, Montreal, QC, Canada

8 The corresponding author's email address: alan.garcia-elfring@mail.mcgill.ca

9

10

11

12

Abstract

Captive-bred ball pythons (*Python regius*) represent a powerful model system for studying the genetic basis of colour variation and Mendelian phenotypes in vertebrates. Although hundreds of Mendelian phenotypes (colour morphs) affecting colouration and patterning have been described for ball pythons, the genes causing these colour morphs remain unknown. Here, we used crowdsourcing of samples from commercial ball python breeders to investigate the genetic basis of a classic phenotype found in the pet trade, the piebald [characterized by dorsolateral patches of unpigmented (white) skin]. We used whole-genome sequencing of pooled samples followed by population genetic methods to delineate the genomic region containing the causal gene. We identified *TFEC* of the MIT-family of transcription factors as a candidate gene. Functional annotation of SNPs identified a nonsense mutation in *TFEC*, which we conclude is the likely causal variant for the piebald phenotype. Our work shows that ball python colour morphs have the potential to be an excellent model system for studying the genetic basis of pigment variation in vertebrates, and highlights how collaborations with commercial breeders can accelerate discoveries.

Introduction

The study of colour variation has a long history in evolutionary biology and genetics (Caro 2017; Cuthill et al. 2017; Harris et al. 2019). Over 150 years ago, crossing yellow and green peas helped Mendel to discover that traits are inherited as discrete hereditary units or “factors” (Mendel 1866; reviewed by Ellis et al. 2011). Once Mendel’s work was rediscovered in 1900, invertebrate colour variation helped establish that Mendel’s “factors”, which we now call genes, were located on chromosomes (Bridges 1919; Harrison 1920). Further, colour variation in vertebrates has provided insights in many fields, particularly the biomedical sciences (e.g. Pingault et al. 1998; Tachibana et al. 2003; Richards et al. 2019), developmental biology (e.g. Cohen et al. 2016; Haupaix et al. 2018), and evolutionary biology (e.g. Schweizer et al. 2018; Barrett et al. 2019; Burgon et al. 2020; reviewed by San-Jose and Roulin 2017). However, our current knowledge of the genetics of vertebrate colouration comes primarily from studies on model species, particularly the mouse (e.g. Baxter et al. 2004; Hoekstra 2006; San-Jose and Roulin 2017; but see Ullate-Agote et al. 2020). This bias in the literature provides a limited view of the ways in which genetic variation can affect vertebrate colouration.

Although mice, as all mammals and birds, have one pigment-producing cell type (melanocytes) which synthesizes melanin, reptiles, like fish and amphibians, have three colour-producing cells or chromatophores (reviewed by Bechtel 1978; Mills and Patterson 2009; Olsson et al. 2013). In addition to melanin-producing cells, called melanophores, reptiles have xanthophores and iridiophores that contribute to colouration. Xanthophores synthesize yellow pteridine-based pigments and can also contain red carotenoid pigments that are obtained from

diet (often called erythrophores). Iridiophores do not contain pigments but instead produce structural colouration by reflecting light on guanine nanocrystals (Teyssier et al. 2015). The three colour-producing cells interact in 3D space (Grether et al. 2004; Saenko et al. 2013) to generate the remarkable range of colour variation observed in reptiles.

We propose that the ball python (*Python regius*; Figure 1a) is an excellent model for studying the genetic basis of vertebrate colouration (Irizarry and Bryden 2016). Ball pythons are native to western Sub-Saharan Africa, with a range extending from Senegal to Uganda. With a maximum length of approximately 6 feet long, ball pythons are small relative to their better-known Asian congener, the Burmese python (*Python bivittatus*). Owing to their small size, docility, and simple husbandry, ball pythons rose in popularity in the pet trade, particularly among reptile enthusiasts. An additional factor that has contributed to the ball python's popularity was the discovery of rare phenotypes that affect colouration. In Ghana, claims of rare wild ball pythons with altered patterning go back as early as 1966. In the early 1990s, one such ball python was caught, a piebald. This individual had patches of unpigmented (white) skin (e.g. Figure 1b-d), a pigment deficiency found in many vertebrates (Ahi and Sefc 2017), including humans (Oiso et al. 2013). However, it would take a few years to rule out environmental factors as a causal explanation for the piebald phenotype in the ball python caught from the wild. By 1998, Peter Kahl had reproduced the piebald phenotype in captivity (called 'proving' by commercial breeders) and showed it had a recessive mode of inheritance (A brief history of Royal Ball Python morphs for beginners, n.d.). This pioneering work of breeders and hobbyists spurred the discovery and propagation of numerous additional and highly diverse colour morphs (e.g. Figure 1b-g). Today, 314 Mendelian phenotypes (i.e., phenotypes that

segregate as a single Mendelian factor or major quantitative trait locus) are bred in captivity (known as ‘basic’ morphs; <http://www.worldofballpythons.com/morphs>), providing a valuable resource for studying the genetics of vertebrate colouration in great detail. Indeed, commercial breeders have taken advantage of epistasis and produced over 6000 unique phenotypes (‘designer morphs’) with different combinations of the Mendelian colour factors.

Ball pythons therefore provide an outstanding opportunity to study the genetic basis of Mendelian phenotypes related to pigmentation and patterning. For example, ball python colour morphs can be characterized by changes to pattern or to pigmentation or both. Phenotypes with changes to pigmentation but not pattern (Figure 1c) are presumably determined by changes to genes with roles in a pigment synthesis pathway. On the other hand, aberrations to pattern, like white-spotting, are thought to occur through mutations in genes with functions in progenitor cell migration and differentiation from the neural crest (reviewed by Mills and Patterson 2009). However, most research on colouration, vertebrate and invertebrate, has focused on pigmentation, particularly melanin-based pigmentation (e.g. San-Jose and Roulin 2018), rather than on patterning (Mills and Patterson 2009). The extensive phenotypic variation in pigment and pattern found in captive-bred ball pythons can therefore help correct the bias (e.g. Ullate-Agote et al. 2020) associated with mammalian models and melanin-based pigmentation in the scientific literature on vertebrate colouration.

