2017-05-25

Comparative Genomics of Mechanisms Underlying Adaptation to High-Altitude Environments in Andean Waterfowl

Allie Marie Graham
University of Miami, graham.allie@gmail.com

UNIVERSITY OF MIAMI

COMPARATIVE GENOMICS OF MECHANISMS UNDERLYING ADAPTATION TO HIGH-ALTITUDE ENVIRONMENTS IN ANDEAN WATERFOWL

By

Allie Marie Graham

A DISSERTATION

Submitted to the Faculty
of the University of Miami
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy

Coral Gables, Florida

August 2017

UNIVERSITY OF MIAMI

A dissertation submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

COMPARATIVE GENOMICS OF MECHANISMS UNDERLYING ADAPTATION TO HIGH-ALTITUDE ENVIRONMENTS IN ANDEAN WATERFOWL

Allie Marie Graham

Approved:

_____
Kevin McCracken, Ph.D.
Kushlan Chair and Associate
Professor of Biology

_____
Barbara Whitlock, Ph.D.
Associate Professor of Biology

_____
William Browne, Ph.D.
Assistant Professor of Biology

_____
Marjorie Oleksiak, Ph.D.
Associate Professor of Marine
Biology and Ecology
Rosenstiel School of Marine and
Atmospheric Science

_____
Zachary Cheviron, Ph.D.
Associate Professor of Biology
University of Montana

_____
Guillermo Prado, Ph.D.
Dean of the Graduate School

GRAHAM, ALLIE MARIE (Ph.D., Biology)

Comparative Genomics of Mechanisms Underlying Adaptation to High-Altitude Environments in Andean Waterfowl (August 2017)

Abstract of a dissertation at the University of Miami.

Dissertation supervised by Professor Kevin G McCracken.
No. of pages in text: (201)

Organisms which reside in high-altitude environments represent ideal species for the study of physiological adaptation, as well as for elucidating the underlying genetic mechanisms associated with such a strong selective constraint (ie. hypoxia). Therefore, the variety of waterfowl species who have independently invaded the altiplano plateaus of the Andean mountains provide crucial representative study organisms in answering such questions. The research encompassed by this dissertation sought to answer questions about [1] the specificity of genetic mechanisms involved in high-altitude adaptation, [2] the potential for evidence of how these molecular adaptations were acquired, either through *de novo* mutations (convergence, parallel evolution) or collateral evolution through hybridization, and finally [3] to what extent, and at what level, is convergent and/or parallel evolution occurring, all in the context of comparing three different Andean waterfowl species, and their respective high- and low- altitude populations: Speckled teal (*Anas flavirostris*), Yellow-billed pintail (*Anas georgica*) and Cinnamon teal (*Anas cyanoptera*). Given the abundance of research on adaptation to low-oxygen environments, my dissertation takes two different approaches: one that focuses on *a priori* candidates, and one that takes a genome-wide scan approach. I explored the role of the mitochondrial genome, alpha- and beta- hemoglobin complexes, and the Hypoxia-

Inducible Factor (HIF) pathway in the three Andean waterfowl species using target-enrichment datasets, and then looked at overall genomic response to high-altitude adaptation in the Speckled teal, using RAD-seq data. Overall, my dissertation showed a high degree of molecular convergence and parallelism on a number of previously identified genetic mechanisms – more specifically, in hemoglobin (β^{ε}) and specific HIF-pathway candidates (*EPAS1*, and *EGLN1*). Ultimately, I was able to show considerable levels of parallelism not only at the pathway level, but at the gene, exon and even nucleotide/amino-acid level. My dissertation suggests that adaptive molecular evolution is highly predictable, especially for adaptations to high-altitude, low-oxygen environments.

ACKNOWLEDGEMENTS

"Vision is the art of seeing what is invisible to others" – Jonathan Swift

First, I would like to thank the many people and provincial and federal governments in Argentina, Peru, and Bolivia who assisted the McCracken lab with fieldwork for many years to collect the tissue samples I utilized in my dissertation. I would also like to thank the Arctic Regional Supercomputing (ARSC) at the University of Alaska Fairbanks assisted with high-performance computing throughout my dissertation, and provided me with excellent/prompt troubleshooting. I would also like to thank Woods Hole Marine Biological Laboratories (Workshop on Molecular Evolution) and Michigan State University (Analyzing Next-Generation Sequencing Data) for hosting incredibly useful workshops that I attended, and have thus applied the lessons learned into my dissertation work. Finally, I would like to thank my advisor (K. McCracken), and my committee members (W. Browne, M. Oleksiak, B. Whitlock, Z. Cheviron) for all their insight, advice and support.

General funding for various parts of my dissertation include the Alaska EPSCoR (NSF EPS-0092040, EPS-0346770), the National Science Foundation (DEB-0444748 and IOS-0949439), Frank M. Chapman Fund at the American Museum of Natural History, Graduate Student Research Grant-In-Aid from the Society for Integrative and Comparative Biology, a UM CAS Graduate Summer Research Award, and a Maytag Fellowship. Specimen collections were carried out under University of Alaska-Fairbanks IACUC protocols #02-01 and #05-05.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 4

Chapter 5

Chapter 1: Introduction

Background

Oxygen homeostasis, a critical organismal biological constraint, is maintained through the coordination of numerous regulatory genes (Semenza 2007b). Oxygen (O_2) occupies a central role in aerobic metabolism, where it serves as the terminal electron acceptor in oxidative phosphorylation (OXPHOS) in the mitochondria. To maintain homeostasis, multicellular eukaryotes have adopted specialized mechanisms to enhance O_2 uptake and distribution, resulting in dynamic respiratory and circulatory systems, capable of responding to changes in O_2 availability on local, organismal, and temporal levels. These changes are mediated in part through the induction of hypoxia-inducible transcription factors (HIF), the regulatory components of which are highly conserved, both in form and function across phyla (Hoogewijs et al. 2007; Gorr et al. 2006; Rytkönen et al. 2011). Moreover, for organisms that rely on aerobic energy production, genes involved in mitochondrial function and energy metabolism, O_2 binding and delivery, and hematopoiesis can be activated (Hoogewijs et al. 2007; Hopkins and Powell 2001). Utilization of these different regulatory pathways provides a mechanism, and logical target for selection, for local adaptation or acclimatization to inadequate O_2 supply/low O_2 conditions (i.e. hypoxia) (Cheviron and Brumfield 2011; Bigham and Lee 2014) (Figure 1.1).

Consequently, determining the mechanisms/pathways by which organisms use to cope with low O_2 supply is required to better understand this fundamental biological constraint. In order to develop an improved understanding of hypoxia response and

adaptation mechanisms, I aim to test how populations of the same species have adapted to cope with variable O_2 supplies by investigating genomic variance in individuals found in both hypoxic and normoxic environments. By including several independent lineages that have colonized the same high-altitude region for different lengths of time, I also tested short and long-term responses to hypoxic environments (i.e., phenotypic plasticity versus molecular genetic change).

High-altitude species found at > 4,000 meters where the O_2 partial pressure (pO_2) is approximately 60% of that at sea level offer an unparalleled system to understand the molecular and physiological bases of adaptation to hypoxic environments. To date, molecular work on high elevation adaption has largely focused on hemoglobin (Hb) (reviewed by Storz & Moriyama 2008); However, studies in a variety of organisms have highlighted several different strategies/mechanisms that have resulted in adaptation to low-O_2 environments and/or temperature stress, including insects (Zhang et al. 2013), fish (Gracey et al. 2004; Li et al. 2013b; Rissanen et al. 2006; Terova et al. 2008), birds (Cheviron et al. 2008; Qu et al. 2013; Scott et al. 2011; Wang et al. 2015) and mammals (Alkorta-Aranburu et al. 2012; Kang et al. 2013; Li et al. 2014; Qiu et al. 2012; Xu et al. 2007; Xu et al. 2011; Yi et al. 2010). These studies have shown that several interconnected genetic pathways, including the OXPHOS and HIF-pathways, are frequently converged upon during the colonization of low-O_2 environments, and may play an integral role in high-altitude adaption. It is clear that these genes/pathways are extensively converged upon via natural selection in response to a similar, or identical, biological constraint. Such patterns of convergence are thought to reflect intrinsic biases in the generation of variation, due to the propensity of development generally increasing

the likelihood by which similar traits evolve in different lineages (Wake 1991; Losos 2011). Thus, the documented predictability of natural selection on a small number of genetic targets, allows us to *a priori* test whether this similarity extends to other high-altitude adapted organisms.

High-altitude adaptation involves coordinated changes in the expression of many genes that are involved in interacting biochemical pathways; thus in order to have a more complete understanding of the mechanisms behind adaptive evolution at high-altitude, further inquiry requires a pathway perspective. By considering candidate pathways, it is possible to examine high-altitude adaptation at the level of the gene interactions (Cheviron and Brumfield 2011; Cork and Purugganan 2004). Therefore, I tested whether similar genes/pathways (ie. HIF, OXPHOS, Hb) underlie high-altitude adaptations in three Andean waterfowl species: the yellow-billed pintail (*Anas georgica),* cinnamon teal (*Anas cyanoptera),* and the speckled teal (*Anas flavirostris flavirostris*/*Anas f. oxyptera*) (Figure 1.2). Genetic signatures for high-elevation adaption were compared and contrasted between species with more recent/shallow divergences, *Anas georgica* and *Anas cyanoptera,* and a species with deeper divergence, *Anas f. flavirostris*/*oxyptera*. Each of these lineages has also independently colonized the same high-altitude wetlands and puna grasslands of the Altiplano and inter-Andean valleys of South America. Thus, although these lineages have invaded hypoxic environments, whether their adaptation was accomplished via the same mechanism, or is due to adaptive changes and/or phenotypic plasticity, remains largely unknown.

Objectives

This dissertation will elucidate the degree to which natural selection is predictable under a constant/severe constraint (i.e., hypoxia) by pinpointing pathways and genes associated with genetic mechanisms for adaptation to high-altitude environments in three Andean waterfowl species. Therefore, I tried to elucidate the extent to which convergent and/or parallel evolution is associated with high-altitude adaptation in three Andean duck species. Due to what is known about the frequency (and depth) of convergence on certain genetic mechanisms, I evaluated the extent to which similar patterns (seen across a multitude of other organisms) match these study organisms.

Specifically, I examined genetic variation at in a variety of genes previously unstudied in these three duck systems, including full mitochondrial genomes, the complete hemoglobin α/β gene clusters, specifically the embryonic paralogs of the adult Hb genes and the *cis*-regulatory upstream regions of each gene, and 26 HIF-pathway genes, in order to understand their possible role in the three species invasion of high-altitude environments. I was able to identify whether the same or different molecular pathways are involved in high altitude adaptation in each taxon by testing for similar/different genomic signatures (i.e., same or different loci/SNPs within the same pathways) between species-pairs that would support molecular convergence through parallel evolution (i.e., new mutations), or collateral evolution through hybridization (Stern 2013).

In chapter 2, I explored the role of genetic variation in the mitochondrial genomes of both high- and low-altitude populations of the three species (cinnamon teal, yellow-billed pintail, speckled teal). Mitochondrial variation, especially non-synonymous

changes, have the potential to affect changes of OXPHOS function, and thus cellular energy production. However, there was no such variation to be found in any of the three species, instead, differences in levels of purifying selection were found between each high- and low-altitude species pair. Such patterns were explained by the possible necessity of high-altitude populations requiring optimal energy production through tighter OXPHOS coupling. I also discussed the relative contributions of demography and selection for the patterns observed, although I argue selection has likely played a strong role given the consistency of results across multiple taxa with varying degrees of difference in demography. While we did not find any genetic variation, nor evidence of positive selection, associated with high-altitude adaptation across the three species, I suggest that my result does not preclude other roles for this organelle in such adaptation.

In chapter 3, I investigate how variation in both the α- and β- gene clusters, with a focus on embryonic paralogs (α^{π} and β^{ε}), along with upstream cis-regulatory regions, were associated high-altitude adaptation. I was able to show that there was (a) significant variation in β^{ε} resulting in nonsynonymous changes in amino-acid composition, and (b) that these amino acid changes were identical between speckled teal and yellow-billed pintail. I speculate on the degree to which that variation is identical via parallel evolution, or potentially through hybridization between their high-altitude populations. I was also able to identify that significant variation in the upstream *cis*-regulatory regions overlapped with transcription-factor binding sites (TFBS), suggesting a potential role for regulation of transcription activity as an additional avenue for adaptation to high-altitude

environments. I argue that exploring these additional avenues has the potential to expand our understanding of the variety of ways organisms can adapt to high-altitude environments.

In chapter 4, I examined whether there is evidence for parallel evolution in the HIF-pathway, not only across the three waterfowl species, but also whether there is evidence of convergence, in relation to high-altitude human populations. I discovered that both speckled teal and yellow-billed pintail had significant numbers of outliers associated with *EPAS1* and *EGLN1*, both of which are considered outliers in both Tibetan and Andean human populations. However, unlike those human populations, my results showed variation in exonic regions (exon 12) of *EPAS1*, which also result in nonsynonymous changes in amino-acid composition; wheras the variation in *EGLN1* was largely intronic. Based on a plethora of functional information associated with these genes in humans, I hypothesize that these variants in *EPAS1* are likely affecting hemoglobin production (either directly or indirectly), thus attenuating the long-term detrimental effects of erythocytosis which accompanies living at high-altitude.

Lastly, although the mitochondria, hemoglobin and the HIF-pathway are frequent targets of adaptations to hypoxic, high-altitude environments, there are numerous lineage specific examples of such adaptations occurring in idiosyncratic ways, dependent on the organism. Therefore, I used genome-wide variation from the speckled teal (*Anas flavirostris*) to explore other avenues for adaptation to such environments through different mechanisms targeting similar physiological systems, in addition to the role of sex-chromosomes in facilitating population divergence. In chapter 5, I identified a set of loci putatively under selection with allele frequencies strongly correlated with high- and

low-altitude habitats – most notably those involved in the insulin-like signaling pathway, bone morphogenesis, metabolic processes through the mitochondria (oxidative phosphorylation), responders to hypoxia-induced DNA damage, and feedback loops to the HIF pathway; these same outliers were also found to be largely concentrated on the Z-sex chromosome. I discuss Z-linked loci and their role in the population differentiation of incipient diverging species; our data suggest that Z-linked loci may be simultaneously under selection due to their mechanistic role in high-altitude adaptation as well as phenotypic divergence.

Overall

My dissertation encompasses genomic approaches attempting to locate the pathways, genes, and polymorphisms underlying mechanisms of adaptation to high-altitude environments, in combination with how variation in those components have contributed to population divergence. As a whole, these chapters attempt to evaluate the degree to which natural selection is predicable (pathway, gene, nucleotide), not only between similar taxa (i.e., ducks), but among long diverged taxa (i.e., ducks vs. humans). Ultimately, I was able to show there was a high degree of molecular convergence, that not only extends to similar pathways (OXPHOS; HIF-pathway), and genes (Hb, β^{ε}; *EPAS1*; *EGLN1*), but to the same exons (*EPAS1*) and even the same amino-acid changes (β^{ε}). However, this was only the case for two of the three species sampled in my dissertation – so although these data suggest high levels of convergence/parallel evolution, in the case of the genes tested here, it does not extend to all Andean waterfowl,

but only two of the three I studies. Altogether, my dissertation highlights the complexity by which adaptation occurs across the genome, due to the role of natural selection in response to an extreme environmental pressure.

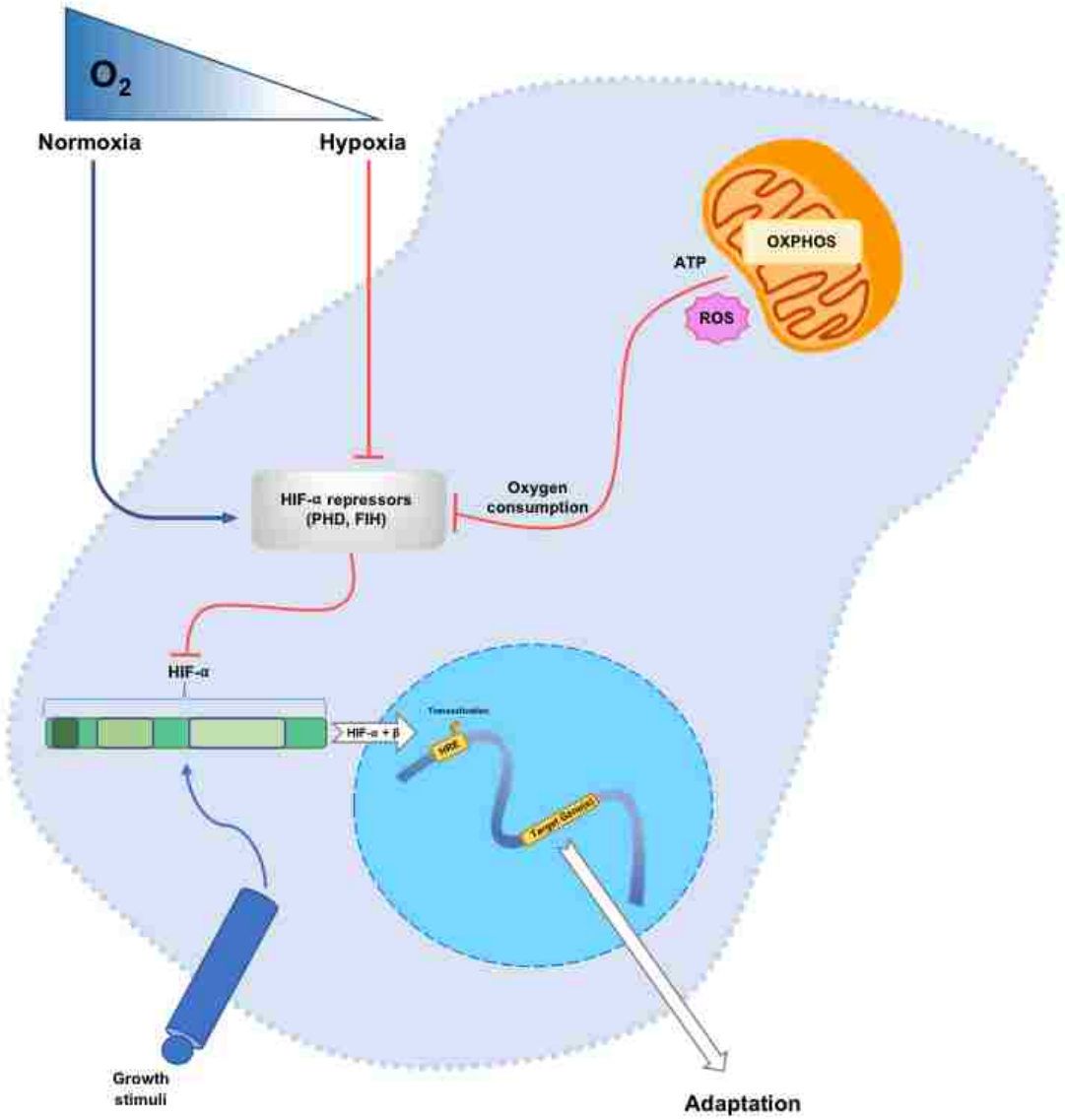

Figure 1.1: General schematic showing of the interconnectivity of the HIF-pathway, OXPHOS pathway, and other gene regulatory elements in a cell. Image modeled after Wenger et al. (2005).