In this study, our goal was to discover the candidate mutation(s) that produce the Mendelian phenotype affecting pattern in ball pythons, the piebald (Figure 1b). We used crowdsourcing of samples from commercial breeders, thereby bridging a gap between

academia and industry. Genes known to cause white-spotting in other vertebrates, like *KIT*, *MITF*, *EDNRB*, and *SOX10* (Fleischman et al. 1991; Baxter et al. 2004; Ahi and Sefc 2017) represent *a priori* candidate genes, but given the differences in pigmentation development between reptiles and previously studied model systems, we employed a whole-genome approach to facilitate discovery of unknown candidate genes. We sequenced whole-genomes of pooled samples and then applied population genetic methods to delineate the genomic region of interest and identify candidate genes. We then functionally annotated SNPs to identify the putative causal mutation. Given the character of the piebald phenotype (i.e., dorsolateral patches of unpigmented skin), we expected the candidate gene to be expressed in the neural crest and associated with chromatophore migration and differentiation.

Methods

Sampling and DNA extraction

We obtained ball python samples by crowd-sourcing from commercial breeders in Canada (Mutation Creation, T. Dot Exotics, The Ball Room Canada, Designing Morphs), who supplied us with shed skin. We then used shed skin samples for 47 piebald individuals (i.e., individuals homozygous for the piebald variant; Table S1) and 52 non-piebald individuals (Table S2). Although individuals from both sets of samples contained additional basic morphs, the only consistent difference between the two was the piebald phenotype. We attempted to maximize the number of individuals that come from different families to minimize the effects of population structure, although there were some exceptions (Table S2). From each sample we

used approximately 0.1 g of shed skin, cut to small pieces using scissors, for DNA extraction. We extracted DNA following a standard phenol-chloroform procedure, with a 24-hour proteinase-K incubation time at 37 °C. Piebald and non-piebald samples were prepared on different working days to avoid contamination. We quantified all samples using a Picogreen® ds DNA assay (Thermo Fisher Scientific, Waltham, USA) on an Infinite® 200 Nanoquant (Tecan Group Ltd. Männedorf, Switzerland).

Sequencing

After DNA extraction, we mixed DNA of individuals (according to phenotype) in equimolar amounts to obtain a single pool for each phenotype, ‘piebald’ and ‘non-piebald.’ We used PCR-based whole-genome libraries for both pools, which were prepared at the McGill University and Genome Quebec Innovation Center, Montreal, Canada. We sequenced 150 bp paired-end reads using two lanes of Illumina HiSeqX.

Bioinformatics

To process raw reads, we applied filters based on read quality and length, keeping reads with a minimum quality of 20 (--quality-threshold 20) and a length of 50 bp (--min-length 50). We then aligned processed reads to the Burmese python (*Python bivittatus*) draft assembly Pmo2.0 (Castoe et al. 2013) using the program *NextGenMap* (Sedlazeck et al. 2013). *NextGenMap* was designed for aligning reads to highly polymorphic genomes or genomes of closely related species. We used *SAMtools* (Li et al. 2009) to convert SAM files to BAM format and remove reads with mapping quality below 20 (samtools view -q 20). We filtered for PCR duplicates using the program *MarkDuplicates* of Picard Tools (Wysoker et al. 2013). We then created a mpileup

file (samtools mpileup -B) from which the synchronized (sync) file was produced using *Popoolation2* (Kofler et al. 2011b). The sync file contains read counts for all nucleotides sequenced in the genome and it is used for subsequent downstream analyses (e.g. F_{ST} scan). We also applied the same protocol as above but instead aligned reads to the chromosome-length Burmese python reference genome, Python_molurus_bivittatus-5.0.2_HiC.assembly (Dudchenko et al. 2017, 2018).

Identification of fixed SNPs and delineation of the candidate genomic region

We applied an F_{ST} scan to identify fixed SNPs across the two pools. For this procedure, we used the *fst-sliding.pl* script of *Popoolation2* (--min-count 10, --min-coverage 20, --max-coverage 500, --min-covered-fraction 0, --window-size 1, --step-size 1, --pool-size 47:52, --suppress-noninformative). We then identified candidate SNPs as those with F_{ST} estimates of 1 and mapped them to genes. We used a custom BASH script to map fixed SNPs to genes in the gene annotation file using scaffold and SNP position. Because the draft assembly is highly fragmented (Castoe et al. 2013), we also applied the same F_{ST} scan on data aligned to the chromosome-length genome assembly (Dudchenko et al. 2017, 2018) to get a better delineation of the genomic region of interest within candidate SNPs. However, this latter assembly is not annotated with genetic features – hence necessitating the use of both assemblies.

Functional annotation of SNPs

We next functionally annotated genome-wide SNPs with the software snpEff (Cingolani et al. 2012) to aid in identifying the putative causal mutation for the piebald phenotype. SnpEff was

designed for annotating and predicting the effects of SNPs, such as amino acid changes. This program provides an assessment of the impact of a SNP, including 'HIGH' (e.g. stop codon), 'MODERATE' (e.g. non-synonymous change), 'LOW' (e.g. synonymous change), or 'MODIFIER' (change in an intergenic area).

Results

SNP dataset

From the BAM files, we obtained an average depth of coverage of 50.5 and 52.6 for the piebald and non-piebald pools, respectively. After filtering, our dataset consisted of 3,095,304 SNPs for the draft assembly and 3,221,285 SNPs in the chromosome-length assembly.

F_{ST} scan for fixed SNPs and candidate genomic region

Across all SNPs, we found an average F_{ST} of 0.03456 – relatively low, indicating that population structure was minimized. Using the draft assembly, we identified 129 fixed SNPs ($F_{ST} = 1.0$) and 369 SNPs with $F_{ST} > 0.9$. Plotting the F_{ST} values (Figure S1) shows peaks across the reference genome, demonstrating the fragmented state of the draft assembly. In the chromosome-length assembly, we found 131 fixed SNPs ($F_{ST} = 1.0$) and 372 SNPs with $F_{ST} > 0.9$. The chromosome-length assembly also shows that 128/131 fixed SNPs and 365/372 SNPs with $F_{ST} > 0.9$ map to a single region on scaffold 7, clearly delineating a genomic region of interest (Figure 2). The 128 fixed SNPs map to a region 8 Mb long (scaffold 7: 49526089-57612101). This region becomes 26.5 Mb when considering the 365 SNPs with $F_{ST} > 0.9$ (scaffold 7: 49246858-76754467).

However, most of these 365 SNPs cluster within the 8 Mb region demarcated by fixed SNPs, with only 11 mapping beyond the bounds of this region (three upstream, eight downstream).

Candidate genes and SNP annotation

We found that 59 fixed SNPs map to 12 different protein-coding genes (*BMT2*, *CAPZA2*, *CNTN1*, *DOCK4*, *FOXP2*, *GPR85*, *KCND2*, *LOC103067393*, *LSMEM1*, *ST7*, *TES*, and *TFEC*). We annotated SNPs for predicted impact (see methods) and found that all fixed SNPs have the lowest impact classification (MODIFIER), often assigned to mutations in non-functional regions, and most map to introns. However, we found other impact classifications among SNPs with F_{ST} greater than 0.9. One synonymous (i.e., silent) SNP (LOW impact, NW_006539084.1 26867, $F_{ST} = 0.92$) is located within the gene *LSMEM1* and one nonsense SNP (HIGH impact, NW_006534020.1 160458, $F_{ST} = 0.96$) is located within the gene *TFEC*. This nonsense variant consisted of a C>T mutation on Arg165, resulting in a premature Opal stop codon (Table S3).

Discussion

We regard the ball python as a powerful new model for understanding the genetic basis of pigmentation in vertebrates. As proof-of-principle, we set out to discover candidate genes and causal mutations for the classic colour morph found in the pet trade, the piebald. Over the last thirty years, ball python breeders have discovered and successfully propagated in captivity more than three hundred Mendelian phenotypes. Yet, researchers working on the genetics of vertebrate colouration have been slow to appreciate the ball python as a powerful new model

(but see Irizarry and Bryden 2016). We set out to bridge this gap. To do so, we appealed to commercial ball python breeders to join our efforts in discovery, starting with uncovering the genetic basis of the piebald phenotype. Commercial breeders have a vested interest in acquiring the ability to identify heterozygotes for recessive phenotypes (i.e., carriers of recessive alleles affecting colouration), and piebald is one such phenotype, although subtle phenotypic differences on heterozygotes are not uncommon (e.g. ring of unpigmented skin, often near the vent).

To elucidate the genetic basis of the piebald phenotype, we combined a case-control experimental design with whole-genome sequencing, population genetic methods, and functional annotation of the candidate SNPs. We identified a premature stop codon in the protein-coding region of the gene *TFEC* as the putative genetic basis for this Mendelian phenotype. *TFEC* belongs to a family of transcription factors (*MITF*) known to be expressed in the neural crest and pigment cells. However, functional validation through gene editing techniques, like CRISPR cas9 (Rasys et al. 2019), is required to establish that the candidate nonsense mutation is the causal variant.

With knowledge of the specific genetic variants underlying pigmentation variation in ball pythons, gene editing could pave the way to produce colour morphs free of secondary phenotypes that are undesired from a commercial or ethical perspective. Genes that contribute to colouration are also expressed in many other tissues derived from neural crest cells (Donoghue et al. 2008). Thus, mutations that alter colour or patterning often alter other traits, such as behaviour (Ducrest et al. 2008; McKinnon and Pierotti 2010). Not surprisingly, there are

ball python colour morphs linked to morphological deformities and neurological dysfunction (Rose and Williams 2014). For example, the incomplete dominant ‘spider’ morph is associated with a neurological syndrome called the ‘head wobble’ (Fox and Hogan 2020). The propagation of colour morphs with correlated traits have implications for the health and welfare of ball pythons in captivity (Rose and Williams 2014) and are a source of current controversy among hobbyists and commercial breeders (Riera 2015). In 2017, the International Herpetological Society in the U. K. banned ball pythons with the spider mutation, which is thought to be homozygous lethal, from their reptile shows (International Herpetological Society 2020). However, it is not known if the correlated trait(s) for any particular colour morph is due to pleiotropic effects or linked variation. If due to linked variation, targeted mutation on a different genetic background (i.e., without the linked variant) could allow the reproduction of colour phenotypes free of behavioural and morphological abnormalities. If due to pleiotropy, targeted mutation at different nearby loci could achieve the same effect.