Figure 1.2: Representative images of (top) speckled teal, *Anas flavirostris* (middle), yellow-billed pintail, *Anas georgica*, (bottom), cinnamon teal, *Anas cyanoptera* –male on left, female on right. "Speckled Teal" by Duncan Wright is licensed under CC BY-SA 3.0, "Yellow-Billed Pintail" by Brian Ralphs is licensed under CC 2.0, "Cinnamon Teal" by Doug Greenberg is licensed under CC BY-NC 2.0.

Chapter 2: Signatures of demography and purifying selection on mitochondrial genome evolution in three high-altitude Andean duck species

Background

The Andes are the world's highest mountains outside of Asia and possess some of the largest high plateaus in the world. The Altiplano, in particular, is the large plateau that extends from northwestern Argentina to southern Peru and Bolivia. It contains wetlands, including Lake Titicaca, that host a large number of waterfowl (Fjeldså 1985; Caziani et al. 2001). These species are ideal for studying the effects of both hypoxia (low O_2), and cold-stress, due to the metabolic demands associated with their high-altitude environment. Previous work on some of these waterfowl has focused on the extent to which variation in hemoglobin (Hb) is associated with high-altitude adaptation. This work has revealed seemingly extensive convergent evolution on specific genetic changes in Hb subunits affecting O_2-binding affinity (Natarajan et al. 2015; McCracken et al. 2010; McCracken et al. 2009a). However, outside of Hb, genetic variation associated with other physiological mechanisms for high-altitude living remain under-investigated in this group, most notably the mitochondrial genome and its role in oxidative phosphorylation (OXPHOS) and the electron transport chain.

Mitochondria serve a critical function in the production of cellular energy, due to genes encoding OXPHOS and parts of the five complexes involved in the electron transport chain. Mitochondrial bioenergetics is exposed to constant natural selection, and is thus considered an important driver of speciation (Gershoni et al. 2009; Das 2006; Breton et al. 2014), especially through its contributions to population divergence in

relation to hypoxic conditions (Zhang et al. 2013; Cheviron et al. 2008; Li et al. 2013b; Xu et al. 2007; Scott et al. 2011; Kang et al. 2013; Luo et al. 2008; Ehinger et al. 2002; Tomasco and Lessa 2011; Yu et al. 2011), as well as temperature stress (Mishmar et al. 2003; Hoekstra et al. 2013; Koevoets et al. 2012; Arnqvist et al. 2010).

To date, many studies have identified adaptive mitochondrial variation associated with a physiological response to O_2 or temperature stress, including differential mitochondrial gene expression and regulation (Gracey et al. 2004; Whitehead and Crawford 2006; Scott et al. 2015), increased density of mitochondria in skeletal muscle (Hoppeler et al. 2003; Scott et al. 2009), increased metabolic capacity and catalytic efficiency (Zhang et al. 2013), as well as genetic variation (Hassanin et al. 2009; Yu et al. 2011; Melo-Ferreira et al. 2014). Other bird species, including waterfowl, have been shown to have adapted to hypoxic, high-altitude conditions via the mitochondrial genome, through genetic variation in protein coding regions (Cai et al. 2013; Zhao et al. 2016; Zhou et al. 2014; Toews et al. 2014; Scott et al. 2011). Ultimately, evidence points to the possibility of both parallel and convergent evolution in mitochondrial adaptations. Thus, it is clear that the mitochondria play a central role in adaptations to high-altitude environment, due to selection on overlapping mechanisms related to hypoxia and temperature stress; therefore, representing a genetic target likely under selection in waterbirds in the high-altitude wetlands of the Andes Mountains.

However, demographic fluctuations, such as population expansion or contraction, have the potential to confound and mimic the genetic signatures of selection (Bamshad and Wooding 2003; Excoffier et al. 2009; Nielsen 2005), especially for the mitochondrial genome (Ballard and Whitlock 2004; Ballard and Pichaud 2014). In the case of

organisms invading a new habitat niche such as at high altitude, the colonizing population will be predicted to have undergone a founder effect likely followed by a population expansion. This scenario is predicted to lead to patterns of nucleotide diversity and site frequency spectra that are very similar or identical to those produced by directional selection, due to an excess of rare alleles, mimicking a recent selective sweep. This is important to note since such a founder event likely occurred for each of the three duck populations studied here, which colonized high-altitude areas from low altitude ancestral environments. In addition, how recently population divergence has occurred can also influence the extent to which selection can be ascertained. Recently diverged populations may not have had enough time for beneficial mutations to accumulate, and therefore not have the genetic variation to adequately estimate the effects of selection, using traditional tests for neutrality, such as dN/dS or Tajima's D (Mugal et al. 2013). In contrast, older diverged populations that have had sufficient time to accumulate genetic variation may allow for a wider variety of tests to demarcate patterns that may have resulted from selection. In the case of the latter, for example, the population would be expected to have accrued non-synonymous substitutions, where the former more recently established populations might not. Finally, the mitochondrial genome itself is expected to display patterns distinct from the nuclear genome, due to being haploid with a four-fold smaller effective population size; therefore, genetic drift is expected to have a disproportionately larger effect on mtDNA than nuclear DNA (Ballard and Whitlock 2004; Dowling et al. 2008) as are factors like philopatry and sex-biased dispersal (Anderson et al. 1992).

With these issues in mind, I compared signatures of demography and selection related to high-altitude adaptation in the mitochondrial genomes of three Andean

waterfowl species: yellow-billed pintail *(Anas georgica)*, cinnamon teal (*Anas cyanoptera)* and speckled teal (*Anas flavirostris*). Each high-altitude population independently colonized the same habitats in the Andes from ancestral low-altitude lineages, thus providing natural replicates of founder events and independent adaptive events to high-altitude, cold-stressed, hypoxic environments (McCracken, *et al*., 2009b). These dabbling duck species also represent contrasting depths of evolutionary separation, since the time that low-altitude populations first colonized high-altitude regions. Previous genetic work on these three species indicate that the yellow-billed pintail and cinnamon teal have low mitochondrial divergence (0.3 - 0.4%), whereas the speckled teal divergence is deeper by at least an order of magnitude (2.7%) (McCracken et al. 2009a; Pereira and Baker 2006; Wilson et al. 2013).

To address these questions, I sequenced and analyzed 60 full mitochondrial genomes in these three species of Andean ducks across low- and high-altitude populations. I then used a variety of neutrality tests to ascertain the roles of purifying and positive selection on both coding and non-coding regions of these mitgenomes. Ultimately, I find evidence of significant purifying selection across the mitochondrial genome, especially in protein-coding genes, potentially acting to maintain the OXPHOS unit's ability to operate optimally through increased negative selection in high-altitude populations. However, I also describe the potentially confounding role of demographic effects involved with using specific types of neutrality tests, although I argue that some of our methods ameliorate that constraint.

Methods

Specimen Collection and DNA Extraction

A total of 60 individuals were used for this study from three different Andean duck species – cinnamon teal, speckled teal and yellow-billed pintail. Sampling from each of the three species is comprised of 20 individuals with 10 individuals from each of the high-altitude populations and 10 individuals from each of the low-altitude populations. For the cinnamon teal, individuals from low-altitude populations are the *A. c. cyanoptera* subspecies (n = 10; elevation range 7-23 m) and from high-altitude are the *A. c. orinomus* subspecies (n = 10; elevation range 3533-3,871 m) (Wilson et al. 2013). For the speckled teal, individuals from low-altitude populations are the *A. f. flavirostris* subspecies (n = 10; elevation range 77-860 m) and from high-altitude are the *A. f. oxyptera* subspecies (n = 10; elevation range 3,211-4,405 m). For the yellow-billed pintail, individuals from both populations are taxonomically identified as *Anas georgica spinicauda*. A total of 20 yellow-billed pintails were collected from low- (n = 10; elevation range 292-914 m) and high-altitude (n = 10; elevation range 3,332-4,070 m). Genomic DNA was extracted from tissue using a DNeasy Tissue Kit (Qiagen, Valencia, California, USA) following manufacturers protocols.

Target-enrichment Sequencing

This study utilized in-solution target capture to selectively enrich libraries for the complete mitochondrial genome prior to NGS sequencing (Gnirke et al. 2009). All steps of the process were performed by MYcroarray (Ann Arbor, MI). A custom MYbaits® biotinylated ssRNA target capture baitset was designed. Specifically, 1,359 120mer

probes at 2x tiling density were designed from five previously published *Anas* mitochondrial genomes (*A. acuta*, NC_024631.1; *A. crecca*, KC771255.1; *A. formosa*, NC_015482.1, *A. platyrhynchos* NC_009684/EU009397.1; and *A. poecilrhyncha* NC_022418.1). DNA samples were then subjected to capture library synthesis and preparation. The prepared libraries were hybridized with the custom mitochondrial baits. Following hybridization, target regions were purified on magnetic beads followed by post-hybridization amplification to add indexing sequences. Sequencing was performed on an Illumina Hi-Seq platform paired-end (100 bp) with a 250-300 bp insert size.

Mitochondrial Genome Assembly and Annotation

Sequences were received pre-parsed by individual with adapters trimmed and quality filtered (Q < 30). Additional adapter trimming was performed utilizing *fastq-clipper* (AGATCGGAAGAGC) and remaining sequences were then filtered by length and quality using *fastq-quality filter* (reads <20 bp, and Q < 30) from the FASTX-Toolkit v. 0.0.13 package (Gordon and Hannon 2010). The mitochondrial genome for each individual was assembled with MITObim version 1.7 (Hahn et al. 2013) using the full mitochondrial genome of the mallard as a reference (*Anas platyrynchnos*; EU009397.1). For genome reconstruction, the "quick" and "trim" options were used, with an iteration limit set to 30. Mapping results were aligned with MAFFT (Katoh et al. 2002) and annotated in Geneious version 8.1.6 (Kearse et al. 2012) based on 90% threshold similarity to the mallard reference (*Anas platyrynchnos*; EU009397.1). Protein coding, tRNA, and rRNA genes were extracted from these annotations and then subjected to individual analyses.

Population Genetic, Neutrality Tests and Selection Analyses

For each species, I calculated pairwise F_{ST} (Weir and Cockerham (1984) for all variable sites between low- and high-altitude populations for full mitochondrial genomes as well as individual gene regions - protein coding, and non-protein coding (D-Loop, tRNA, rRNA) - in each of the three species using Arlequin v. 3.5 (Excoffier and Lischer 2010) and MEGA v7 (Kumar et al. 2016). All protein-coding genes were subjected to various selection analyses, whereas any tRNAs with significant SNPs were subjected to structural analyses.

These analyses were aimed at dissecting out the respective roles of different demographic and selective processes on the mitochondria, specifically signatures of population expansion and or contraction or positive and purifying selection. I tested for evidence of different types of selection using [1] Tajima's D [2] codon-based Z-tests, both implemented in MEGA v7, and [3] codon-based selection estimators SLAC and MEME applied through HyPhy (Hypothesis Testing Using Phylogenies; (Pond and Frost 2005a, 2005b).

Tajima's D (Tajima 1989; Nei and Kumar 2000) compares the mean number of pairwise differences and number of segregating sites, and is an unbiased estimator of θ, which under neutrality should be equal to nucleotide diversity (π). Tajima's $D < 0$ signifies excess high frequency polymorphisms and is indicative of a recent selective sweep or purifying selection and/or founder event followed by population expansion. Tajima's $D > 0$ signifies low levels of both low and high frequency polymorphism and suggests either balancing selection or population contraction. We also implemented neutrality tests based on the site-frequency spectrum using DnaSP (Librado and Rozas

2009). In addition, I utilized a Z-test to test hypotheses about the types of selection present between the two populations – neutral ($d_N = d_S$), positive ($d_N > d_S$) and purifying selection ($d_N < d_S$). These analyses were performed averaging within groups, estimating variance via bootstrap method (500 replications), using the Nei-Gojobori method (Nei and Gojobori 1986).

I also examined evidence for selection at the level of specific codons using the SLAC (Single Likelihood Ancestor Counting) method and MEME (Mixed Effects Model of Evolution) through the program HyPhy. SLAC assesses selection using ω as a proxy for nonsynonymous/synonymous (dN/dS) ratios; the ω ratio is a measure of natural selection acting on the protein, with values of $\omega < 1, = 1$, and > 1 signifying negative/purifying selection, neutral evolution, and positive selection, respectively. These modules use different methods to estimate ω (dN/dS) at every codon in the alignments and report which codons show evidence of positive or negative selection, using significance levels. SLAC calculates the expected and observed numbers of synonymous and nonsynonymous substitutions to infer selection using maximum-likelihood phylogenetic tests (Pond and Frost 2005b). Unlike SLAC, MEME allows ω to vary across codons as well as across branches of the phylogeny, allowing it to detect a small proportion of branches that are evolving under positive selection or under episodic selection (Murrell et al. 2012). Episodic selection refers to selective pressures that are sustained in a finite fashion in episodes across evolutionary time; this is compared to pervasive selective pressures, which are considered sustained selection over an extended period of time. To avoid a high false-positive rate, due to the reduced number of sequences, sites with P values < 0.1 for both models were considered significant (Pond

and Frost 2005b). My data, in conjunction with six outgroups were used in these analyses of each of the three species: *A. acuta*, NC_024631.1; *A. clypeata*, NC_028346.1: *A. crecca*, KC771255.1; *A. formosa*, NC_015482.1, *A. platyrhynchos* EU009397.1; and *A. poecilrhyncha* NC_022418.1.

Lastly, mitochondrial tRNAs are non-protein coding regions whose structure is highly susceptible to small changes in genetic variation. Thus, the structure of the tRNAs were assayed in ARWEN (Laslett and Canbäck 2008) using the online server (http://mbio-serv2.mbioekol.lu.se/ARWEN/). Locations of SNPs were then identified based on general structural information as well as whether SNPs overlapped with any known mutagenic SNPs in humans (Lott et al. 2013; Wittenhagen and Kelley 2003).

Results

Assembly Features and Genome Organization

The average coverage for the three duck mitochondrial genomes was 232x with an average of 41,635 reads per individual mitochondrial assembly containing an average 48.6% GC content. The final consensus genome size for each species was 16,601bp (cinnamon teal), 16,602 bp (speckled teal), and 16,616 bp (yellow-billed pintail), respectively (Table 2.1). The mitochondrial genome sequences for each individual will be deposited in NCBI GenBank once submitted for publications. Each of the mitochondrial genomes from the three species contained NADH6 and 8 tRNAs transcribed from the light strand, with the remaining 12 protein coding genes, 14 tRNAs and the 12S and 16S rRNAs transcribed on the heavy strand (Figure 2.1 - 2.3). This composition and

arrangement of the mitochondrial genome is typical of birds and also matches other Anseriformes species ranging from 16,594 bp to 16,608 bp in size (Yan et al. 2015; Zhou et al. 2015; Pan et al. 2014; Hu et al. 2015).

Population Genetic Analyses

Both cinnamon teal and yellow-billed pintail revealed low genetic differentiation, and nucleotide diversity across the full mitochondrial genome. For the cinnamon teal, the overall F_{ST} between high- and low-altitude groups was 0.021, with 64 segregating sites. For the yellow-billed pintail, F_{ST} between high- and low-altitude groups was 0.000, with 139 segregating sites. In contrast, for the speckled teal, F_{ST} between high- and low-altitude groups was 0.714, with 206 segregating sites.

The F_{ST} patterns for each SNP for each of the three species were representative of their genetic differentiation, with the cinnamon teal and yellow-billed pintail showing no individual polymorphic sites that were significantly differentiated between high- and low-altitude populations, whereas the speckled teal had 119 such sites. SNPs that were significantly different ($P < 0.05$) were present in both noncoding (D-Loop, l-rRNA, tRNA-Ala, tRNA-Thr) and almost all coding regions (ND1-6, COX1-3, ATP6 and ATP8). Of those sites, several were responsible for nonsynonymous amino-acid changes: two in ND5, three in ND3, one in COX1, and one in ATP6. Otherwise, the rest of the variation in protein-coding genes was mostly located in 3rd codon positions, which are selectively "silent" effecting no amino acid changes.

A maximum-likelihood tree of all 60 individuals including the 6 outgroups (Figure 2.4) was constructed using GARLI (Bazinet et al. 2014; Zwickl 2006), in order to show general phylogenetic relationships between the species/populations and the outgroups used for HYPHY, which produces its own phylogenetic trees to test for evidence of selection (see next section).

Protein Coding Genes – Demography and Selection Analyses

Both demography and selection pressures can cause similar patterns for most commonly used tests of neutrality, especially for tests based on Tajima's D and the site-frequency spectrum, potentially leading to spurious inference of selection when applying neutrality tests (Thornton and Jensen 2007; Przeworski 2002). However, tests based on comparisons of divergence and/or variability between different classes of mutation (i.e., codon-based Z-scores, HYPHY) are considered less likely to be affected by demographic processes like population expansion or decline, founder events, and bottlenecks. Therefore, this study incorporates both types of neutrality tests, and considers demographic processes as the null hypothesis (H_0). Specifically, the H_0 for each high-altitude population is founder event followed by population expansion, whereas the H_0 for each low-altitude population is an older more established stable population with larger N_e.

All three species had significantly negative Tajima's D for both high- and low-altitude populations for the full mitochondrial genome (Figure 2.5). The cinnamon teal had 7 out of 13 genes with segregating sites; of those 7 genes, all showed negative D values for the total population as well as the high- and low-altitude sub-populations

treated separately. For yellow-billed pintail, 10 out of 13 genes had segregating sites, with each of those genes having negative D values for all populations, except for ATP6 which had positive D values (Figure 2.5). For speckled teal, 11 of 13 genes had segregating sites, with 6 genes showing positive D values for the total population, but negative D values for the high- and low-altitude populations treated separately, except for 6 genes for which the low-altitude population had no segregating sites and 1 gene had no segregating sites in the high-altitude population. The same patterns for each species were similar when parsed out by mitochondrial complexes (Figure 2.6), as well as among the tRNA, rRNA, and D-loop non-coding gene regions (Figure 2.7).

The site-frequency spectrum of the full mitochondrial genomes of each species thus showed evidence for non-neutrality (Figure 2.8). Cinnamon teal and yellow-billed pintail showed an excess of low-frequency SNPs, as well as paucity of high-frequency SNPs; whereas speckled teal showed an excess of low and high-frequency SNPs segregating between the high- and low-altitude populations. Similar patterns were seen in the Tajima's D (see above), which are similar to the site-frequency spectrum metrics (Durrett 2008; Ronen et al. 2013).

Selection was also assayed at the level of specific codons using HyPhy (Pond and Frost, 2005; Pond and Frost, 2005), which uses ω as a proxy for dN/dS. The output from the SLAC analysis showed that both cinnamon teal and yellow-billed pintail had no codons under significant positive selection, whereas speckled teal had eight codons under significant positive selection across five genes (COX2, CYTB, ND1, ND5, ND6) (Table 1). The MEME analysis showed evidence of episodic diversifying selection across the three species, with most genes showing at least one significant codon (Table 2.2): the

cinnamon teal (16), speckled teal (15), and yellow-billed pintail (9). There were also several instances of diversifying selection targeting the same codon, in two of the species (ND1, codon 17), and two codons in particular present in all three species (COX1, codon 273; ND4, codon 54) (Table 2.2). Overall, the results from the two analyses (SLAC, MEME) show more examples of purifying selection, as compared to positive selection (posterior value 0.1) in all three species; although there was some evidence of diversifying selection in a few mitochondrial genes (ATP6, COX1, ND1, ND4; Table 2.3).

Of those genes identified through these analyses, speckled teal was the only species that had significant nonsynonymous changes between high- and low altitude populations, in both ATP6 and COX1, and members of the NAD complex, although not ND1 and ND4. Thus, the evidence for positive selection I find here should be taken more as indicative of positive selection at that site on this species across both high- and low-altitude populations, rather than between populations.

A codon-based Z-test was used to test hypotheses about the type of selection (neutral, positive, purifying) predominating between high- and low-altitude populations. For all three species, the Z-test for individual protein coding genes showed evidence for either neutrality or purifying selection (Table 2.4). None of the analyses in any of the three species yielded significant results for positive selection. However, the low-altitude speckled teal population showed significantly fewer genes under purifying selection (3), relative to the high-altitude speckled teal population (10; $X^2 = 7.5385$, $P < 0.01$). This pattern was not seen in either the cinnamon teal or the yellow-billed pintail ($P > 0.05$). However, comparing the incidence of purifying selection across all low-altitude

populations to all high-altitude populations showed evidence of more widespread purifying selection ($X^2 = 6.477$, $P < 0.01$). In addition, across all three species, COX2 and ND1 genes appear to be under purifying selection in the high-altitude populations but evolving neutrally in low altitude populations, whereas the ND4-6 genes appear to be under purifying selection in both high- and low-altitude populations. Overall, neutrality tests with different sensitivities to demographic effects showed significant purifying selection in the protein-coding OXPHOS genes of high-altitude populations.

Non-protein Coding Genes – Selection, and Structure Analyses

Because tRNA and rRNA genes do not produce an amino acid sequence, quantifying nonsynonymous or synonymous mutations is not possible; thus selection could only be assessed using Tajima's D. Both tRNA and rRNA regions showed evidence of purifying selection and/or population expansion as compared to both the non-coding D-loop and protein-coding regions. Otherwise, each high-altitude population of each species showed patterns consistent with purifying selection.

For the speckled teal, the tRNAs that had SNPs with significant F_{ST} were tRNA-Ala (alanine) and tRNA-Thr (threonine) (Figure 2.9) - the SNP is in the anticodon stem close to the T-arm (ie. TΨC arm) in tRNA-Thr, and the SNP is in the t-arm next to the start of the accepter stem in tRNA-Ala. There is currently no known function associated with these specific SNPs in those regions, at least in human literature. However, in general, SNP changes in these tRNAs have been associated with mitochondrial myopathy and encephalomyopathy (Lott et al. 2013). Both the tRNA-Ala and the tRNA-Thr showed

a 1 bp SNP difference between high- and low altitude populations. However, there were no significant polymorphisms in any of the tRNAs for cinnamon teal or for yellow-billed pintail.

Discussion

Effects of Demography and Selection on Mitochondrial Genome Evolution

In this study, I present a comparative analysis of mitochondrial genome evolution in three species of Andean waterfowl across their low- and high-altitude distribution in South America. My results suggest evidence for a role of purifying selection across the mitochondria for each population, although the pattern in which this was strongest was different for each species. These results are based on a combination of nucleotide- and codon-based analyses, which all suggest a major role for a specific selective pressure (i.e., purifying selection).

However, a major issue with using mitochondrial genomes results from the difficulty in accounting for population demography, including patterns resulting from population expansion or contraction, founder events and bottlenecks, and unequal migration rates between populations, which when not accounted for can lead to erroneous interpretations of signatures of selection (Ballard and Whitlock 2004; Excoffier et al. 2009; Bamshad and Wooding 2003). Another factor that makes it difficult to parse out the roles of demography and selection on the mitochondrial genomes is that these duck species are female philopatric, meaning that male-biased dispersal partly underlies the structuring (i.e., high F_{ST}) observed in the mitochondria (Flint et al. 2009; Peters et al.

2007; Avise et al. 1992; Zink and Barrowclough 2008; Peters et al. 2012), especially in the speckled teal, which has a much deeper estimated divergence time between its high- and low-altitude populations.

Therefore, we acknowledge these potentially confounding factors and have employed a combination of tests both susceptible to and relatively unsusceptible to demographic forces to guard against such bias (i.e., tests based on allelic frequencies and the site frequency spectrum vs. tests based on divergence or variability between different classes of mutation; Table 2.4). However, ultimately, there are few such tests available that are not constrained by demographic effects (Nielsen 2005), especially in so recently diverged species (Mugal et al. 2013).

Potential Patterns of Selection in the Three Andean Duck Species

For the cinnamon teal and yellow-billed pintail, recent divergence has resulted in little or no difference between the high- and low-altitude populations in polymorphism across the entirety of the mitochondrial genome. However, the mitochondrial genomes of both species show evidence for purifying selection overall (Table 2.4; Figures 2.5 - 2.7), with the strongest negative selection being exerted on the coding regions for cinnamon teal, and noncoding regions (tRNAs, rRNAs) for the yellow-billed pintail. The role of purifying selection is potentially striking, since not a single occurrence of a fixed (or close to fixed) non-synonymous change was detected between the high- and low-altitude populations of either of these two duck species.

It is clear that high- and low-altitude populations of cinnamon teal and yellow-billed pintail are utilizing other molecular and physiological mechanisms in low-O_2

environments (Wilson et al. 2013; McCracken et al. 2009a; McCracken et al. 2009b). This may be especially true for the cinnamon teal, where high-altitude populations are classified as separate subspecies, due to differences in plumage patterns and body size (Wilson et al. 2008; Wilson et al. 2010). In contrast, the speckled teal shows potentially stronger evidence for mitochondrial and oxidative phosphorylation (OXPHOS) adaptations to high-altitude hypoxic environments. Despite a deeper divergence (2.7%) between the high- and low-altitude populations, mitochondrial genome divergence is extreme compared to nuclear DNA (nuDNA $F_{ST} = 0.06$, mtdna $F_{ST} = 0.71$), suggesting that despite high mtDNA divergence, there is still significant gene flow between populations. In this species, the results suggest a pattern of non-neutral processes (Tajima's D, Codon-based Z tests, HYPHY tests) playing a role in divergence, which mirrors previous research on the evolution of Hb-O_2 affinity in this particular species (Natarajan et al. 2015).

Ultimately, my results also show a significant role for purifying selection in both high- and low-altitude populations of all three species (Figures 2.5 - 2.7), but especially in high-altitude populations (Table 2.4). This suggests that adaptation to hypoxic environments in these duck species has been demarcated by differences in the strength of purifying selection (e.g., strong at high altitude vs. weak at low altitude). However, it is important to note that, although the three species seem to be experiencing purifying selection, the data suggests that at high altitude purifying selection plays a more prominent role in the OXPHOS protein coding genes (Table 3; $X^2 = 6.477, P < 0.01$). This pattern of selection suggests that the mitochondrial genome has played a role in adaptation to high-altitude, hypoxic environments.

These results are perhaps not so surprising because mitochondrial genome evolution may largely be driven by purifying selection (Ballard and Whitlock 2004; Dowling et al. 2008; Castellana et al. 2011). It has been suggested that adaptation of nuclear and mitochondrial proteins, at the level of respiratory function is stringent due to how tightly intertwined the functions of the mitochondrial and nuclear genomes are (Wolff et al. 2014; Dowling et al. 2008; Rand et al. 2004; Blier et al. 2001). The compatibility between OXPHOS mitochondrial-nuclear subunits is particularly apparent when looking at population-level divergence, where it has effectively created isolating barriers leading to population structuring and speciation (Ellison and Burton 2006; Meiklejohn et al. 2013; Niehuis et al. 2008; Bolnick et al. 2008). Thus, given the functional importance of the genes encoded by the mitochondrial genome, selection for optimal catalytic capacity/regulatory efficiently could lead to stronger patterns of purifying selection.

Oxidative Phosphorylation and High-altitude Adaptation

The role of mitochondria in adaptation to the hypoxic and cold environments that accompany living at high altitude is thought to relate to efficiency (Solaini et al. 2010). Energy released from electron transport is used to create an electrochemical gradient to generate ATP, as well as generate heat to maintain body temperature (Cooper 2000), with efficiency of OXPHOS determining the balance between ATP and heat generation, occurring through proton/electron leakage (Jastroch et al. 2010). Tightly coupled OXPHOS subunits result in maximum ATP and minimum heat generation, whereas loosely coupled mitochondria generate more heat, at the expense of ATP production.

Thus, it has been suggested that the latter is preferable and important for survival in cold environments, especially in mammals (Gershoni et al. 2009; Mishmar et al. 2003; Welch et al. 2014), which use non-shivering thermogenesis (NST) for heat production. However, birds typically do not use NST, and instead use shivering, which is an ATP dependent form of muscular activity (Bicudo et al. 2001). Thus, these results suggest a potential selective pressure for tightly coupled OXPHOS subunits resulting in maximum ATP production, at the expense of heat generation.

Overall, this might explain why the high-altitude populations appear to be under stronger evolutionary constraint (i.e., increased purifying selection) at the level of mitochondrial function (Wang et al. 2011). In fact, not only has mitochondrial efficiency been shown to underlie adaptations to extreme environments, like high-altitude locations (Zhang et al. 2013; Zhao et al. 2013a), but it has been suggested that the mtDNA of strongly locomotive species has experienced stronger purifying selection to maintain efficient energy metabolism (Shen et al. 2009; Shen et al. 2010; Mitterboeck and Adamowicz 2013). In addition, it has been shown that bird species that live at altitudes above 1,000 m have basal metabolic rates (BMR) that are ~10% higher of those that live at lower altitudes (McNab 2012, 2003); however, recent evidence suggests that other thermophysical traits, not BMR, limit altitudinal distributions of tropical birds (Londono et al. 2015; Londoño et al. 2016). Thus, conservation of mitochondrial catalytic efficiency could be the reason behind such pervasive purifying selection for the three species in this study and especially for their higher altitude populations.

However, it is important to note other several other possibilities for a lack of selective signal in the two most recently diverged species (cinnamon teal and yellow-

billed pintail). First, there are additional metabolic constraints placed on birds, especially related to flight (Norberg 2012; Hedenstrom and Alerstam 1995), which could provide some insight as to why both the cinnamon teal and yellow-billed pintail show widespread purifying selection, coupled with low genetic variation. Therefore, it is possible that the patterns seen in these two species are unconnected to high-altitude living, but are rather due to inherent constraints of being a bird.

Second, it is also possible that enough time has not elapsed for genetic variation in the mitochondria to have accumulated through mutation, and phenotypic plasticity may be more important during the "immediate" years following high altitude colonization. The physiological toll of hypoxia is system-wide in its effect; thus over time mechanisms may be selected for, through a series of independent adaptations across different functional levels (physiological, gene pathways/gene regulation, genetic variation), of which OXPHOS activity is frequently targeted (Storz et al. 2010; Seebacher et al. 2010). Thus, phenotypic plasticity may be playing a major role in these species with recently diverged populations, where not enough time has passed in order for genetic assimilation or genetic accommodation to occur (Lande 2009; Pigliucci et al. 2006; Levis and Pfennig 2016). Thus, although I did not find any strong selective signals for either cinnamon teal or yellow-billed pintail in relation to the mitochondrial genome, my results do not rule out a role for mitochondria, including organelle density in specific muscle tissue, changes in expression or biogenesis, and oxidative capacity, which have been identified as important mechanisms in other organisms (Gracey et al. 2004; Whitehead and Crawford 2006; Hoppeler et al. 2003; Dawson et al. 2016).

Conclusions

Although it is difficult to parse out the relative contributions of demography and selection on the patterns observed in these species, I argue that selection has likely played a strong role given the consistency of results across multiple taxa with varying degrees of difference in demography. In conclusion, these three Andean waterfowl species have successfully colonized and, are thriving in the same high-altitude environment, potentially due to a repeated pattern of purifying selection acting to maintain the OXPHOS unit's ability to operate optimally through increased negative selection in high-altitude populations. This is especially true for the speckled teal, which shows distinct signs of differential selective pressures acting on the mitochondrial genome between high- and low-altitude populations; specifically, a substantial increase in purifying selection likely due to the invasion of high-altitude niches.

However, this lack of any strong evidence for positive selection in these results, with an emphasis on purifying selection, doesn't rule out a more proactive role for mitochondria in the adaptation of these three species to high-altitude environments. Therefore, although my results do not suggest a role for positive selection, or specific genetic variation in the mitochondrial genome, as correlating with high-altitude adaptation, it does not address the other ways in which mitochondria may be involved in these species. In the future, the role of mitochondria in the evolution of a wider variety of other Andean species should be studied further to elucidate the extent of its involvement in their adaptation to high-altitude environments.

Table 2.1: General statistics (coverage, number of reads) for individual mitochondrial genome assemblies.

Specimen ID	Population	Species	Coverage Max	Coverage Avg	# Reads
REW 251	High	Yellow-billed pintail	312	168.68	30552
REW 247	High	Yellow-billed pintail	293	160.48	28788
REW 144	High	Yellow-billed pintail	519	226.22	40806
REW 112	High	Yellow-billed pintail	203	110.54	19851
KGM 1076	High	Yellow-billed pintail	392	224.07	40403
KGM 567	High	Yellow-billed pintail	223	117.66	21211
KGM 551	High	Yellow-billed pintail	326	202.38	36514
KGM 507	High	Yellow-billed pintail	193	91.32	16236
KGM 491	High	Yellow-billed pintail	248	135.31	24394
KGM 425	High	Yellow-billed pintail	237	139.94	25265
KGM 780	Low	Yellow-billed pintail	429	269.37	48421
KGM 765	Low	Yellow-billed pintail	382	206.77	37305
KGM 742	Low	Yellow-billed pintail	367	190.40	34269
KGM 733	Low	Yellow-billed pintail	344	181.85	32680
KGM 731	Low	Yellow-billed pintail	373	187.55	33761
KGM 714	Low	Yellow-billed pintail	670	399.93	72069
KGM 324	Low	Yellow-billed pintail	465	251.23	45324
KGM 309	Low	Yellow-billed pintail	502	277.91	49805
KGM 306	Low	Yellow-billed pintail	626	336.84	60621
KGM 274	Low	Yellow-billed pintail	639	361.15	65177
Average	-	Yellow-billed pintail	387.15	211.98	38172.6
KGM 1129	High	Speckled teal	416	197.46	35694
KGM 437	High	Speckled teal	376	204.56	36615
KGM 449	High	Speckled teal	195	87.72	15237
KGM 484	High	Speckled teal	492	246.75	43997
KGM 502	High	Speckled teal	423	210.49	37722
KGM 543	High	Speckled teal	446	226.76	40437
REW 092	High	Speckled teal	226	106.18	18850
REW 132	High	Speckled teal	801	390.63	70521
REW 219	High	Speckled teal	784	352.82	63346
REW 237	High	Speckled teal	557	269.59	48021
KGM 267	Low	Speckled teal	51	22.10	3766
KGM 275	Low	Speckled teal	349	187.11	33318
KGM 285	Low	Speckled teal	241	120.34	21609
KGM 319	Low	Speckled teal	345	176.23	31574
KGM 699	Low	Speckled teal	355	168.27	30176

KGM 727	Low	Speckled teal	500	259.59	46228
KGM 735	Low	Speckled teal	384	203.62	36583
KGM 747	Low	Speckled teal	454	232.55	41964
KGM 778	Low	Speckled teal	709	360.11	65231
KGM 790	Low	Speckled teal	365	186.88	33560
Average	-	Speckled teal	423.45	210.49	37722.45
REW 286	High	Cinnamon teal	1213	701.88	127107
REW 272	High	Cinnamon teal	624	314.80	56953
REW 269	High	Cinnamon teal	686	318.91	57665
REW 259	High	Cinnamon teal	564	248.89	45050
REW 255	High	Cinnamon teal	465	196.95	35407
REW 253	High	Cinnamon teal	448	217.55	39296
REW 238	High	Cinnamon teal	594	302.78	54882
KGM 533	High	Cinnamon teal	370	176.62	31446
KGM 527	High	Cinnamon teal	559	282.10	49996
KGM 486	High	Cinnamon teal	602	357.37	64468
REW 316	Low	Cinnamon teal	470	233.27	42192
REW 305	Low	Cinnamon teal	446	207.61	36960
REW 301	Low	Cinnamon teal	510	258.86	46856
REW 235	Low	Cinnamon teal	465	196.95	35407
REW 207	Low	Cinnamon teal	512	242.82	43473
REW 206	Low	Cinnamon teal	603	321.49	58399
REW 203	Low	Cinnamon teal	427	189.37	33764
REW 200	Low	Cinnamon teal	448	237.39	42937
REW 193	Low	Cinnamon teal	535	245.06	43508
REW 081	Low	Cinnamon teal	354	173.50	30501
Average	-	Cinnamon teal	544.75	271.21	48813.35

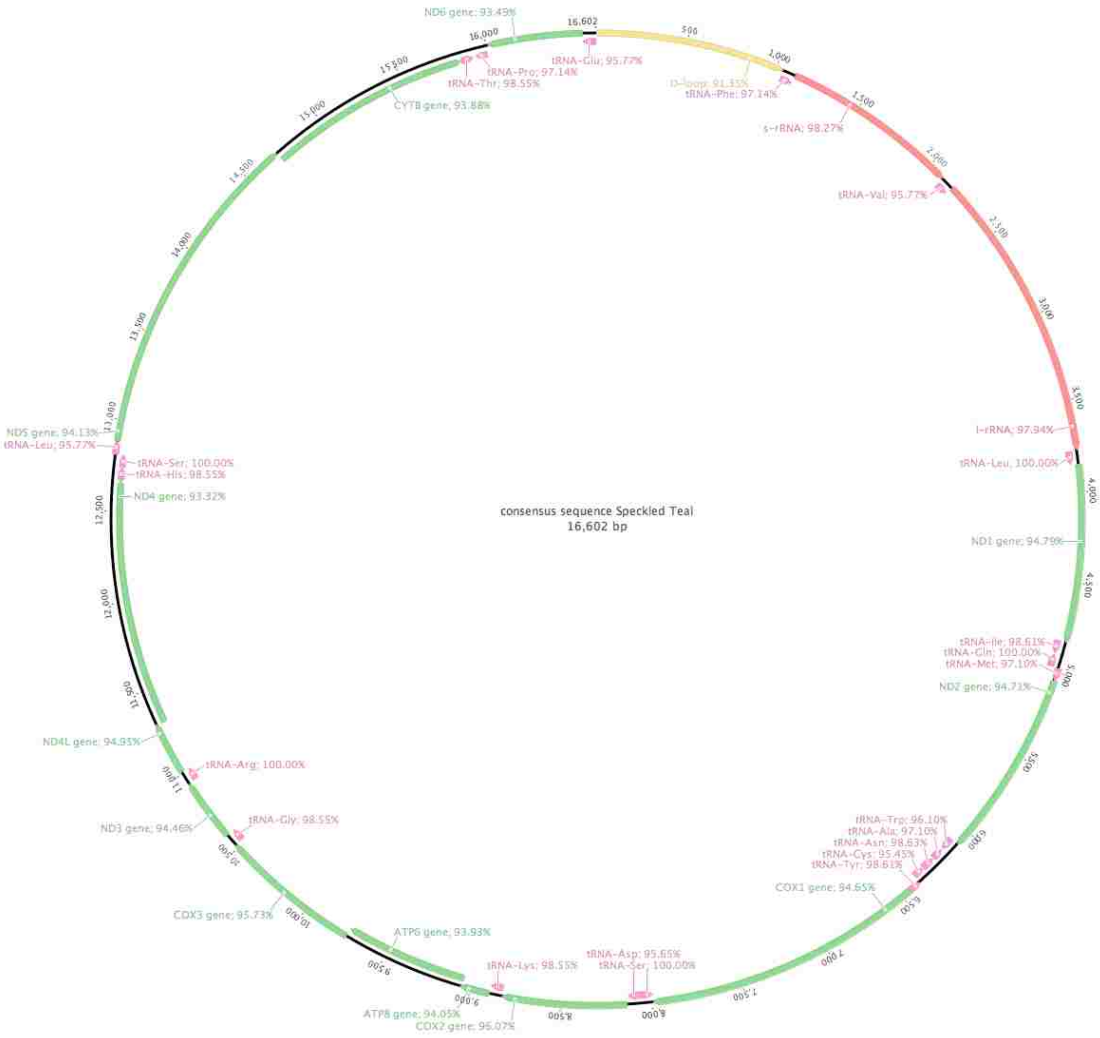

Figure 2.1: Consensus mitochondrial genome with gene annotations for the speckled teal (*Anas flavirostris*). Percentages shown indicate percent identity to the reference (mallard; *Anas platyrhynchos*). Yellow = D-loop; red = rRNAs; green = protein coding gene regions; pink = tRNAs.