The wide variety of ball python phenotypes found in the pet trade offers the opportunity to study how genes control vertebrate pigmentation and pattern formation. Changes to pigmentation are often caused by mutations in genes involved in the pigment synthesis pathway (Mills and Patterson 2009; Hubbard et al. 2010). For example, albinism occurs when an individual inherits two copies of a gene involved in melanin synthesis (e.g. OCA2) that is non-functional. Snakes with two non-functional OCA2 genes lack dark melanin-based pigment though contain yellow pigments produced in the xanthophores (Saenko et al. 2015). Changes to pattern, on the other hand, are often attributed to genes with functions in cell migration from the neural crest and differentiation into colour-producing cells (Mills and

Patterson 2009). However, Mendelian phenotypes characterized by changes to either pigmentation or patterning are likely to map to the protein-coding regions of genes. Genomic research on the most extensively studied vertebrate, humans, shows that Mendelian phenotypes are generally caused by mutations arising in protein-coding regions and affecting protein function (Cooper et al. 2010; Chong et al. 2015). The progenitor cells of chromatophores originate in the neural crest and mutations to these genes result in partial or total loss of pigment and structural colouration. Mutations to genes such as *MITF*, *KIT*, *EDNRB*, and *SOX10*, which are expressed in the neural crest, results in white spotting or a complete lack of pigment across a wide range of vertebrates, including fish, amphibians, birds, and mammals (Baxter et al. 2004; Kelsh 2006; George et al. 2016; Ahi and Sefc 2017; Woodcock et al. 2017; Camargo-Sosa et al. 2019; Goding and Arnheiter 2019). These genes represented *a priori* candidates for the piebald phenotype in ball pythons, but did not show differentiation between our phenotypic pools.

We identified a putative causal mutation for the piebald phenotype in ball pythons located in a new candidate gene for vertebrate pigmentation: the gene *TFEC* (Figure 3, Table S3). This variant, which produces a premature stop codon, was nearly fixed ($F_{ST} = 0.96$). Complete differentiation between the piebald and non-piebald pools was prevented by a single read for the reference allele that was sequenced in the piebald samples. It is possible this is the result of a sequencing error or the inclusion of an unknown heterozygote. *TFEC* was not one of the candidate genes expected *a priori*. This particular gene has not been implicated in mammalian colouration. However, *TFEC* is part of the MIT-family of transcription factors known to be involved in animal patterning. For example, this family consists of the gene *MITF*, which

forms heterodimers with the other three members, *TFEB*, *TFE3*, and *TFEC* (Goding and Arnheiter 2019). In mouse models, double knockout mutations to *MITF* and *TFE3* cause all-white phenotypes (Steingrímsson et al. 2002). Interestingly, the same study showed that knocking out *TFEC* did not result in a phenotype different from the wild-type mouse. There is evidence that shows *TFEC* is expressed in the retinal pigment epithelium of mammals (Rowan et al. 2004). No evidence suggests, however, that *TFEC* is expressed in cells that descend from the mammalian neural crest, which includes pigment-producing cells. In contrast, recent evidence from the zebrafish model shows that *TFEC* is highly expressed in neural crest cells, specifically progenitor pigment cells and iridophores (Petratou et al. 2019; Saunders et al. 2019). In fact, *TFEC* is expressed in two dorsolateral patches on the torso during embryogenesis (Lister et al. 2011; Petrato et al. 2018; Petrato et al. 2019), reminiscent of where the white-spotting occurs in the piebald phenotype in ball pythons. The expression patterns of *TFEC* in zebrafish, together with our results, makes a strong case that *TFEC* is the piebald gene in ball pythons and that a premature stop codon is the likely causal mutation.

Conclusion

In this study, we aimed to uncover, for the first time, the genetic basis of a ball python colour morph. We discovered that a premature stop codon in the gene *TFEC* is the likely cause for the Mendelian piebald phenotype found in ball pythons from the pet trade. The lack of evidence for *TFEC* affecting mammalian pigmentation highlights the need to study the genetic basis of colouration in non-mammalian vertebrate models. Future work will be aimed at demonstrating

causality through direct functional manipulation. These efforts would produce the first genetically engineered ball python, a step that would open the door to safe reproduction of other colour morphs through targeted mutation.

Acknowledgments

RDHB was supported by an NSERC Discovery Grant and Canada Research Chair. We thank Mutation Creation, T Dot Exotics, The Ball Room Canada, and Designing Morphs for supplying samples.

Data accessibility

The data that support the findings of this manuscript will be uploaded to Dryad.

Contributions

AGE, HR, APH, and RDHB conceived and designed the study. HR collected samples. AGE performed the molecular work. AGE performed the bioinformatics and analyzed the genomic data. AGE wrote the manuscript with input from all authors.

REFERENCES

- Ahi, E.P. and Sefc, K.M., 2017. A gene expression study of dorso-ventrally restricted pigment pattern in adult fins of *Neolamprologus meeli*, an African cichlid species. *PeerJ*, 5, p.e2843.
- Barrett, R.D., Laurent, S., Mallarino, R., Pfeifer, S.P., Xu, C.C., Foll, M., Wakamatsu, K., Duke-Cohan, J.S., Jensen, J.D. and Hoekstra, H.E., 2019. Linking a mutation to survival in wild mice. *Science*, 363(6426), pp
- Baxter, L.L., Hou, L., Loftus, S.K. and Pavan, W.J., 2004. Spotlight on spotted mice: a review of white spotting mouse mutants and associated human pigmentation disorders. *Pigment Cell Research*, 17(3), pp.215-224..499-504.
- Bechtel, H.B., 1978. Color and pattern in snakes (Reptilia, Serpentes). *Journal of Herpetology*, pp.521-532.
- Bridges, C.B., 1919. The genetics of purple eye color in *Drosophila*. *Journal of Experimental Zoology*, 28(2), pp.265-305.
- A brief history of Royal Ball Python morphs for beginners. Retrieved from <http://bagnallsreptiles.weebly.com/history-into-morphs.html>
- Burgon, J.D., Vieites, D.R., Jacobs, A., Weidt, S.K., Gunter, H.M., Steinfartz, S., Burgess, K., Mable, B.K. and Elmer, K.R., 2020. Functional colour genes and signals of selection in colour-polymorphic salamanders. *Molecular ecology*, 29(7), pp.1284-1299.
- Camargo-Sosa, K., Colanesi, S., Müller, J., Schulte-Merker, S., Stemple, D., Patton, E.E. and Kelsh, R.N., 2019. Endothelin receptor Aa regulates proliferation and differentiation of Erb-dependent pigment progenitors in zebrafish. *PLoS genetics*, 15(2), p.e1007941.
- Caro, T., 2017. Wallace on coloration: contemporary perspective and unresolved insights. *Trends in ecology & evolution*, 32(1), pp.23-30.
- Castoe, T.A., de Koning, A.J., Hall, K.T., Card, D.C., Schield, D.R., Fujita, M.K., Ruggiero, R.P., Degner, J.F., Daza, J.M., Gu, W. and Reyes-Velasco, J., 2013. The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. *Proceedings of the National Academy of Sciences*, 110(51), pp.20645-20650.
- Chong, J.X., Buckingham, K.J., Jhangiani, S.N., Boehm, C., Sobreira, N., Smith, J.D., Harrell, T.M., McMillin, M.J., Wiszniewski, W., Gambin, T. and Akdemir, Z.H.C., 2015. The genetic basis of Mendelian phenotypes: discoveries, challenges, and opportunities. *The American Journal of Human Genetics*, 97(2), pp.199-215.

- Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X. and Ruden, D.M., 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(2), pp.80-92.
- Cohen, M.A., Wert, K.J., Goldmann, J., Markoulaki, S., Buganim, Y., Fu, D. and Jaenisch, R., 2016. Human neural crest cells contribute to coat pigmentation in interspecies chimeras after in utero injection into mouse embryos. *Proceedings of the National Academy of Sciences*, 113(6), pp.1570-1575.
- Cooper, D.N., Chen, J.M., Ball, E.V., Howells, K., Mort, M., Phillips, A.D., Chuzhanova, N., Krawczak, M., Kehrer-Sawatzki, H. and Stenson, P.D., 2010. Genes, mutations, and human inherited disease at the dawn of the age of personalized genomics. *Human mutation*, 31(6), pp.631-655.
- Cuthill, I.C., Allen, W.L., Arbuckle, K., Caspers, B., Chaplin, G., Hauber, M.E., Hill, G.E., Jablonski, N.G., Jiggins, C.D., Kelber, A. and Mappes, J., 2017. The biology of color. *Science*, 357(6350), p.eaan0221.
- Donoghue, P.C., Graham, A. and Kelsh, R.N., 2008. The origin and evolution of the neural crest. *Bioessays*, 30(6), pp.530-541.
- Ducrest, A.L., Keller, L. and Roulin, A., 2008. Pleiotropy in the melanocortin system, coloration and behavioural syndromes. *Trends in ecology & evolution*, 23(9), pp.502-510.
- Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P. and Aiden, E.L., 2017. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*, 356(6333), pp.92-95.
- Dudchenko, O., Shamim, M.S., Batra, S., Durand, N.C., Musial, N.T., Mostofa, R., Pham, M., Hilaire, B.G.S., Yao, W., Stamenova, E., Hoeger, M., Nyquist, S.K., Korchina, V., Pletch, K., Flanagan, J.P., Tomaszewicz, A., McAloose, D., Estrada, C.P., Novak, B.J., Omer, A.D., Aiden, E.L., 2018. The Juicebox Assembly Tools module facilitates de novo assembly of mammalian genomes with chromosome-length scaffolds for under \$1000. *bioRxiv* 254797. <https://doi.org/10.1101/254797>.
- Ellis, T.N., Hofer, J.M., Timmerman-Vaughan, G.M., Coyne, C.J. and Hellens, R.P., 2011. Mendel, 150 years on. *Trends in plant science*, 16(11), pp.590-596.

- Fleischman, R.A., Saltman, D.L., Stastny, V. and Zneimer, S., 1991. Deletion of the c-kit protooncogene in the human developmental defect piebald trait. *Proceedings of the National Academy of Sciences*, 88(23), pp.10885-10889.
- Fox, B.C. and Hogan, S., 2020. Hypopigmentation as the Cause for the Neurological Disorder “Spider Wobble” in Python regius.
- George, A., Zand, D.J., Hufnagel, R.B., Sharma, R., Sergeev, Y.V., Legare, J.M., Rice, G.M., Schwoerer, J.A.S., Rius, M., Tetri, L. and Gamm, D.M., 2016. Biallelic mutations in MITF cause coloboma, osteopetrosis, microphthalmia, macrocephaly, albinism, and deafness. *The American Journal of Human Genetics*, 99(6), pp.1388-1394.
- Goding, C.R. and Arnheiter, H., 2019. MITF—the first 25 years. *Genes & development*, 33(15-16), pp.983-1007.
- Grether, G.F., Kolluru, G.R. and Nersissian, K., 2004. Individual colour patches as multicomponent signals. *Biological Reviews*, 79(3), pp.583-610.
- Harris RB, Irwin K, Jones MR, Laurent S, Barrett RD, Nachman MW, Good JM, Linnen CR, Jensen JD, Pfeifer SP. The population genetics of crypsis in vertebrates: recent insights from mice, hares, and lizards. *Heredity*. 2019 Aug 9:1-4.
- Harrison, J.H., 1920. Genetical studies in the moths of the geometrid genus *Oporabia* (*Oporinia*) with a special consideration of melanism in the lepidoptera. *Journal of Genetics*, 9(3), pp.195-280.
- Haupaix, N., Curantz, C., Bailleul, R., Beck, S., Robic, A. and Manceau, M., 2018. The periodic coloration in birds forms through a prepattern of somite origin. *Science*, 361(6408), p.eaar4777.
- Hoekstra, H.E., 2006. Genetics, development and evolution of adaptive pigmentation in vertebrates. *Heredity*, 97(3), pp.222-234.
- International Herpetological Society. (2020) *IMPORTANT NOTICE – BANNED MORPHS*. June 19, 2020. <https://www.ihs-web.org.uk/news/show-news/important-notice-banned-species/>
- Irizarry, K.J. and Bryden, R.L., 2016. In silico analysis of gene expression network components underlying pigmentation phenotypes in the Python identified evolutionarily conserved clusters of transcription factor binding sites. *Advances in bioinformatics*, 2016.
- Kelsh, R.N., 2006. Sorting out Sox10 functions in neural crest development. *Bioessays*, 28(8), pp.788-798.