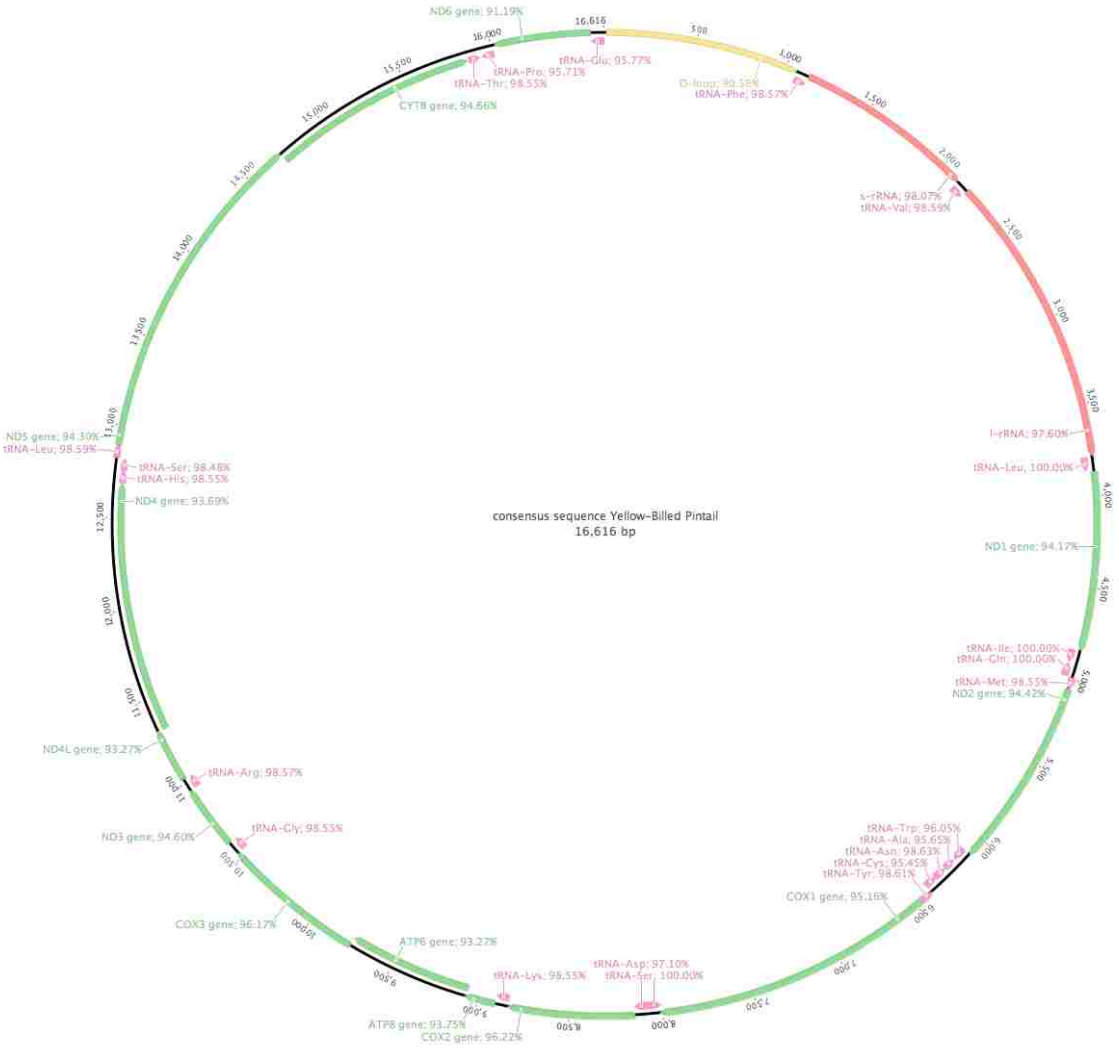

Figure 2.2: Consensus mitochondrial genome with gene annotations for the yellow-billed pintail (*Anas georgica*). Percentages shown indicate percent identity to the reference (mallard; *Anas platyrhynchos*). Yellow = D-loop; red = rRNAs; green = protein coding gene regions; pink = tRNAs.

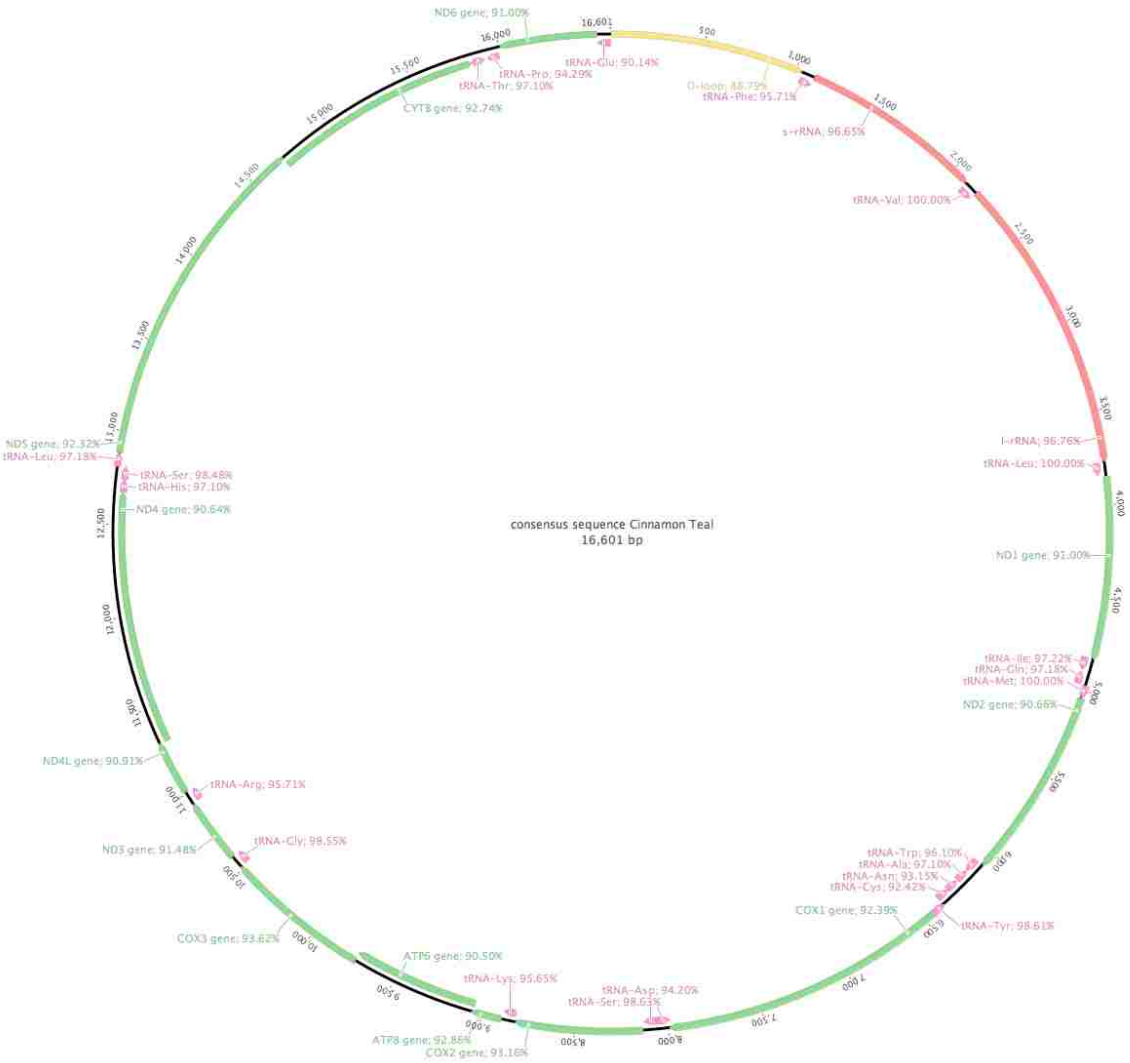

Figure 2.3: Consensus mitochondrial genome with gene annotations for the cinnamon teal (*Anas cyanoptera*). Percentages shown indicate percent identity to the reference (mallard; *Anas platyrhynchos*). Yellow = D-loop; red = rRNAs; green = protein coding gene regions; pink = tRNAs.

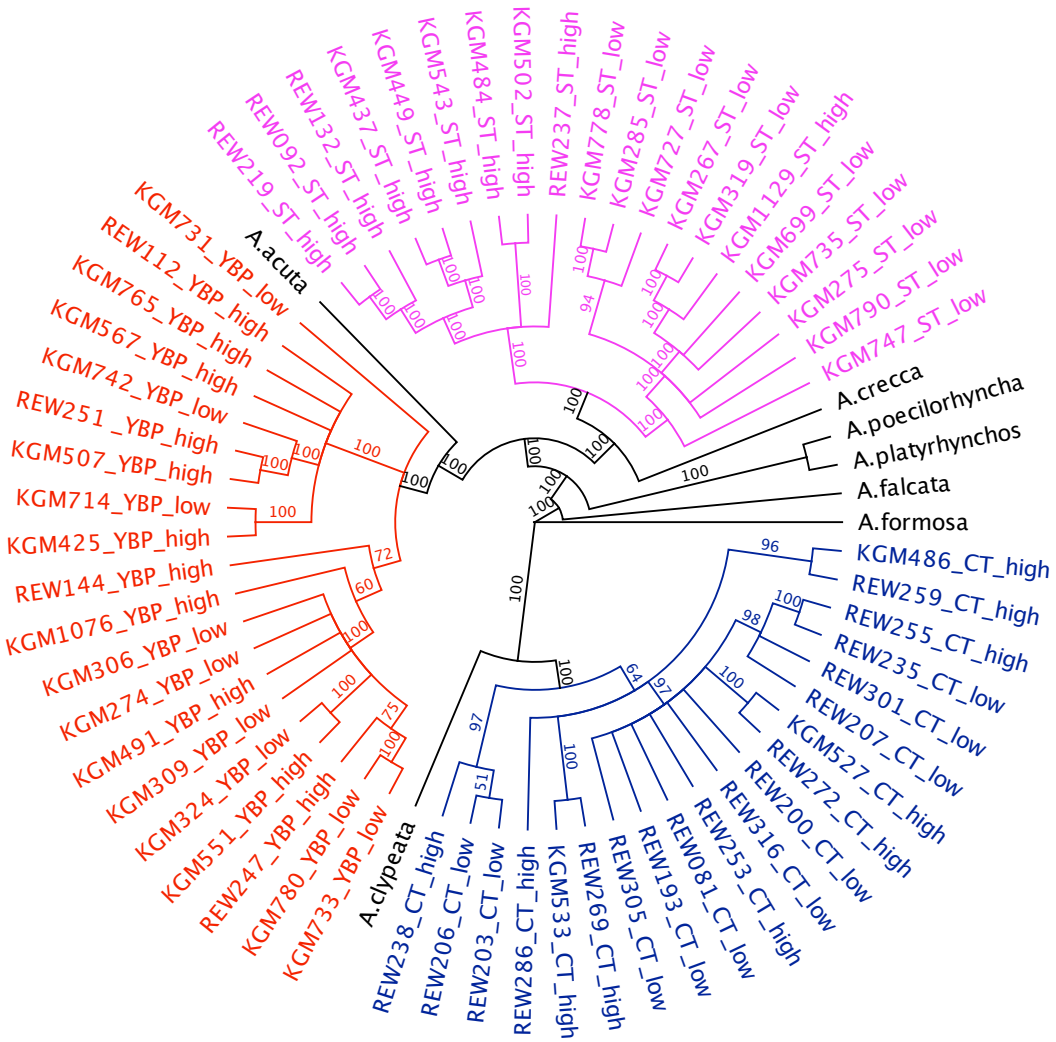

Figure 2.4: Maximum-likelihood tree (GARLI) using the full mitochondrial genomes of the speckled teal (pink), yellow-billed pintail (red) and cinnamon teal (blue), along with other *Anas* outgroups. Numbers at nodes represent percent consensus support; "high" = individuals from high-altitude populations, "low" = individuals from low-altitude populations.

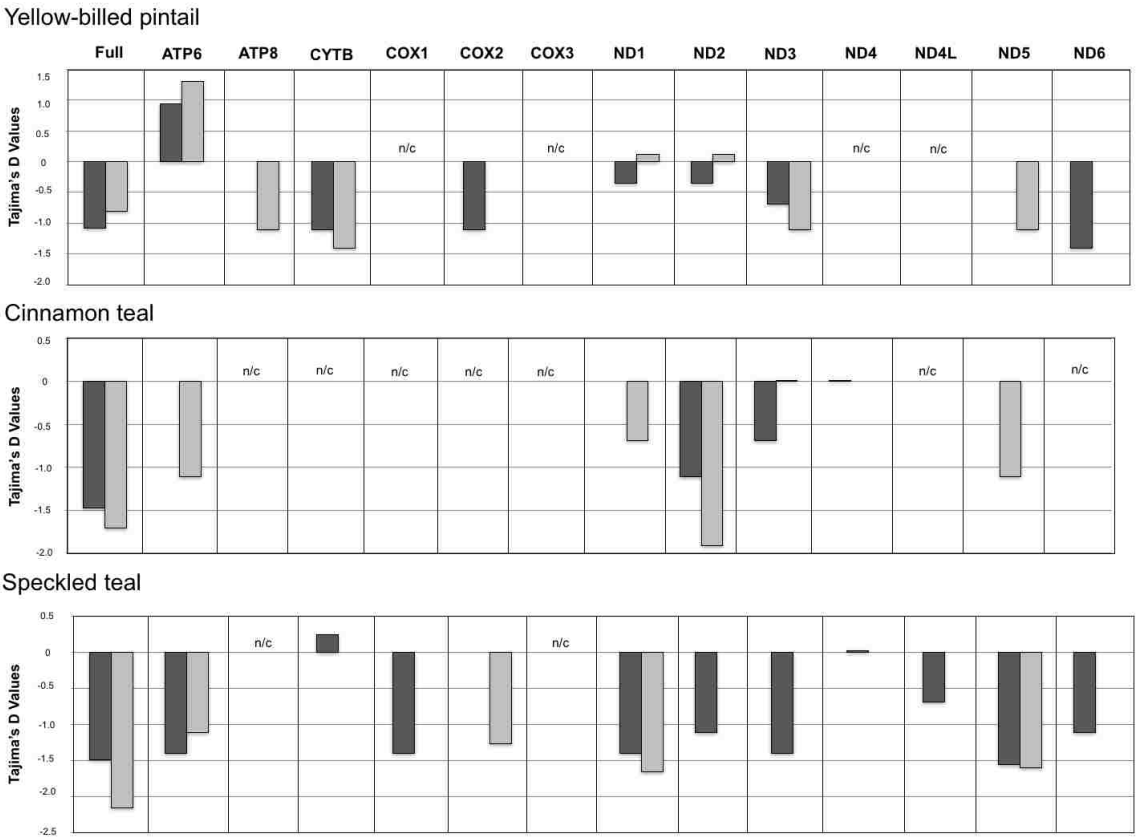

Figure 2. 5: Tajima's D for individual protein-coding genes across each high- (dark grey) and low-altitude population (light grey) of the three species; nc = not calculated, due to a lack of nucleotide variation.

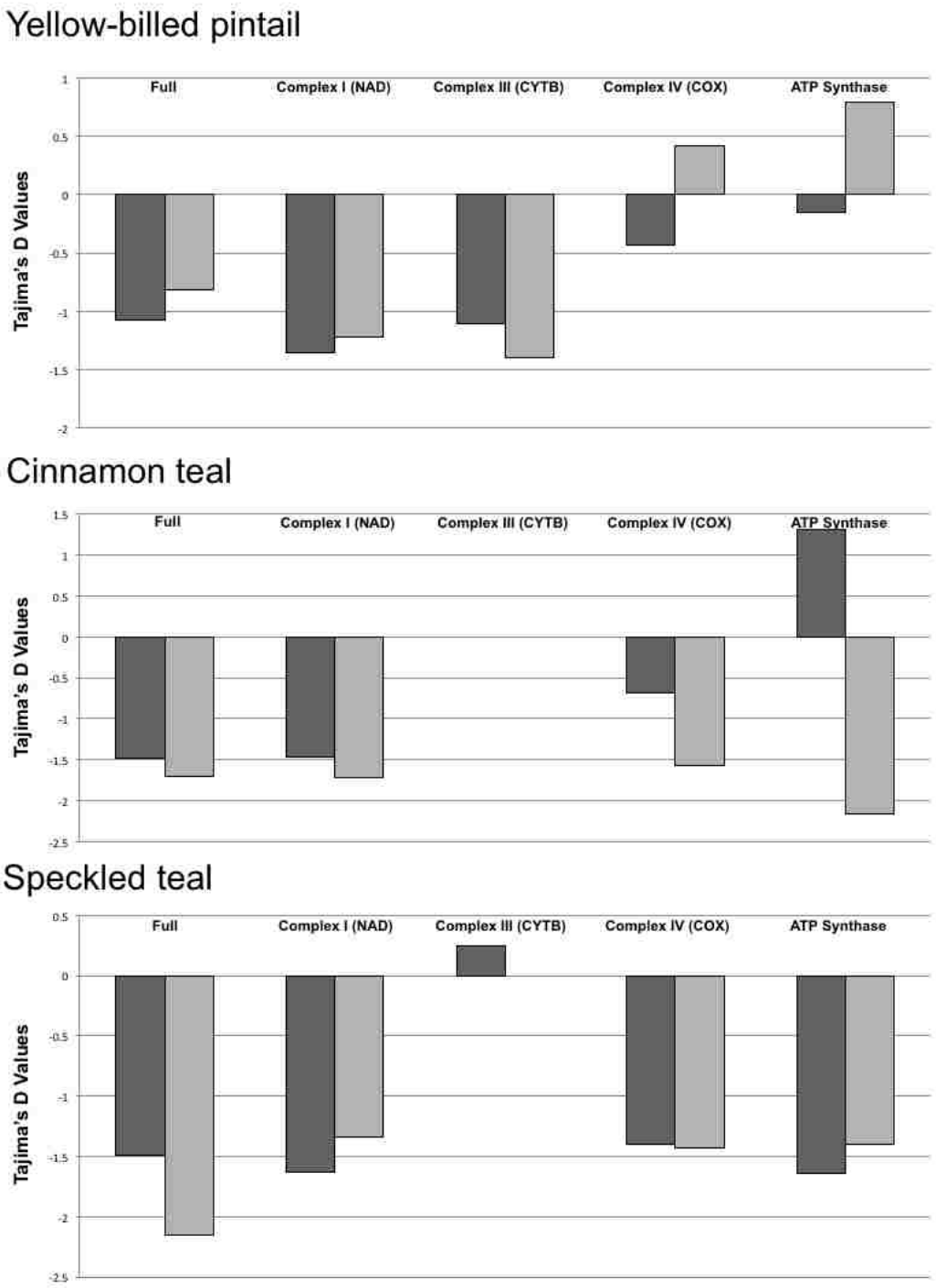

Figure 2.6: Tajima's D for protein-coding genes across five of the mtDNA complexes in high- (dark grey) and low-altitude populations (light grey) of the three species.

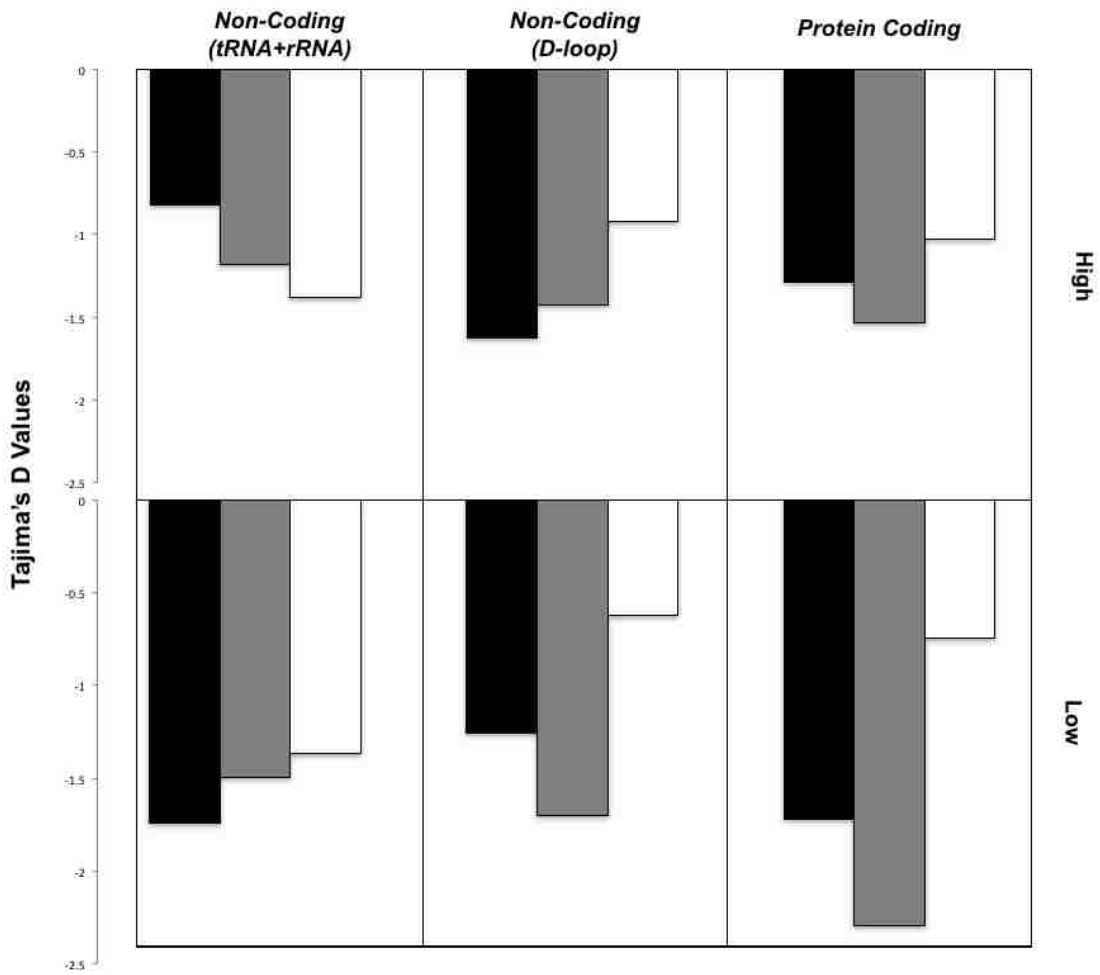

Figure 2.7: Tajima's D for protein-coding and non-protein coding portions (D-loop, tRNAs, rRNAs) of the genome; black bars = cinnamon teal, grey = speckled teal, white = yellow-billed pintail.

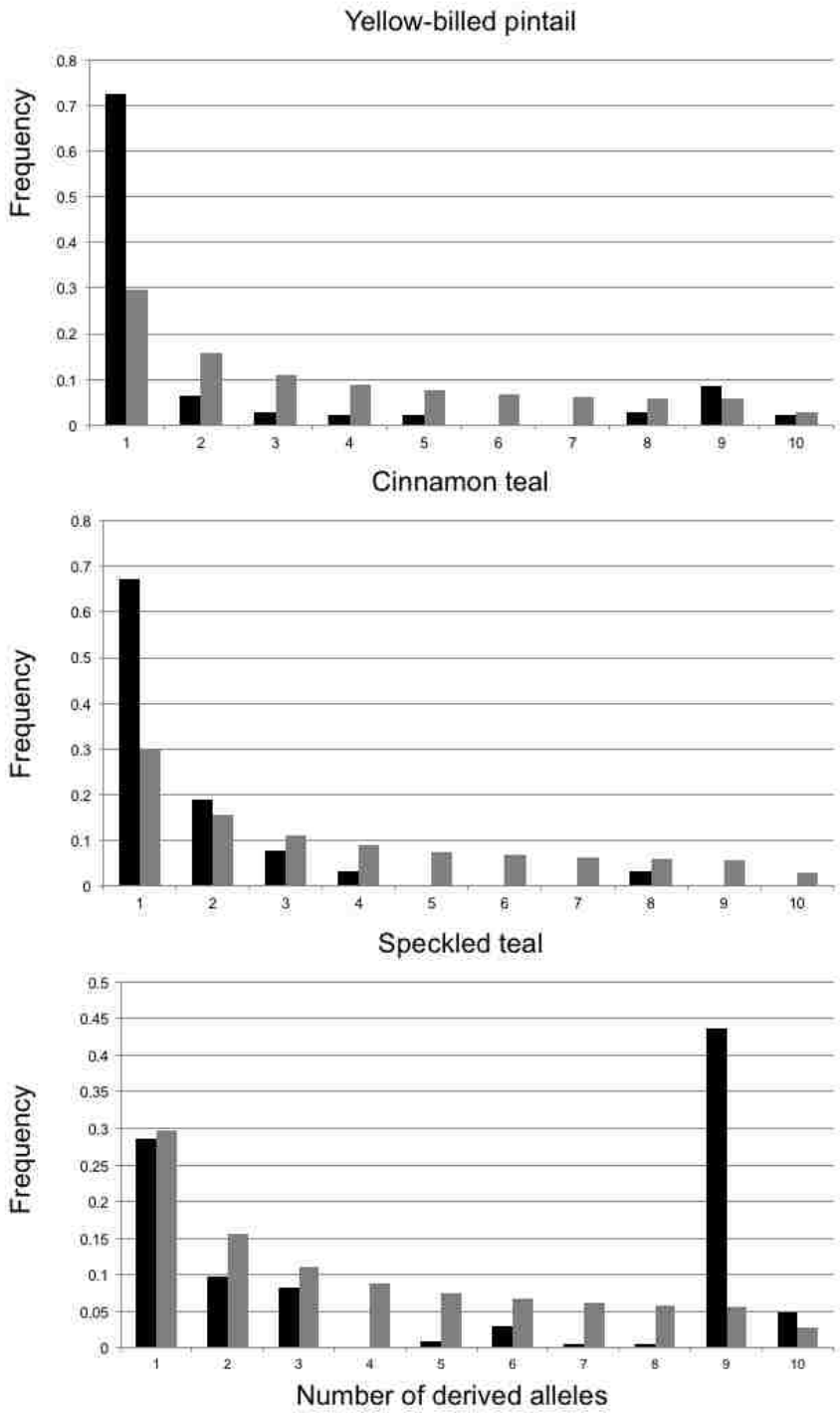

Figure 2.8: Site-frequency spectrum across the full mitochondrial genome for each species (combined high- and low-altitude populations); black bars = frequency observed, grey bars = frequency expected.

Gene	Species	Codon	P-value
ATP6	CT	21	0.051
	CT	46	0.098
	ST	46	0.097
	ST	60	0.058
	YBP	39	0.042
COX1	CT	273	0.001
	ST	273	0.007
	YBP	273	0.007
CYTB	CT	109	0.005
	ST	103	0.039
	YBP	234	0.078
ND1	CT	17	0.071
	CT	315	0.032
	ST	-	-
	YBP	17	0.052
	YBP	184	0.032
	YBP	264	0.010
ND4	CT	54	0.073
	ST	54	0.035
	YBP	54	0.071

Table 2.2: Results of HYPHY- MEME (Mixed Effects Model of Evolution) codon selection analyses; codons which showed evidence of selection in multiple species are highlighted in grey. CT = cinnamon teal; ST = speckled teal; YBP = yellow-billed pintail.

Gene	Codon	P-value	Normalized dN-dS
COX2	167	0.014	0.0968919
	165	0.030	0.194741
CYTB	114	0.016	0.0189694
	132	0.065	0.0507863
ND1	92	0.013	0.0555833
ND5	549	0.026	0.0270495
	561	0.037	0.236315
ND6	117	0.027	0.101825

Table 2.3: Results from HYPHY - SLAC (Single Likelihood Ancestor Counting) codon selection results for speckled teal; cinnamon teal and yellow-billed pintail had none.

Species	Pop.	Selection Test	ATP6	ATP8	Cox1	Cox2	Cox3	CyTB	ND1	ND2	ND3	ND4	ND4L	ND5	ND6
Yellow-billed pintail	High	Positive	0.304	1	1	1	1	1	1	1	1	1	1	1	1
	Low	Positive	0.493	1	1	1	1	1	1	1	1	1	1	1	1
	High	Purifying	1	0.064	0.028	0.011	0.044	0.27	0.156	0.009	0.35	0.003	0.086	0.012	0.001
	Low	Purifying	1	0.097	0.03	0.078	0.056	0.042	0.493	0.119	0.131	0.024	0.083	0.006	0.021
Cinnamon teal	High	Positive	1	1	1	1	1	1	1	0.145	1	1	1	1	1
	Low	Positive	1	1	1	1	1	1	1	1	1	1	1	1	1
	High	Purifying	0.146	1	0.067	1	1	1	0.048	1	0.395	0.007	1	0.049	0.042
	Low	Purifying	0.112	1	0.073	1	1	1	0.296	1	0.165	0.013	1	0.03	0.131
Speckled teal	High	Positive	1	1	1	1	1	1	1	1	0.73	1	1	1	1
	Low	Positive	1	1	1	1	1	1	1	1	1	1	0.147	1	1
	High	Purifying	0.03	0.146	0	0.039	0.001	0.027	0.001	0.007	1	0	0.098	0	0.006
	Low	Purifying	0.266	1	0.012	1	0.018	0.151	0.147	0.041	1	0.146	1	0.136	0.076

Table 2.4: Tests for positive ($H_o = d_N > d_S$) and negative selection ($H_o = d_N < d_S$) through codon-based Z-test across protein-coding gene regions; grey highlighted cells correspond to significant values ($P < 0.05$).

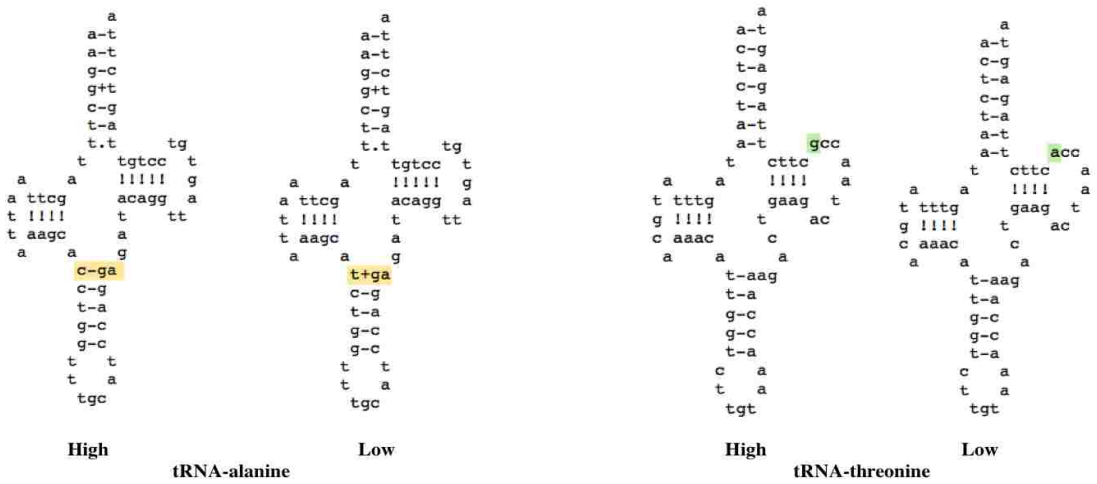

Figure 2.9: t-RNA schematics from ARWEN for tRNA-ala and tRNA-thr. Changes in nucleotide composition between high- and low-altitude populations of speckled teal.

Table 2.5: Outline of how selection and demography can be explained using common tests for neutrality in relation to the results of those analyses in this study

	Effect explained by… (Y/N)				
Effect	Demography (neutral hypothesis)	Selection (alternative hypothesis)	Tests Used	Test robust to demographic effects?	Sp.
Low divergence estimate	Y (recent time since divergence)	Y (purifying/negative directional selection)	F_{ST}	N	CT, YBP
High divergence estimate	Y ("non-recent" time since divergence, female philopatry structuring mtDNA)	Y (positive directional selection)	F_{ST}	N	ST
Rare alleles present at high frequencies (Increased proportion of low frequency variants)	Y (Population expansion after a bottleneck, founder event)	Y (Recent selective sweep, linkage to a swept gene, purifying/negative directional selection)	Tajima's D	N	CT, YBP, ST
			Site-frequency Spectrum	N	CT, YBP
Rare alleles present at low frequencies (Increased proportion of high frequency variants)	Y (population contraction, bottleneck)	Y (Balancing selection, positive directional selection)	Site-frequency Spectrum	N	ST
dN < dS (Less non-synonymous variation than synonymous variation)	N	Y (purifying/negative selection)	Codon-based Z-Test for Selection	Y	CT, YBP, ST
			SLAC (Single likelihood ancestor counting)	Y	CT, YBP, ST
			MEME (Mixed Effects Model of Evolution)	Y	CT, YBP, ST

Chapter 3: Embryonic β^{ε}-hemoglobin (HBE) and upstream transcription factor binding sites are associated with high-altitude adaptation in Andean ducks

Background

Hemoglobin (Hb) has been a flagship molecular marker in studies of adaptation to high altitude (Storz and Moriyama 2008). The Hb protein is a hetero-tetramer containing two α- and two β-chains encoded by corresponding members of the α- and β-globin gene families (Storz and Moriyama 2008; Hoffmann et al. 2010; Stamatoyannopoulos 2005). Both globin gene families have undergone a complex history of gene duplication and divergence over the course of vertebrate evolution (Hoffmann et al. 2008; Storz et al. 2013), providing a source of variation on which adaptation can occur (Storz and Moriyama 2008; Hoffmann et al. 2008).

In short, Hb is responsible for the transport of molecular O_2 from the respiratory system to tissues, where it is then released for cellular use (Dickerson and Geis 1983). At high elevations, the low partial pressure of oxygen (PO_2) results in a reduction in O_2 saturation of arterial blood (Beall 2001). In the absence of physiological or compensatory mechanisms that improve O_2 transport, low O_2 (hypoxia) in the tissues can be detrimental to an animal's metabolism and thus capacity for sustained physical activity.

Waterfowl, as a group of birds, have featured prominently in studies of Hb function and evolution in response to high-altitude adaptation. Amino acid substitutions have been found that increase O_2-binding affinity of the HbA (major) isoforms by affecting α- and β- subunit interactions and sensitivity to allosteric cofactors like IP_5 (inositolpentaphosphate) (Weber et al. 1993; Perutz 1983). In waterfowl, the α^A and β^A

subunit genes (HBA and HBB, respectively) exhibit a high degree of parallel amino-acid substitution resulting in increased Hb-O_2 affinity in the high-altitude lineages (McCracken et al. 2009a; McCracken et al. 2009b; Natarajan et al. 2015). Other bird species, as well as mammals, show differences in Hb-O_2 affinity caused by specific amino acid substitutions in the Hb subunits, consistent with parallel evolution in response to low-O_2 environments (Projecto-Garcia et al. 2013; Storz et al. 2009; Revsbech et al. 2013; Cheviron et al. 2014; Tufts et al. 2014; Nery et al. 2013). Although the function of Hb in adapting to low- O_2 environments is well known, most studies have relied on sequencing only the genes encoding the adult Hb protein isoforms (i.e., α^A, α^D, β^A in birds), excluding from analysis the embryonic genes in the two globin clusters (α^π, β^ε), as well as the up-stream regulatory elements containing transcription factor binding sites (TFBS).

Embryonic hemoglobin

Synteny is considerably conserved among avian genomes, in part due to relatively low rates of gene turnover in multigene families (Backström et al. 2008; Ellegren 2010). Overall, this same pattern is generally apparent in the α- and β-globin clusters, with a high degree of stability and conservation in Hb gene family evolution across avian lineages (Opazo et al. 2015). Comparing the globin clusters across birds, the α-globin cluster is the more stable, whereas the β-globin has undergone changes in gene composition. Both chicken and mallard for example have the same complement of α-globin genes (α^E/sometimes also known as α^π, α^D, α^A). However, the chicken complement of β-globin includes four genes (β^H, β^A and 2 copies of *epsilon*/β^ε), whereas

the mallard complement includes three (*rho*/β^{P}, β^{A} and β^{ε}) (Opazo et al. 2015). However, the current inclusion of β^{P} in the mallard β-globin cluster is uncertain, due to possible pseudogenization or truncation, which is a common occurrence in the hemoglobin clusters (Hoffmann et al. 2010; Opazo et al. 2015; Storz et al. 2013).

Nonetheless, in general, adult erythroid cells express Hb tetramers comprised of two subunits, one each of α and β respectively (Bulgarella et al. 2009). In the case of ducks, the HbA ($\alpha_2^{A}\beta_2^{A}$) is the most common adult isoform comprising between 83-68% of all blood cells, whereas HbD ($\alpha_2^{D}\beta_2^{A}$) is the least common isoform comprising 31-19% of blood cells (Figure 3.1)(Natarajan et al. 2015; Opazo et al. 2015). However, these adult isoforms are not substantially represented during embryonic development; the embryonic/fetal Hb isoforms instead are the most common. In mammals, embryonic Hb is present both very early in development and for a relatively short period of time (i.e., first 6-weeks of development) with gene expression eventually transitioning to fetal Hb, which stays up-regulated until about 12-weeks postnatally. A similar process occurs in birds. During development, erythroid cells are comprised of some combination of embryonic subunits (ε, π), and adult globin subunits: HbP ($\alpha_2^{\pi}\beta_2^{\varrho}$), HbM ($\alpha_2^{D}\beta_2^{\varepsilon}$), HbE ($\alpha_2^{A}\beta_2^{\varepsilon}$), HbL ($\alpha_2^{D}\beta_2^{H}$) and HbH ($\alpha_2^{A}\beta_2^{H}$). The primitive erythroid cells produce these five globins combinations, until 5-18 days post-egg laying, which then transitions to an up-regulation of primarily adult globin isoforms (Chapman et al. 1982; Cirotto et al. 1987; Chapman et al. 1980, 1981; Gou et al. 2007). However, such a characterization of duck hemoglobin complements are currently unavailable, though are likely comprised of some combination of HbM, HbE and HbL, since there is no β^{H} in the duck genome (Opazo et al. 2015; Cirotto et al. 1980). It is interesting to note that embryonic Hbs are sometimes

up-regulated in the presence of hypoxic conditions in some mammals (Risso et al. 2012; Reynafarje et al. 1975); however, the same phenomenon is currently unknown in bird species.