397 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and
398 Durbin, R., 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*,
399 25(16), pp.2078-2079.

400 Lister, J.A., Lane, B.M., Nguyen, A. and Lunney, K., 2011. Embryonic expression of zebrafish MITF
401 family genes *tfe3b*, *tfeb*, and *tfec*. *Developmental Dynamics*, 240(11), pp.2529-2538.

402 Mills, M.G. and Patterson, L.B., 2009, February. Not just black and white: pigment pattern
403 development and evolution in vertebrates. In *Seminars in cell & developmental biology*
404 (Vol. 20, No. 1, pp. 72-81). Academic Press.

405 McKinnon, J.S. and Pierotti, M.E., 2010. Colour polymorphism and correlated characters:
406 genetic mechanisms and evolution. *Molecular ecology*, 19(23), pp.5101-5125.

407 Mendel, G. (1866) Versuche u"ber pflanzen-hybriden. *Verhandlungen der naturforschungs*
408 *Vereins* 4, 3-47

409 Oiso, N., Fukai, K., Kawada, A. and Suzuki, T., 2013. Piebaldism. *The Journal of Dermatology*,
410 40(5), pp.330-335.

411 Olsson, M., Stuart-Fox, D. and Ballen, C., 2013, June. Genetics and evolution of colour patterns
412 in reptiles. In *Seminars in cell & developmental biology* (Vol. 24, No. 6-7, pp. 529-541).
413 Academic Press.

414 Petratos, K., Subkhankulova, T., Lister, J.A., Rocco, A., Schwetlick, H. and Kelsh, R.N., 2018. A
415 systems biology approach uncovers the core gene regulatory network governing
416 iridophore fate choice from the neural crest. *PLoS genetics*, 14(10), p.e1007402.

417 Petratos, K., Spencer, S.A., Kelsh, R.N. and Lister, J.A., 2019. The MITF paralog *tfec* is required in
418 neural crest development for fate specification of the iridophore lineage from a
419 multipotent pigment cell progenitor. *bioRxiv*, p.862011.

420 Pingault, V., Bondurand, N., Kuhlbrodt, K., Goerich, D.E., Pr"ehu, M.O., Puliti, A., Herbarth, B.,
421 Hermans-Borgmeyer, I., Legius, E., Matthijs, G. and Amiel, J., 1998. SOX10 mutations in
422 patients with Waardenburg-Hirschsprung disease. *Nature genetics*, 18(2), pp.171-173.

423 Rasys, A.M., Park, S., Ball, R.E., Alcala, A.J., Lauderdale, J.D. and Menke, D.B., 2019. CRISPR-Cas9
424 Gene Editing in Lizards through Microinjection of Unfertilized Oocytes. *Cell reports*,
425 28(9), pp.2288-2292.

426 Richards, J.R., Yoo, J.H., Shin, D. and Odelberg, S.J., 2019. Mouse models of uveal melanoma:
427 Strengths, weaknesses, and future directions. *Pigment Cell & Melanoma Research*.

428 Riera, D., 2015. BALL PYTHON GENETICS: MORPH COMBINATIONS TO BE CAUTIOUS OF.
429 <https://www.rierareptiles.com/ball-python-genetics-anomalies/>

430 Rose, M.P. and Williams, D.L., 2014. Neurological dysfunction in a ball python (*Python regius*)
431 colour morph and implications for welfare. *Journal of Exotic Pet Medicine*, 23(3),
432 pp.234-239.

433 Rowan, S., Chen, C.M.A., Young, T.L., Fisher, D.E. and Cepko, C.L., 2004. Transdifferentiation of
434 the retina into pigmented cells in ocular retardation mice defines a new function of the
435 homeodomain gene *Chx10*. *Development*, 131(20), pp.5139-5152.

436 San-Jose, L.M. and Roulin, A., 2017. Genomics of coloration in natural animal populations.
437 *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1724),
438 p.20160337.

439 San-Jose, L.M. and Roulin, A., 2018. Toward understanding the repeated occurrence of
440 associations between melanin-based coloration and multiple phenotypes. *The American*
441 *Naturalist*, 192(2), pp.111-130.

442 Saenko, S.V., Teyssier, J., Van Der Marel, D. and Milinkovitch, M.C., 2013. Precise colocalization
443 of interacting structural and pigmentary elements generates extensive color pattern
444 variation in *Phelsuma* lizards. *BMC biology*, 11(1), p.105.

445 Saenko, S.V., Lamichhaney, S., Barrio, A.M., Rafati, N., Andersson, L. and Milinkovitch, M.C.,
446 2015. Amelanism in the corn snake is associated with the insertion of an LTR-
447 retrotransposon in the *OCA2* gene. *Scientific reports*, 5, p.17118.

448 Saunders, L.M., Mishra, A.K., Aman, A.J., Lewis, V.M., Toomey, M.B., Packer, J.S., Qiu, X.,
449 McFaline-Figueroa, J.L., Corbo, J.C., Trapnell, C. and Parichy, D.M., 2019. Thyroid
450 hormone regulates distinct paths to maturation in pigment cell lineages. *Elife*, 8,
451 p.e45181.

452 Schweizer, R.M., Durvasula, A., Smith, J., Vohr, S.H., Stahler, D.R., Galaverni, M., Thalmann, O.,
453 Smith, D.W., Randi, E., Ostrander, E.A. and Green, R.E., 2018. Natural selection and
454 origin of a melanistic allele in North American gray wolves. *Molecular biology and*
455 *evolution*, 35(5), pp.1190-1209.