In humans, although embryonic, fetal and adult Hb have very similar overall structures, fetal/embryonic Hbs generally are characterized by having higher O_2 affinity that enhances the O_2 diffusion gradient from the maternal circulation across the placenta to the fetal circulation. This can arise from amino acid substitutions at positions altering the arrangement of intersubunit contacts (Perutz 1989; Manning et al. 2007) or that influence interactions with allosteric effectors (He and Russell 2001; Hofmann and Brittain 1996). It has also been suggested that higher O_2-affinity of embryonic Hb in oviparous species (i.e., birds, reptiles) enhances the transfer of O_2 from the atmosphere through the pores of the eggshell, across the allantoic membrane to the embryo (Isaacks and Harkness 1980). Adaptive changes in the adult Hbs at high-altitude thus mimic their duplicated fetal/embryonic counterpart genes, by more left-shifted O_2-saturation curves, causing a greater Hb-O_2 binding affinity, characterized by ½ saturation of the Hb at a lower PO_2 (i.e., P_{50}). However, it is currently unknown to what extent avian embryonic Hb is also under selection for similar properties at high altitude. For example, if embryonic Hb is left-shifted relative to adult Hb due to constraints placed on development in the egg, do we expect to see further increases in Hb-O_2 affinity in embryonic Hb of species native to high altitude compared to low altitude?

Upstream regulatory elements

Transcription factors are recruited to the upstream regulatory elements of genes through transcription factor binding sites (TFBS; Figure 3.2). These establish DNA-protein interactions between regulatory sequences and specific DNA-binding transcription factors, either up-regulating, down-regulating, or inhibiting transcription (Spitz and Furlong 2012; Lenhard et al. 2012). Genetic variation in TFBS thus has the potential to modify transcriptional regulation in different ways ranging from binary responses to subtle modifications of spatiotemporal patterns of gene expression (Erceg et al. 2014). Therefore, genetic variations occurring within TFBS are often considered as regulatory mutations.

Like most genes, the timing and expressional levels of genes within the α- and β-Hb gene clusters are regulated by a suite of transcription factors (Cao and Moi 2002; Sankaran and Orkin 2013; Knezetic and Felsenfeld 1993; Merika and Orkin 1995). Such interactions have been found to change Hb expression (Gilman et al. 1988; De Gobbi et al. 2006), including adaptive responses to hypoxic environments (Gorr et al. 2004; Gou et al. 2007). Differences in gene expression, driven by transcription factors, furthermore are thought to be a major contributor to speciation and phenotypic diversity (Kasowski et al. 2010; Borneman et al. 2007; Boyle et al. 2014). However, the extent to which they generally are associated with the proximal cis-promoter regions upstream of each of the globin genes, has been largely untouched in the literature.

In summary, both embryonic Hb and the TFBS associated with proximal-promoter regions across these gene clusters might have been affected by selection in the same way polymorphisms in the major Hb genes have been selected for increased Hb-O_2

affinity. In this study, I evaluated whether variation in embryonic paralogs (α^{π}, β^{ε}) of the α- and β-Hb clusters, as well as the proximal promoter regions containing TFBS for all Hb genes, are associated with high-altitude adaptation in three Andean duck species. I used statistical outlier tests to determine polymorphisms with elevated F_{ST} values, thus exhibiting significantly elevated differentiation, and compared it to a background of genome wide reference sequences (ie. RAD-seq). I also modeled changes in protein structure associated with significant coding-region changes, and additionally assessed whether variants in the cis-regulatory regions interacted with predicted TFBS regions.

Methods

Specimen collection and DNA extraction

A total of 60 individuals were used for this study from three different Andean duck species: cinnamon teal (*Anas cyanoptera*), speckled teal (*A. flavirostris*), and yellow-billed pintail (*A. georgica*). Sampling from each of the three species is comprised of 20 individuals with 10 individuals from each of the high-altitude populations and 10 individuals from each of the low-altitude populations. For the cinnamon teal, individuals from low-altitude populations are the *A. c. cyanoptera* subspecies (*n* = 10; elevation range 7-23 m) and from high-altitude are the *A. c. orinomus* subspecies (*n* = 10; elevation range 3533-3,871 m) (Wilson et al. 2013). For the speckled teal, individuals from low-altitude populations are the *A. f. flavirostris* subspecies (*n* = 10; elevation range 77-860 m) and from high-altitude are the *A. f. oxyptera* subspecies (*n* = 10; elevation range 3,211-4,405 m). For the yellow-billed pintail, individuals from both populations are taxonomically identified as *Anas georgica spinicauda*. A total of 20 yellow-billed pintails

were collected from low- (n = 10; elevation range 292-914 m) and high-altitude (n = 10; elevation range 3,332-4,070 m). Genomic DNA was extracted from tissue using a DNeasy Tissue Kit (Qiagen, Valencia, California, USA) following manufacturers protocols.

Target-enrichment sequencing, sequence assembly and annotation

I utilized in-solution target capture to selectively enrich libraries for the α and β-globin cluster genes and upstream *cis*-located transcription factor binding sites prior to NGS sequencing (Gnirke et al. 2009). All steps of the process were performed by MYcroarray (Ann Arbor, MI). Custom MYbaits® biotinylated ssRNA target capture baitsets were designed. Specifically, a custom set of 210 (α-globin) and 579 (β-globin) 120mer probes at 2x tiling density was designed from the *Anas platyrhynchos* (mallard) genome (BGI_duck_1.0, GCA_000355885.1). The DNA samples were sonicated, size-selected, and converted to Illumnia libraries using dual indexes (P5, P7) during library amplification. These libraries were then enriched using the MYbaits® v3.0 procedure (Gnirke et al. 2009), quantified using qPCR, and then pooled for sequencing. Sequencing was performed on an Illumina Hi-Seq v4 platform paired-end (100 bp) with a 250-300 bp insert size.

Sequences were received pre-parsed by individual with adapters trimmed and quality filtered (Q < 30). Additionally, the sequencing data was trimmed of adapter sequences (*fastq-clipper*), then filtered by length and quality using the FASTX-Toolkit v. 0.0.13 package (Gordon and Hannon 2010). A reference-guided assembly was performed by Bowtie2 (Langmead and Salzberg 2012) under highest sensitivity settings in the

Geneious package (Kearse et al. 2012), against the α-globin (NW_004678373, scaffold 2065) and β-globin (NW_004682656, scaffold 6035) sequences of the mallard. Thus, the annotations were pulled from NCBI's most recent version of the scaffold (July 2015). Consensus sequences were then extracted for each individual, and aligned separately by species using MAFFT (Katoh et al. 2002). The intronic and exonic Hb regions for each of the α subunits (α^{π} = LOC101800713; α^{D} = LOC101800520; α^{A} = LOC101800332) and β subunits (β^{ε} = LOC101798290; β^{A} = LOC101798111) clusters were then annotated using genomic features available through the mallard genome on NCBI.

It is important to note here, that there is a discrepancy specifically associated with β^{P}, in that annotations associated with β-globin (NW_004682656, scaffold 6035) do not list it as a gene, but rather an alternate isoform of the gene model for β^{A}; therefore, it was not included in the current set of analyses as a gene.

Proximal promoter regions were designated as the region 1,000 bp upstream of annotated transcription start sites (TSS) for each gene, based on general definitions from previous studies (Tabach et al. 2007; Bajic et al. 2006; Linhart et al. 2008).

RAD-Seq reference data

To generate a genome-wide reference data set to compare HIF-pathway genes to, I used RAD-Seq (Restriction Site Associated DNA Sequencing). Genomic DNA was normalized to around 2ng/uL using a Qubit Fluorometer (Invitrogen, Grand Island, New York, USA) and submitted to Floragenex (Eugene, Oregon, USA) for sequencing. In short, genomic DNA was first digested with the 8 base-pair SbfI restriction enzyme (CCTGCAGG) followed by barcode and adapter ligation. Individually barcoded samples

were multiplexed and sequenced on the Illumina HiSeq 2000 with single-end 100 bp sequencing chemistry. Following the run, RAD sequences were de-multiplexed and trimmed to yield resulting RAD sequences of 90 bp. The methods used by Floragenex are described in (Hohenlohe et al. 2010; Miller et al. 2007; Baird et al. 2008), and are summarized below.

Genotypes at each nucleotide site were determined using the VCF_popgen v.4.0 pipeline to generate a customized VCF 4.1 (variant call format) database with parameters set as follows: minimum AF for genotyping = 0.075, minimum Phred score = 15, minimum depth of sequencing coverage = 10x, and allowing missing genotypes from up to 2 out of 20 individuals (10%) at each site. To filter out base calls that were not useful due to low quality scores or insufficient coverage, genotypes at each nucleotide site were inferred using the Bayesian maximum likelihood algorithm. The genotyping algorithm incorporates the site-specific sequencing error rate, and assigns the most likely diploid genotype to each site using a likelihood ratio test and significance level of $p = 0.05$. Sites for which the test statistic is not significant are not assigned a genotype for that base in that individual, effectively removing data from sites for which there are too few high-quality sequencing reads. The sequencing coverage and quality scores were then summarized using the software VCFtools v.0.1.11 (Danecek et al. 2011). Custom perl scripts written by M. Campbell (University of Alaska Fairbanks) were used to retain bi-allelic sites only and converted to VCF file format. Ultimately, these SNPs were then incorporated into the final Hb dataset used for outlier analyses (see next section).

Population genetic analyses

Basic genetic diversity estimates and F_{ST} between low- and high-altitude populations were calculated using Arlequin (Excoffier et al. 2005). These analyses were performed for separate partitions of the data, including [1] individual gene regions, with a focus on embryonic Hb subunits (π, ε) and [2] potential upstream proximal promoters of each gene (1,000 bp). Variant calls were created separately in order to be fed into outlier analyses - a custom pipeline (https://github.com/amgraham07) was created to remove orphan sequences, and assemble sequences against the α-globin (NW_004678373, scaffold 2065) and β-globin (NW_004682656, scaffold 6035) sequences of the mallard reference, using BWA v0.7.15 (Li and Durbin 2009). The Samtools package v1.3.1, including BCFtools v.1.3.1 (Li et al. 2009) was then used to create a VCF file, as well as provide assembly statistics (i.e., bp-by-bp coverage). These programs used in the pipeline called SNP variants that were different against the mallard reference, including indels (insertion/deletion); however, the indel information was excluded in the final dataset since the softwares used do not accommodate indels.

I used a Bayesian approach as implemented in BayeScan v. 2.1 to again identify putatively selected loci, and to broadly assess linkage disequilibrium across the cluster. BayeScan uses a logistic regression model to separate locus-specific effects of selection from demographic differences (Foll and Gaggiotti 2008). For each SNP, BayeScan estimates the posterior distribution under neutrality $\alpha = 0$ and separately allowing for selection $\alpha \neq 0$ and computes the posterior odds ratio (PO) as a measure of support for the model of local adaptation relative to neutral demography.

Significant variants from the AMOVA which resulted in nonsynonymous variation in either of the embryonic Hb subunits were then located on the human embryonic protein structure to predict potential change in function (Protein Data Bank ID: 1a9w) using PyMOL 1.7.4 (DeLano 2002); the low- and high-altitude genotypes were modeled using the "mutagenesis" function.

Finally, significant variants from the AMOVA located in the potential upstream promoters of each gene were assayed in a separate manner to determine overlap with transcription factor binding sites (TFBS) (see next section).

Transcription factor binding site identification

Two different assessments were employed in an effort to evaluate the extent to which regulation of Hb genes might occur via TFBS. The current promoter set for Hb gene clusters in chicken (*Gallus gallus*) are not clear, and are unknown in mallard (*Anas platyrhnchos)*, therefore computational predictions were utilized.

The locations of potential transcription factor binding sites were performed in MatInspector (Cartharius et al. 2005; Quandt et al. 1995) using the following options: matrices with 0.95 core similarity in vertebrates, in addition to searching for corresponding promoters in chicken. The α- and β-globin clusters of the reference mallard sequence, as well as each of the 60 individuals were subjected to these analyses. Specifically, the regions 1,000 bp upstream of each gene in the clusters were designated as the proximal-cis regulatory region(s), as defined in the previous section. Summaries, or counts of the predicted TFBS in the form of TFBS matrices (or matrix families) summaries, were created between the high- and low-altitude populations. For each of the

TFBS matrices, Mann-Whitney-U tests were performed to assess significant differences in predicted counts of TFBS. Any significantly different ($P > 0.05$) families were queried for their associated Gene Ontology (GO) terms as characterized by Genomatrix's Matrix Library (version 9.3, March 2015).

In addition to general trends associated with TFBS number, significant SNPs found in the AMOVA ($P < 0.05$), which also fell within the proximal-cis regulatory regions, were queried for whether they overlapped (start, end, or anchor position) with any TFBS locations identified by MatInspector.

Results

Hemoglobin assembly statistics

Coverage was calculated for both the Bowtie2 (contiguous sequence assembly) and BWA (VCF pipeline) assemblies. For Bowtie2, the average coverage for the three duck α-globin clusters was 1,022x with an average of 110,779 reads per individual assembly containing an average 49.60% GC content; whereas, the average coverage for the three duck β-globin clusters was 275x with an average of 120,586 reads per individual assembly containing an average 49.06% GC content. For BWA, across the three species, the per base-pair coverage was 279x covering 92.5% of the β-globin cluster, and 525x covering 97.5% of the α-globin cluster (Table 3.1-3.3).

Population differentiation of coding regions and upstream regions

Pairwise F_{ST} between low- and high-altitude populations was calculated for each individual gene, or upstream TFBS region, as well as across the entire Hb clusters (Table

3.4). These regions were parsed separately in order to deduce whether there were significant correlations with coding regions, putative upstream regulatory regions, or both. In addition, each Hb cluster was also subjected to an outlier analysis through BayeScan.

For the α-globin cluster in cinnamon teal the highest divergence was found in the α^A ($F_{ST} = 0.831$). For speckled teal, both the α^D and α^A genes had moderate levels of divergence ($F_{ST} = 0.532$, and 0.477); whereas, the yellow-billed pintail showed no significant differentiation in any of the three α-globin genes (Table 3.4). The upstream putative TFBS regions of the α-globin gene cluster did not show similar divergence estimates, despite only being 1,000 bp away from sites with elevated F_{ST}. Two of the three species showed lower overall F_{ST} for their upstream α^D and α^A regions. However, the putative upstream TFBS region of the embryonic α^π gene had a much higher F_{ST} (0.676) in the speckled teal, with neither cinnamon teal nor yellow-billed pintail showing similar levels (Table 3.4). For the β-globin cluster, speckled teal and yellow-billed pintail showed moderate to high F_{ST} for both β and β^ε. The upstream putative TFBS regions of the β-globin genes mirrored the F_{ST} estimates for their corresponding individual gene estimates with the cinnamon teal β and β^ε showing no signs of differentiation, whereas both speckled teal and yellow-billed pintail showed extremely high F_{ST} (ST, $F_{ST} = 0.646$, and 0.825; YBP, $F_{ST} = 0.608$, and 0.941; Table 3.4).

Bayesian outlier analyses showed large amounts of significant values (Log10(PO) < 0.5) starting at the epsilon gene and continuing until the end of the HBB gene, also covering the intergenic region between the two genes, covering close to 17k bp (Figure 3.3). It is also important to note that there is a pronounced dip in Log10(PO) values

following HBB, then an increase in values on the 5' end of the cluster (the beta globin sequence are antisense), likely due to the presence of another gene on the scaffold olfactory receptor 51G2; although no such significance is found within another olfactory receptor gene (51L1) present on the 3' end of the scaffold.

Embryonic Hb genes and amino-acid substitutions

SNPs with significantly elevated F_{ST} ($P > 0.05$) were found in the embryonic/fetal β^{ε} gene. This was the case for yellow-billed pintail and speckled teal, but not for cinnamon teal (sees previous section). Thus, sequences were then examined for corresponding changes in their amino-acid composition (nonsynonymous substitutions) in the three exons, as well as differences in amino-acid biochemistry between the ancestral and derived residues (e.g. hydrophobic vs. hydrophilic, acidic vs. basic, aliphatic vs. aromatic). There were four amino-acid changes that were identical in both species: β^{ε}14Ser $\rightarrow$ Gly, β^{ε}15Ile $\rightarrow$ Leu, β^{ε}17Ser $\rightarrow$ Gly, and β^{ε}126Ala $\rightarrow$ Thr. For speckled teal, all four sites were fixed for different amino acids between populations (F_{ST} = 1.0), whereas allele frequencies were very highly divergent but not fixed in yellow-billed pintail (F_{ST} = 0.778) (Table 3.5). The first three of these at positions 14, 15, and 17 occurred in the first helical turn of the first α subunit, and the last one at position 126 in the last helical turn (Figure 3.4, 3.5). The locations of these amino-acid substitutions were located on the PDB human embryonic protein structure.

Transcription factor binding sites (TFBS) and change in TFBS identity

First, the upstream regions of each gene in both clusters were assayed for potential TFBS based on matrices with a 0.95 core similarity in vertebrates. This was performed for all individuals to evaluate the extent to which genetic variation may, or may-not, have changed TFBS identity between high- and low- altitude populations in each species. The TFBS matrices that had significant deviations in counts across the different populations varied widely from transcription factors that have general functions, to transcription factors with very specific functions (Table 3.6, 3.7). However, given the known physiological role of Hb, the specific functions of each of the TFBS matrices were queried using a set of key words: "blood", "HIF", "hypoxia", "hemato-", "erythro-" or "oxygen". Of the TFBS matrices that had significant deviations between high- and low-altitude populations, between 41.9-45.6% had a reference to those key words in their GO terms (α, 18/43 TFBS matrices; β, 21/46 TFBS matrices).

In terms of parallel patterns seen across all three species, the results showed little-to-no similarities between types of TFBS matrices that had different representation between high- and low-altitude populations across the three species; however, there was more substantial overlap across the β-globin cluster, especially between speckled teal and yellow-billed pintail (Table 3.7). In addition, across all matrix/TF families that differed between high- and low- altitude populations, between 62.3-69.8% were cases in which the high-altitude populations had more instances of TFBS of that category.

Transcription factor binding sites and overlap with genetic variation

Second, the upstream regions of each gene in both the α- and β-globin clusters were assayed for site-by-site F_{ST}. Single nucleotide polymorphisms (SNPs) with significant F_{ST} ($P < 0.05$), in addition to overlapping with any TFBS locations (start, end, or anchor position), were considered a potential candidate.

In the α-globin cluster, the upstream region associated with the with the α gene, had four significant sites in the speckled teal, whereas the cinnamon teal had one significant site (Table 3.8, 3.9). The upstream region associated with α^{π} (embryonic α globin) had one significant site in both speckled teal and yellow-billed pintail. The most variation was seen upstream of α^{D}, with five significant sites for cinnamon teal, two sites for yellow-billed pintail and one for speckled teal. In contrast, the β-globin cluster upstream regions showed considerably more genetic variation, with the upstream regions associated with β^{A} and β^{ε} having multiple significant sites associated with them in both the speckled teal (β^{A}, 21 SNPs; β^{ε}, 24 SNPs) and yellow-billed pintail (β^{A}, 6 SNPs; β^{ε}, 26 SNPs), whereas the cinnamon teal had none (Table 3.8, 3.10).

These same variants also overlapped significantly with predicted TFBS locations, between 83-93% of SNPs overlapping with a start/stop/anchor position of a potential transcription factor (Tables 3.9, 3.10). In addition, across both clusters these variants were also largely associated with multiple TFBS - with two or more TFBS in which start/stop/anchor positions overlapped with each other (α, 14/17 = 82.3%; β, 27/34, 79.4%; Tables 3.9, 3.10).

Discussion

In this study, I present a comprehensive analysis of the Hb gene clusters of Andean ducks, describing their embryonic Hb genes and upstream promoter regions of each protein-coding gene. These results show evidence for previously undescribed mechanisms for adaptation to high-altitude environments (i.e., hypoxia) via genetic variation in [1] the embryonic β^{ε}-globin gene and [2] proximal promoter regions, specifically in transcription factor binding sites (TFBS), associated with transcription factors known to have Hb-pertinent gene ontology functions.

Embryonic Hb and high-altitude adaptation

High-altitude presents strong selective pressure on reproductive fitness, in a variety of organisms, through hypoxia and cold temperatures. One primary physiological constraint that is sometimes overlooked is on fetal/embryonic development. For example, humans from low-altitude populations who live in high-altitude environments (>2,500 m) are associated with increased incidence of preeclampsia and intrauterine growth restriction (IUGR), low infant birth weight, both leading to increased infant mortality (Moore et al. 2004). More relevantly, this same phenomenon has been observed in birds and waterfowl species in particular, with low egg hatchability for individuals from low-altitude, presumably due to embryonic susceptibility to hypoxia (Visschedijk et al. 1980; Monge and Leon-Velarde 1991; Leon-Velarde et al. 1984; Carey et al. 1994; Gou et al. 2007). Thus, embryonic development in high-altitude environments potentially imposes a more severe selective constraint than on an adult's fitness because all individuals must pass through this bottleneck. This could result in multiple adaptations across different

life-stages, perhaps to a greater degree in early development. Therefore, embryonic Hb may also be under similarly strong selection for increased O_2-binding, in the same way adult waterfowl Hb has been shown to have increased Hb-O_2 affinity (Natarajan et al. 2015).

The results of this study suggest potential mechanisms of adaptations to high-altitude through embryonic Hb in speckled teal and yellow-billed pintail, specifically on the β^{ε} subunit. Allele frequencies that were highly differentiated between populations also were found to occur in exonic regions resulting in amino-acid replacements in the first and last α-helices of the β^{ε}-chain subunit (Table 3.5; Figure 3.4). Based on currently published structure of embryonic human hemoglobin (Protein Data Bank - PDBe › 1a9w), the locations of changes found in this study are peripheral to the locations of metal/heme binding (i.e., β^{ε} 64, β^{ε} 93) (Figure 3.5). However, this does not rule out the possibility of those amino-acid replacements, seen in high-altitude populations, having functional consequences on O_2-binding affinity, as even slight changes resulting in small structural differences such as the elimination of a single H-bond can cause profound changes in Hb-O_2 affinity (Natarajan et al. 2015). In this case, three of the four amino-acid replacements show substantially different biochemical properties than their ancestral low-altitude allelic variants, and can be considered non-conservative mutations. The changes seen in this study thus represent differences in polarity, which are theoretically less likely to occur due to functional constraint associated with purifying selection (Taylor 1986; Swanson 1984; Ng and Henikoff 2006), and their presence suggests an adaptive function of those replacements.

Finally, although our data cannot be phased across the gene clusters, the variants seen in β^{ε} for both species are in linkage disequilibrium with the adult $\beta^{A}116$ and $\beta^{A}133$ Hb variants that have been identified in prior studies (McCracken et al. 2009a; McCracken et al. 2009b). Potential reasons for this stem from general proximity between these genes (~4.1k bp), and population subdivision/population bottlenecks, in combination positive epistasis, and/or depressed recombination. Epistasis, or non-additive interactions between mutations, has been documented in the Hb clusters previously, although within the same gene, or between the α- and β-globin subunits (Tufts et al. 2014; Natarajan et al. 2013; Projecto-Garcia et al. 2013). Between genes, epistasis is known to occur due to interactions of those two genes, usually as a part of a multi-component protein, if the protein encoded by one gene modified the other (i.e., directly), or if both genes encode for components of the same pathway/network (i.e., indirectly)(Phillips 2008). The other possibility is linkage disequilibrium due to depressed recombination rates across the epsilon and HBB.

There are also several potential demographic reasons for increased LD, including population subdivision, bottle-necks and gene-flow, although these explanations are generally for genome-wide LD levels. Population subdivision and gene flow can increase LD in subpopulations whenever allele frequencies differ among those populations. The decay of LD under recombination alone can be greatly reduced, and if selection maintains differences in allele frequencies at two or more loci among those subpopulations, LD in each subpopulation will persist (Nei and Li 1973; Li and Nei 1974; Slatkin 1975). Changes in population size can also increase LD, so this is especially true for colonizing

species which tend to undergo repeated bottlenecks, resulting in a loss of haplotype diversity, and thus generally resulting in increased LD (Slatkin 2008).

Ultimately, the Bayesian outlier analyses (Figure 3.3) suggests pronounced linkage disequilibrium (LD) across ~17k bp covering both the epsilon and HBB genes on the β-globin cluster for both speckled teal and yellow-billed pintail. Overall, there seems to be lower levels of LD in speckled teal, relative to the yellow-billed pintail, presumably due to longer divergence times between its' respective high- and low-altitude populations; this would longer periods of time available for recombination to disrupt LD patterns, previously established by the founder effect/selective sweep.

β^{ε}: Parallel evolution or hybrid introgression?

Speckled teal and yellow-billed pintail showed the exact same amino acid changes at the same positions in the β^{ε} subunit gene (Table 3.5), suggesting either [1] parallel evolution at these sites or [2] gene-flow between these two high-altitude populations resulting in hybrid introgression. The same pattern of parallelism has been documented in these same species in their adult β^{A} subunit, with $\beta^{A}13$, $\beta^{A}116$, and $\beta^{A}133$ also showing identical amino-acid replacements (McCracken et al. 2009a; McCracken et al. 2009b), with the latter two substitutions contributing to functional change that increases Hb-O_2 affinity (Natarajan et al. 2015). Previous work has established that speckled teal and yellow-billed pintail periodically hybridize (McCracken and Wilson 2011), therefore, it seems likely that similar patterns across this entire gene cluster became established through gene-flow and introgression following initial founder event of one or the other species at high altitude. In the case of the β^{ε} subunit the derived high-

altitude genotypes are fixed in the speckled teal populations, but not in the yellow-billed pintail populations (Table 3.5), although this is not the case for the β^A genotypes (McCracken et al. 2009a; McCracken et al. 2009b). Similar examples of adaptive introgression have been identified across plants and animals (Hedrick 2013; Racimo et al. 2015; Goulet et al. 2017; Medugorac et al. 2017), including variation associated with high-altitude living (Huerta-Sánchez et al. 2014).

Transcription factor binding sites and high-altitude adaptation

Gene expression is largely governed by transcription factors (TFs), proteins that recognize short DNA sequence motifs, whose DNA binding specificities are key components of gene regulatory processes. Genetic variation in TFBS have the potential to cause changes in genetic regulation of downstream targets in a number of ways, including [1] novel DNA-binding specificities that can be modified through protein-protein interactions via cofactors that are changed by modified DNA binding preferences, [2] changes in TF binding architecture that can affect the stability of TF binding behavior, due to changes in secondary binding through protein-protein interactions, [3] base pairs flanking a TFBS can influence TF binding through their effects on DNA shape, and [4] the sequence context may influence TF binding through its effect on nucleosome formation (Levo and Segal 2014).

It is well known that nucleotide variation in regulatory regions is considered an important component for disease risk because variation in binding sites may alter gene expression level and likely contribute to variation in human disease risk (Maurano et al. 2012; Epstein 2009). However, such variation is not always deleterious, instead,

frequently providing an important basis for phenotypic variation leading to heritable regulatory changes (Wray 2007; Rockman and Kruglyak 2006; Stranger et al. 2007; Wittkopp and Kalay 2011). Transcription-factor binding sites show remarkable plasticity and rapid divergence (Borneman et al. 2007; Odom et al. 2007; Tirosh et al. 2008; Tuch et al. 2008; Nadimpalli et al. 2015), therefore promoting evolution of regulatory processes through gene expression changes.

The analyses of the TFBS associated with each of the Hb cluster genes, suggests that there is ample genetic variation available in the upstream regions (Table 3.8), which have the potential to alter TF binding, by any of the mechanisms outlined above. In general, the upstream regulatory regions of the β-globin genes displayed the most genetic variation across all three species. The results also suggested that such variation may be responsible for either changing TFBS identity, or creating new TFBS, as evidenced by a general increase in TFBS matrix predictions represented in high-altitude populations. These same families that were overrepresented in high-altitude populations were largely those whose GO terms involved oxygen, hypoxia and/or blood production (Table 3.6, 3.7). Thus, all together, the results could suggest that genetic variation is altering TFBS identity in these species, in turn changing the binding behavior of the TFs that are directly involved in Hb expression. However, it is unclear whether that change results in increased affinity for TF binding and increased Hb expression, or if that variation results in sub-optimal binding targets, thus changing gene expression. These results are thus rather speculative, but do hint at potentially another avenue for which population

divergence and adaptation may have occurred in these species, other than through substitutions in protein coding regions. The methods used in this study are preliminary and highlights an understudied element of the Hb gene clusters.

Conclusions

High-altitude adaptation in multiple Andean waterfowl species has previously been illustrated by examples in which Hb has evolved increased O_2 affinity by convergent and parallel mechanisms in O_2-poor environments. This study further highlights the importance of genetic variation associated with gene regions that have not been studied before, specifically, embryonic Hb and proximal-promoter regions across both the α- and β-globin gene clusters. Furthermore, it underlines the role of additional avenues for adaptations to O_2-depleted environments, not always through de novo mutations, but potentially through adaptive introgression in the embryonic β^{ε} globin gene, and potentially across the entire β-globin cluster. In the latter case, site-directed mutagenesis could be employed to assay the degree to which Hb-O_2 affinity is affected by the four amino-acid replacements observed in high-altitude populations of speckled teal and yellow-billed pintail. To address the roles of transcriptional regulation through transcription-factor binding sites, Chromatin Immunoprecipitation Sequencing (ChIP-Seq) could be utilized to identify DNA binding sites, and cross-verify the computationally predicted sites, from this study. Exploring these additional avenues has the potential to expand our understanding of the variety of ways organisms can adapt to high-altitude environments.

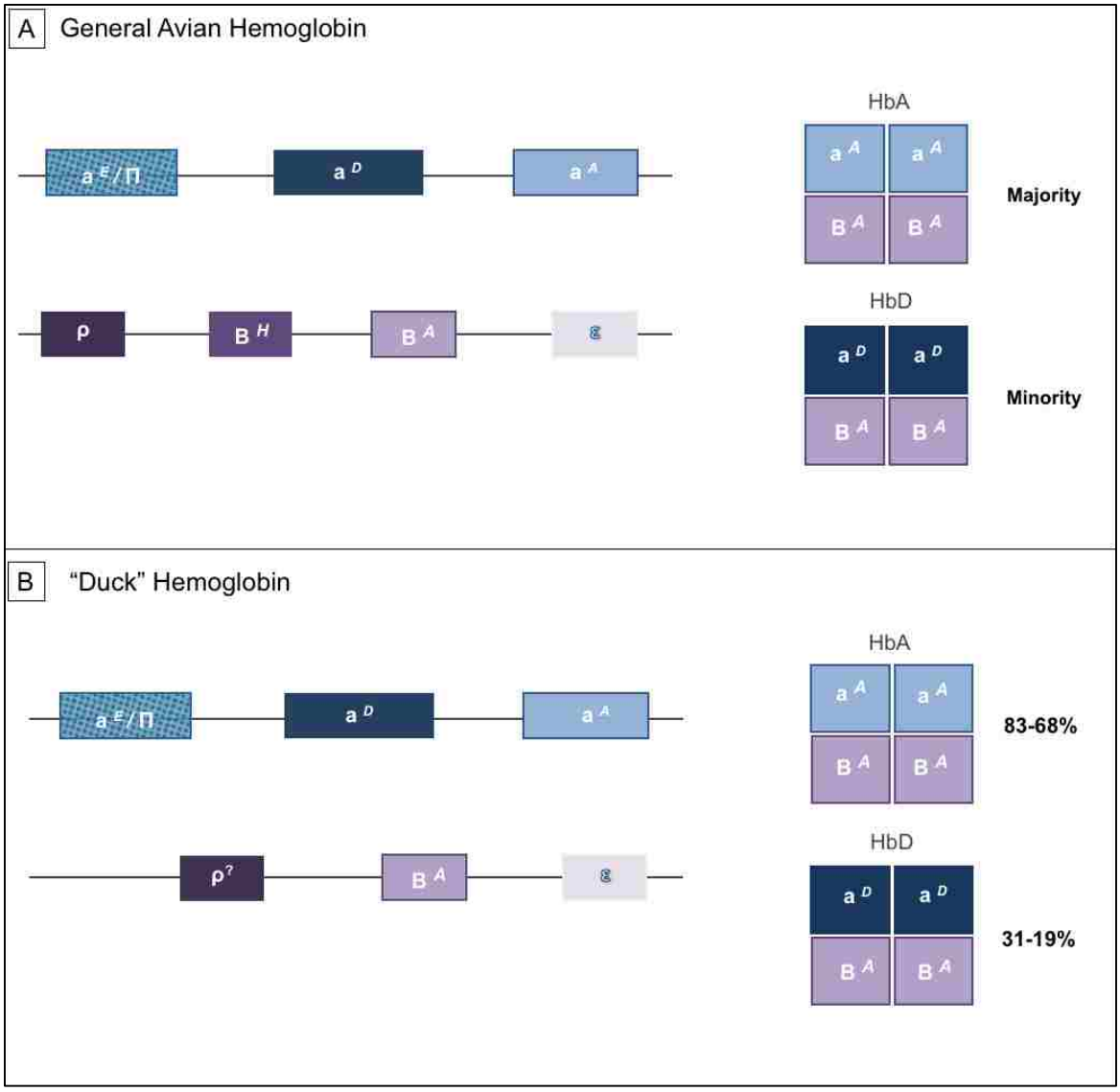

Figure 3.1: (A) General avian model from Opazo et al. (2015) and (B) general duck hemoglobin combined from Opazo et al. (2015) and NCBI annotations and adult isoform Hb percentages from Natarajan et al. (2015).

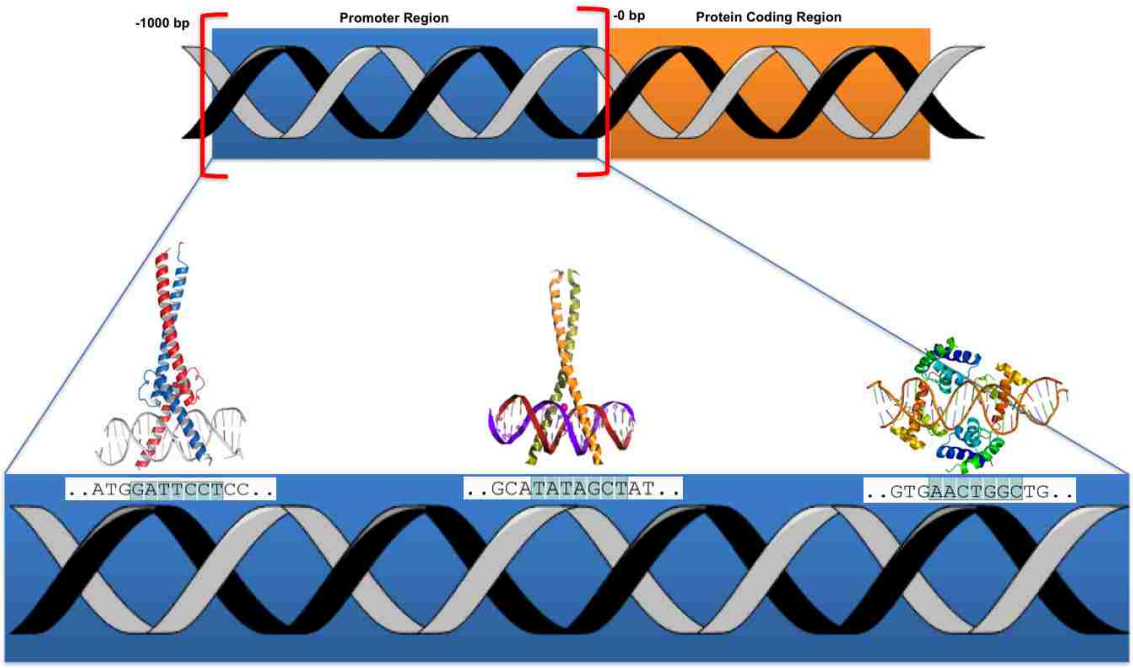

Figure 3.2: Example of a gene and its corresponding upstream promoter region; in this example, the promoter region, there are three TFBS (blue highlighted as anchor position), with their corresponding TF structure (TF structures are examples; image permissions through CC BY-SA 3.0).