456 Sedlazeck, F.J., Rescheneder, P. and Von Haeseler, A., 2013. NextGenMap: fast and accurate
457 read mapping in highly polymorphic genomes. *Bioinformatics*, 29(21), pp.2790-2791.

458 Steingrímsson, E., Tessarollo, L., Pathak, B., Hou, L., Arnheiter, H., Copeland, N.G. and Jenkins,
459 N.A., 2002. *Mitf* and *Tfe3*, two members of the *Mitf*-*Tfe* family of bHLH-Zip transcription

460 factors, have important but functionally redundant roles in osteoclast development.
461 Proceedings of the National Academy of Sciences, 99(7), pp.4477-4482.

462 Tachibana, M., Kobayashi, Y. and Matsushima, Y., 2003. Mouse models for four types of
463 Waardenburg syndrome. *Pigment cell research*, 16(5), pp.448-454.

464 Teyssier J, Saenko SV, Van Der Marel D, Milinkovitch MC. Photonic crystals cause active colour
465 change in chameleons. *Nature communications*. 2015 Mar 10;6:6368.

466 Ullate-Agote, A., Burgelin, I., Debry, A., Langrez, C., Montange, F., Peraldi, R., Daraspe, J.,
467 Kaessmann, H., Milinkovitch, M.C. and Tzika, A.C., 2020. Genome mapping of a LYST
468 mutation in corn snakes indicates that vertebrate chromatophore vesicles are lysosome-
469 related organelles. *Proceedings of the National Academy of Sciences*, 117(42),
470 pp.26307-26317.

471 Woodcock, M.R., Vaughn-Wolfe, J., Elias, A., Kump, D.K., Kendall, K.D., Timoshevskaya, N.,
472 Timoshevskiy, V., Perry, D.W., Smith, J.J., Spiwak, J.E. and Parichy, D.M., 2017.
473 Identification of mutant genes and introgressed tiger salamander DNA in the laboratory
474 axolotl, *Ambystoma mexicanum*. *Scientific reports*, 7(1), pp.1-10.

475 Wysocki, A., Tibbetts, K. and Fennell, T., 2013. Picard tools version 1.90. [http://picard.](http://picard.sourceforge.net)
476 [sourceforge. net](http://picard.sourceforge.net) (Accessed 14 December 2016), 107(17), p.308.

477

478 Figures

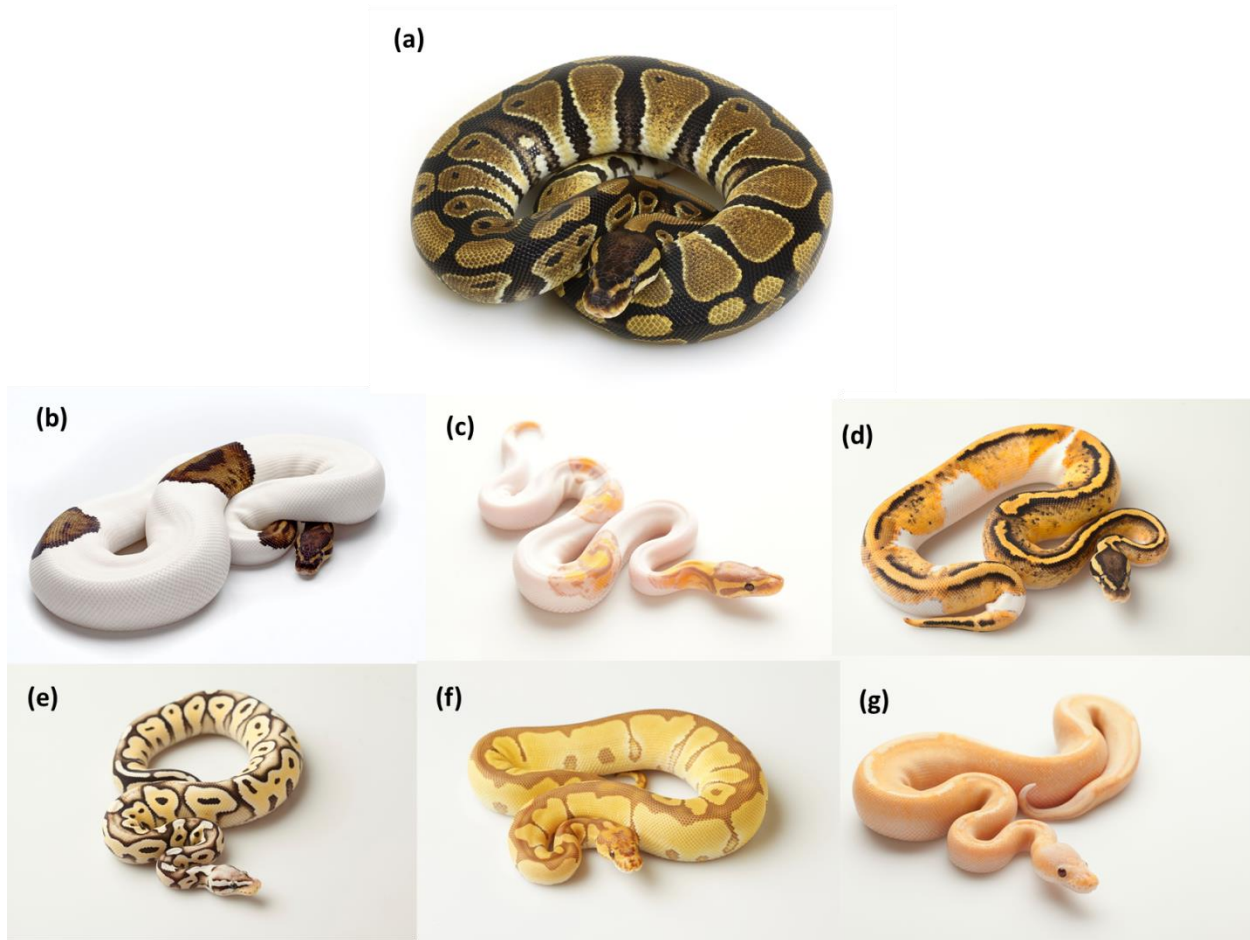


Figure 1. A small sample of the phenotypic variation found in captive-bred ball pythons (*Python regius*).
 (a) wild-type, (b) piebald, (c) banana piebald, (d) pastel piebald, (e) pastel HRA enhancer, (f) ultramel
 clown, (g) banana champagne. Photo credit (b-g): Designing Morphs.

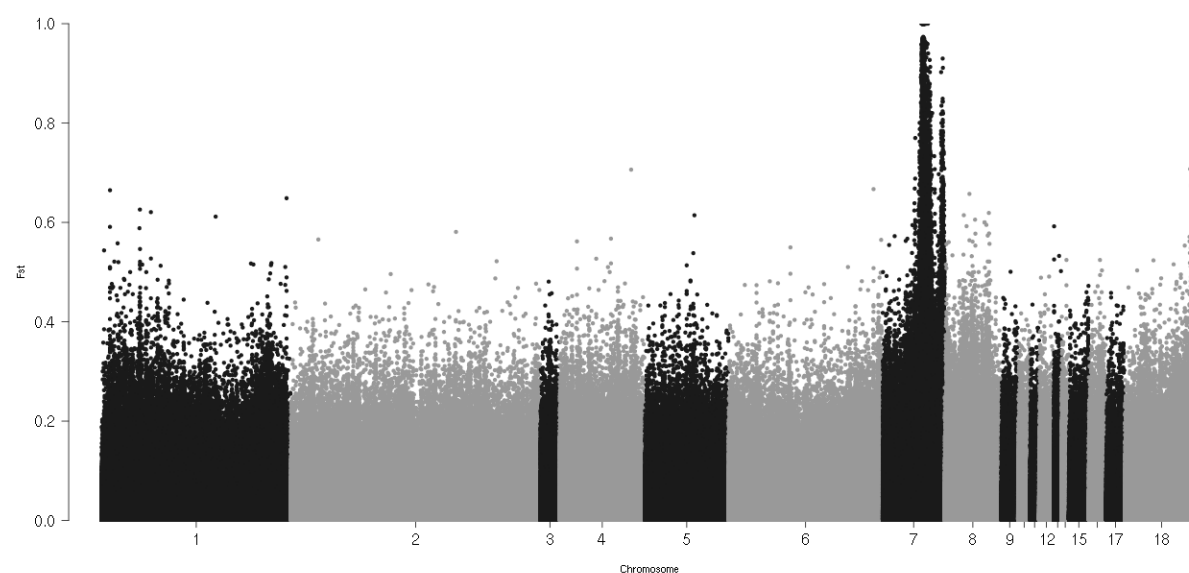


Figure 2. F_{st} plot between piebald and non-piebald samples on the chromosome-length assembly (Python_molurus_bivittatus-5.0.2_HiC.assembly).

492

5' GAT ACA CGA GCA 3'
Asp Thr Arg Ala

C>T Arg165

5' GAT ACA TGA GCA 3'
Asp Thr **Stop** Ala

493

494 Figure 3. A premature stop codon on Arg165 is the
495 putative causal variant for the piebald phenotype.

496

497

498

Supplementary figures

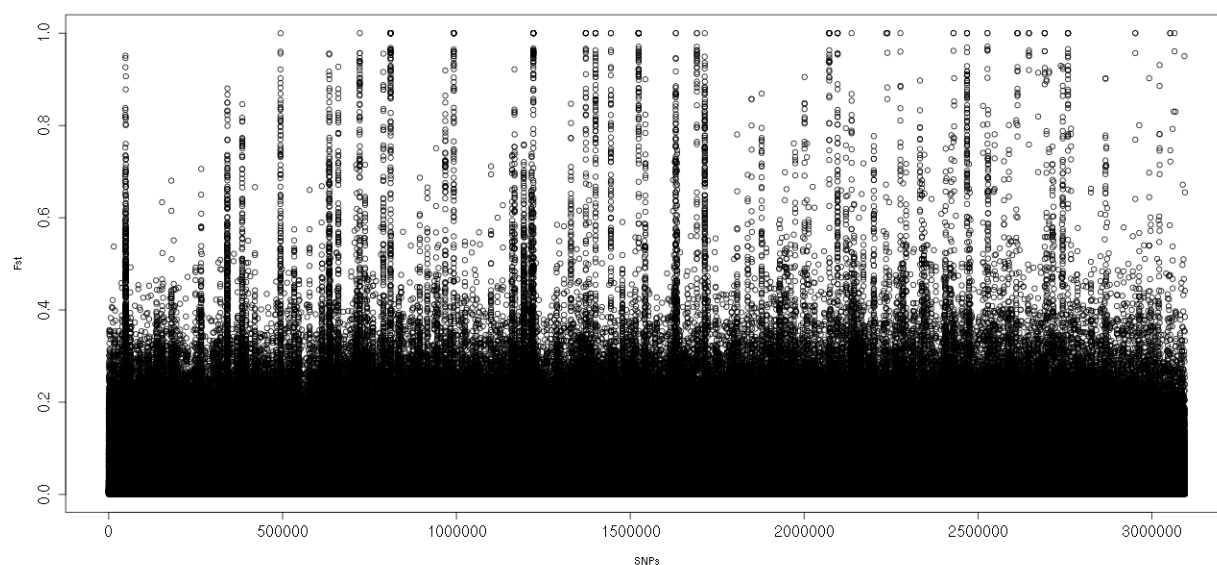


Figure S1. Fst plot of SNPs mapped to the Burmese python draft assembly (Pmo2.0). The 129 fixed SNPs in this plot were used to map to candidate genes.