Table 3.1: Assembly statistics for the hemoglobin clusters for yellow-billed pintail

		α-Hb				β-Hb			
		Bowtie2		BWA		Bowtie2		BWA	
Specimen ID	Pop	Coverage Avg	Reads	Coverage Avg	Percentage	Coverage Avg	Reads	Coverage Avg	Percentage
REW251	High	1074.4	119,913	570.98	98.0%	313	137,080	315.49	93.1%
REW247	High	989.6	113,542	554.93	98.8%	297	130,837	304.34	92.1%
REW144	High	1106	118,015	541.11	98.0%	298.2	129,687	298.21	92.3%
REW112	High	830.5	98,422	496.61	98.2%	266.5	116,604	266.05	92.4%
KGM1076	High	788.9	88,603	448.27	98.6%	228	100,226	227.99	93.5%
KGM567	High	1224.6	123,848	532.59	98.3%	264.6	116,106	265.53	93.3%
KGM551	High	888.5	98,544	488.74	98.3%	242	106,117	240.98	93.6%
KGM507	High	955.3	96,989	445.14	98.0%	200.7	88,161	202.88	92.6%
KGM491	High	919.1	88,707	408.17	97.9%	177.1	77,724	178.45	93.1%
KGM425	High	687.3	89,692	474.16	98.2%	264.9	116,172	263.60	93.0%
KGM780	Low	1101.5	118,945	550.24	98.2%	293.3	128,486	290.57	93.4%
KGM765	Low	1288	122,901	538.62	98.1%	287.1	125,915	283.43	93.6%
KGM742	Low	1238.4	128,813	578.86	98.6%	319.5	140,047	315.95	93.6%
KGM733	Low	1323.3	136,267	604.25	98.9%	332.9	146,270	330.96	93.5%
KGM731	Low	1308.9	137,473	608.26	98.7%	332.3	145,875	329.50	93.5%
KGM714	Low	828.1	91,113	456.81	98.2%	227.2	99,576	225.61	93.6%
KGM324	Low	947.8	104,139	503.30	98.6%	251.8	110,780	250.39	93.6%
KGM309	Low	1062.2	109,328	509.12	98.4%	266.3	116,975	264.39	93.6%
KGM306	Low	1346.4	131,573	561.76	98.4%	298.7	131,119	294.65	93.6%
KGM274	Low	1029.8	112,381	532.40	98.0%	297.2	129,496	292.79	93.1%
Average	-	1046.93	111460.40	520.22	98.3%	272.92	119662.65	272.09	93.2%

Table 3.2: Assembly statistics for the hemoglobin clusters for speckled teal

		α-Hb				β-Hb			
		Bowtie2		BWA		Bowtie2		BWA	
Specimen ID	Pop	Coverage Avg	Reads	Coverage Avg	Percentage	Coverage Avg	Reads	Coverage Avg	Percentage
KGM1129	High	920.5	96,854	471.30	97.8%	230.1	101,120	232.98	93.3%
KGM437	High	1309.2	126,635	557.32	97.5%	280.5	122,951	284.22	92.4%
KGM449	High	53.8	7,369	63.86	95.2%	14	6,236	14.89	90.3%
KGM484	High	1363.4	128,739	570.60	96.4%	280.9	123,517	284.34	92.4%
KGM502	High	1358.7	132,126	573.55	97.7%	302.6	132,800	306.41	92.5%
KGM543	High	1153	117,212	542.15	98.3%	271.5	118,440	268.86	93.6%
REW092	High	243.6	34,256	228.69	96.0%	101	44,428	102.23	91.9%
REW132	High	684.3	84,093	457.57	96.8%	246.7	108,231	245.26	92.4%
REW219	High	961.4	111,232	559.33	97.2%	304.6	133,426	337.93	92.3%
REW237	High	807	104,068	559.33	97.2%	335.1	147,060	337.93	92.3%
KGM267	Low	10.7	1,731	3.66	78.0%	3	1,321	17.89	80.0%
KGM275	Low	1309	134,142	609.35	98.2%	347.2	151,736	340.57	93.3%
KGM285	Low	770.5	82,524	418.28	97.5%	190.5	83,346	190.81	93.4%
KGM319	Low	1868.6	200,558	862.23	98.8%	532	232,840	517.15	93.6%
KGM699	Low	957.2	107,025	523.37	97.4%	263.6	115,820	265.07	92.6%
KGM727	Low	1167.8	122,249	569.09	97.1%	293	127,763	288.84	93.2%
KGM735	Low	1229.4	121,848	549.40	98.0%	274.5	120,176	273.11	93.5%
KGM747	Low	1435.8	141,313	624.13	98.5%	357.7	156,444	351.68	93.6%
KGM778	Low	1561.9	176,622	812.96	97.3%	476.9	209,529	469.17	93.5%
KGM790	Low	1362.6	134,573	603.16	97.6%	310.3	135,589	307.95	93.6%
Average	-	1026.42	108258.45	507.97	96.4%	270.79	118638.65	271.86	92.2%

Table 3.3: General assembly statistics for the hemoglobin clusters for cinnamon teal

		α-Hb				β-Hb			
		Bowtie2		BWA		Bowtie2		BWA	
Specimen ID	Pop	Coverage Avg	Reads	Coverage Avg	Percentage	Coverage Avg	Reads	Coverage Avg	Percentage
REW286	High	818.1	89,110	476.78	93.5%	223.8	97,307	222.20	92.4%
REW272	High	1248.9	134,983	611.40	97.9%	339.7	148,597	341.27	92.4%
REW269	High	1606.1	169,413	710.07	97.6%	393.5	171,843	395.19	92.3%
REW259	High	855	100,350	508.92	92.6%	264.1	115,581	270.91	90.7%
REW255	High	624.2	78,814	442.55	98.7%	227.8	99,984	233.19	92.0%
REW253	High	1214.2	134,660	612.62	98.0%	351	153,828	356.47	91.2%
REW238	High	607.5	77,548	683.99	98.2%	202.6	88,865	395.22	92.2%
KGM533	High	1019.9	101,750	443.05	98.2%	200.1	87,896	201.16	93.0%
KGM527	High	873.9	90,381	421.81	97.8%	193.5	85,258	195.32	93.2%
KGM486	High	845.1	88,221	437.98	97.9%	187.1	81,870	191.52	91.4%
REW316	Low	939.3	120,269	598.11	98.7%	328.9	144,608	331.10	92.0%
REW305	Low	969.7	113,999	542.12	97.8%	286.5	125,620	285.43	92.4%
REW301	Low	1197.6	132,132	605.98	98.4%	339.2	148,850	339.81	92.3%
REW235	Low	1196	124,235	529.64	98.2%	273.1	117,045	275.16	90.9%
REW207	Low	1006.4	124,588	600.16	98.4%	341.8	149,426	345.97	91.0%
REW206	Low	915.8	115,907	566.16	98.9%	319.1	140,711	325.20	92.1%
REW203	Low	905.6	100,840	482.26	98.7%	255.8	111,943	259.05	92.4%
REW200	Low	1146.3	129,971	604.34	98.4%	325	142,683	327.59	93.5%
REW193	Low	1247	145,855	639.45	98.4%	365.3	160,039	367.16	91.9%
REW081	Low	629.1	79,378	418.05	98.1%	222.2	97,184	224.31	92.1%
Average	-	1002.51	113857.58	550.46	97.9%	285.07	124833.21	297.95	92.1%

Table 3.4: F_{ST} for genes, and their upstream regulatory elements for each of the three species (CT, cinnamon teal; ST, speckled teal; YBP, yellow-billed pintail)

Region	Species	Gene	Gene F_{ST}	Upstream F_{ST}
α	CT	α^A	0.831	0.333
	CT	α^D	0.173	0.162
	CT	π	0.275	0
	ST	α^A	0.532	0.676
	ST	α^D	0.477	0.184
	ST	π	0.084	0.026
	YBP	α^A	-0.049	0
	YBP	α^D	-0.016	0.039
	YBP	π	0.024	0.039
β	CT	β^A	-0.051	-0.027
	CT	β^ε	-0.045	-0.100
	ST	β^A	0.729	0.825
	ST	β^ε	0.749	0.646
	YBP	β^A	0.676	0.941
	YBP	β^ε	0.573	0.608

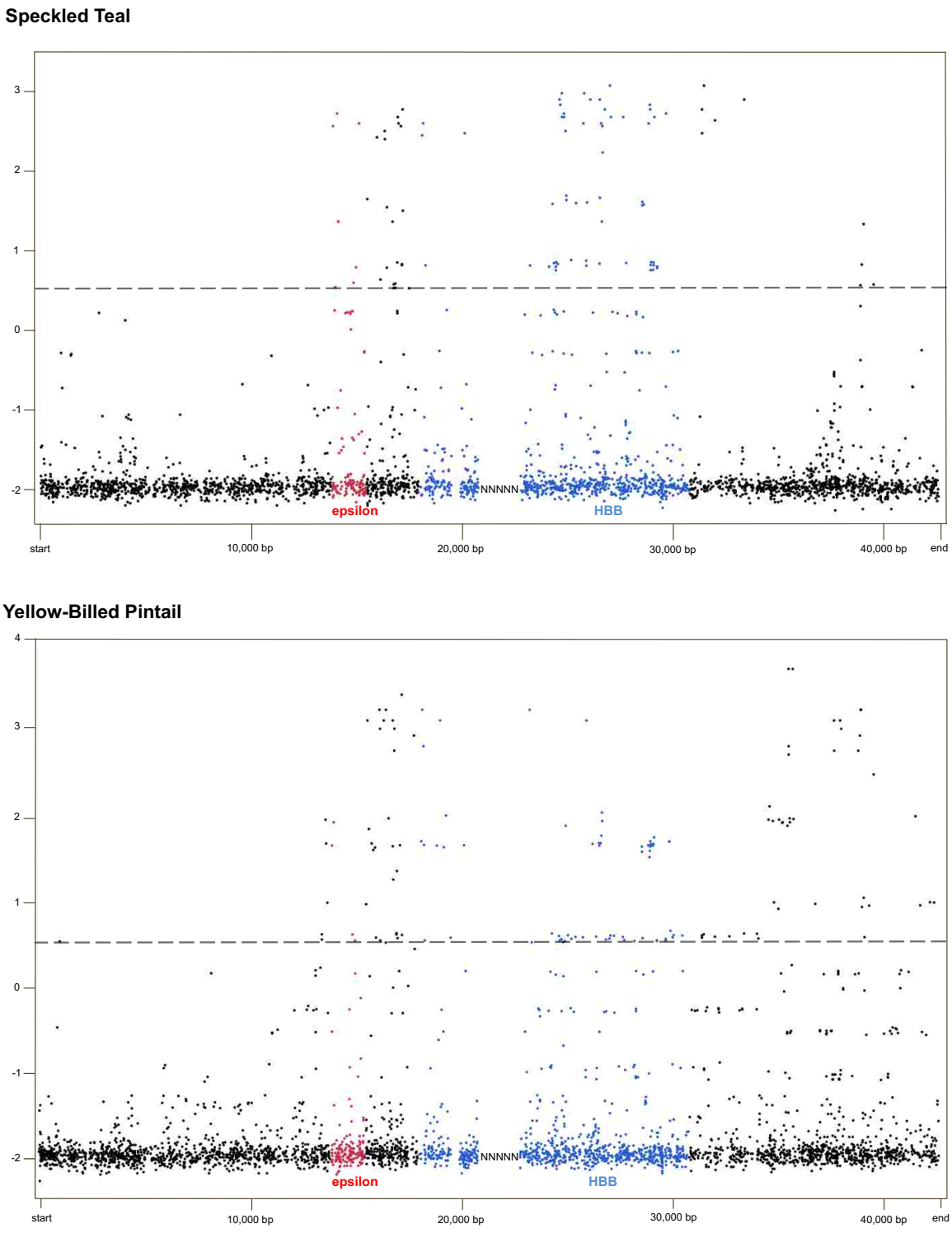

Figure 3.3: Bayesian outlier analysis across the beta-globin cluster for speckled teal (top) and yellow-billed pintail (bottom; both are 5' to 3') – dashed line, Log10(PO) > 0.5 significance threshold; "NNNNN", run of N's in the Mallard scaffold.

Table 3.5: Information about variation in β^{ε}-globin fetal/embryonic gene, including the (1) base-pair position on the full sequence (including introns and exons) (2) base-pair position on the concatenated exon sequence (3) amino-acid position on the polypeptide sequence. In addition, the identities of the amino-acids for each population, as well as the F_{ST} for each species are identified.

Full seq pos.	Coding nuc pos.	aa pos	High aa	Low aa	F_{ST} - ST	F_{ST} - YBP
66	40	14	Gly	Ser	1.00	0.778
69	43	15	Leu	Ile	1.00	0.778
75	49	17	Gly	Ser	1.00	0.733
1468	376	126	Thr	Ala	1.00	0.753

Table 3.6: TFBS matrices (ie. matrix families) whose counts (MWU, $P > 0.05$) were different between high- and low- altitude populations for three genes on the α-globin cluster for each of the three species (CT, cinnamon teal; ST, speckled teal; YBP, yellow-billed pintail). The bolded/starred matrix families are those who have Hb pertinent GO terms.

Upstream gene	Species	Matrix family	Description of matrix family identity
α	CT	V$RXRF	RXR heterodimer binding sites
α	CT	V$AP4R	AP4 and related proteins
α	CT	**V$PERO***	Peroxisome proliferator-activated receptor
α	CT	**V$NR2F***	Nuclear receptor subfamily 2 factors
α^D	CT	**V$AP1R***	MAF and AP1 related factors
α^D	CT	V$NF1F	Nuclear factor 1
α^D	CT	V$ZF57	KRAB domain zinc finger protein 57
α^D	CT	V$KLFS	Krueppel like transcription factors
α^D	CT	**V$HAND***	Twist subfamily of class B bHLH transcription factors
α^D	CT	**V$MYOD***	Myoblast determining factors
α^D	CT	**V$AP2F***	Activator protein 2
α^D	CT	V$PRDF	Positive regulatory domain I binding factor
α^π	CT	**V$AP1F***	AP1, Activating protein 1
α^π	CT	**V$ETSF***	Human and murine ETS1 factors
α^π	CT	**V$NKXH***	NKX homeodomain factors
α^π	CT	**V$SMAD***	Vertebrate SMAD family of transcription factors
α	ST	**V$SORY***	SOX/SRY-sex/testis determinig and related HMG box factors
α	ST	V$INSM	Insulinoma associated factors
α	ST	O$TF3C	RNA polymerase III transcription initiation factor complex (TFIIIC)
α	ST	V$E2FF	E2F-myc activator/cell cycle regulator
α	ST	V$NOLF	Neuron-specific-olfactory factor
α	ST	V$PURA	Pur-alpha binds both single-stranded and douple-stranded DNA in a sequence-specific manner
α	ST	**V$RBPF***	RBPJ - kappa
α	ST	**V$STAT***	Signal transducer and activator of transcription
α^D	ST	**V$EBOX***	E-box binding factors
α^D	ST	V$SF1F	Vertebrate steroidogenic factor
α^D	ST	**V$CP2F***	CP2-erythrocyte Factor related to drosophila Elf1
α^D	ST	V$ZF02	C2H2 zinc finger transcription factors 2
α^D	ST	V$NRSF	Neuron-restrictive silencer factor
α^D	ST	V$DUXF	Double homeobox factors
α^D	ST	**V$EGRF***	EGR/nerve growth factor induced protein C & related factors
α^D	ST	V$CLOX	CLOX and CLOX homology (CDP) factors
α^π	ST	V$MIZ1	Myc-interacting Zn finger protein 1
α^π	ST	**V$ETSF***	Human and murine ETS1 factors
α^π	ST	V$CEBP	Heterodimer of CEBP epsilon and ATF4
α^π	ST	**V$STAT***	Signal transducer and activator of transcription
α^π	ST	V$BARB	Barbiturate-inducible element box from pro+eukaryotic genes
α^π	ST	V$MYT1	MYT1 C2HC zinc finger protein
α^π	ST	V$NEUR	NeuroD, Beta2, HLH domain
α^π	ST	**V$HAND***	Twist subfamily of class B bHLH transcription factors
α^π	ST	V$HNF1	Hepatic Nuclear Factor 1

α^{π}	YBP	V$GCMF	Chorion-specific transcription factors with a GCM DNA binding domain
α^{π}	YBP	V$ZF02	C2H2 zinc finger transcription factors 2
α^{π}	YBP	V$MZF1	Myeloid zinc finger 1 factors
α^{π}	YBP	V$CTCF	CTCF and BORIS gene family, transcriptional regulators with 11 highly conserved zinc finger domains
α^{π}	YBP	V$NOLF	Neuron-specific olfactory factor
α^{π}	YBP	V$MAZF	Myc associated zinc fingers
α^{π}	YBP	**V$ETSF***	Human and murine ETS1 factors
α^{π}	YBP	V$ZBED	Zinc finger BED domain-containing protein
α^{π}	YBP	**V$MYBL***	Cellular and viral myb-like transcriptional regulators
α^{π}	YBP	**V$ETSF***	Human and murine ETS1 factors
α^{π}	YBP	V$MIZ1	Myc-interacting Zn finger protein 1
α^{π}	YBP	**V$HIFF***	Hypoxia inducible factor, bHLH/PAS protein family

Table 3.7: TFBS matrices (ie. matrix families) whose counts (MWU, $P > 0.05$) were different between high- and low- altitude populations for two genes on the β-globin cluster for each of the three species (CT, cinnamon teal; ST, speckled teal; YBP, yellow-billed pintail). The bolded/starred matrix families are those who have Hb pertinent GO terms.

Upstream gene	Species	Matrix family	Description of matrix family identity
β^{ε}	CT	**V$EGRF***	EGR/nerve growth factor induced protein C & related factors
β^{ε}	ST	**V$HBOX***	Homeobox transcription factors
β^{ε}	ST	V$RREB	Ras-responsive element binding protein
β^{ε}	ST	V$CAAT	CCAAT binding factors
β^{ε}	ST	**V$LEFF***	LEF1/TCF
β^{ε}	ST	V$STAF	Selenocysteine tRNA activating factor
β^{ε}	ST	**V$E2FF***	E2F-myc activator/cell cycle regulator
β^{ε}	ST	**V$PAX6***	PAX-4/PAX-6 paired domain binding sites
β^{ε}	ST	**V$EBOX***	E-box binding factors
β^{ε}	ST	**V$NR2F***	Nucleoside diphosphate kinase
β^{ε}	ST	V$AP2F	Activator protein 2
β^{ε}	ST	**V$SORY***	SOX/SRY-sex/testis determinig and related HMG box factors
β^{ε}	ST	V$NDPK	Nucleoside-diphosphate kinases
β^{ε}	ST	**V$KLFS***	Krueppel like transcription factors
β^{ε}	ST	V$PAX5	PAX-2/5/8 binding sites
β^{ε}	ST	V$ZTRE	Zinc transcriptional regulatory element
β^{ε}	ST	V$CTCF	CTCF and BORIS gene family, transcriptional regulators with 11 highly conserved zinc finger domains
β^{ε}	ST	**V$NRSF***	Neuron-restrictive silencer factor
β^{ε}	ST	V$MZF1	Myeloid zinc finger 1 factors
β^{ε}	ST	V$PLAG	Pleomorphic adenoma gene
β^{ε}	ST	**V$ETSF***	Human and murine ETS1 factors
β^{ε}	YBP	V$PAX5	PAX-2/5/8 binding sites
β^{ε}	YBP	V$CAAT	CCAAT binding factors
β^{ε}	YBP	V$NOLF	Neuron-specific olfactory factor
β^{ε}	YBP	V$NFKB	Nuclear factor kappa B/c-rel
β^{ε}	YBP	V$RREB	Ras-responsive element binding protein
β^{ε}	YBP	**V$LEFF***	LEF1/TCF
β^{ε}	YBP	V$ZF02	C2H2 zinc finger transcription factors 2
β^{ε}	YBP	**V$HBOX***	Homeobox transcription factors
β^{ε}	YBP	**V$GLIF***	GLI zinc finger family
β^{ε}	YBP	V$PLAG	Pleomorphic adenoma gene
β^{ε}	YBP	**V$PAX6***	PAX-4/PAX-6 paired domain binding sites
β^{ε}	YBP	**V$KLFS***	Krueppel like transcription factors
β^{ε}	YBP	V$AP2F	Activator protein 2
β^{ε}	YBP	V$MEF3	MEF3 binding sites
β^{ε}	YBP	V$NBRE	NGFI-B response elements, nur subfamily of nuclear receptors
β^{ε}	YBP	V$SP1F	GC-Box factors SP1/GC
β^{ε}	YBP	V$STAF	Selenocysteine tRNA activating factor
β^{ε}	YBP	V$GRHL	Grainyhead-like transcription factors
β^{ε}	YBP	V$BEDF	BED subclass of zinc-finger proteins
β^{ε}	YBP	**V$SORY***	SOX/SRY-sex/testis determinig and related HMG box factors
β^{ε}	YBP	**V$NR2F***	Nuclear receptor subfamily 2 factors
β^{ε}	YBP	**V$EBOX***	E-box binding factors
β^{ε}	YBP	**V$NRSF***	Neuron-restrictive silencer factor
β^{ε}	YBP	V$CTCF	CTCF and BORIS gene family, transcriptional regulators with 11 highly conserved zinc finger domains
β^{ε}	YBP	V$PCBE	REB core-binding element
β^{ε}	YBP	**V$CEBP***	Ccaat/Enhancer Binding Protein

β^{ε}	YBP	V$GABF	GA-boxes
β^{ε}	YBP	V$MIZ1	Myc-interacting Zn finger protein 1
β^{ε}	YBP	V$RU49	Zinc finger transcription factor RU49, zinc finger proliferation 1 - Zipro1
β^{A}	CT	V$BRNF	Brn POU domain factors
β^{A}	CT	**V$LEFF***	LEF1/TCF
β^{A}	CT	**V$NFAT***	Nuclear factor of activated T-cells
β^{A}	CT	V$NGRE	"Negative" glucocoticoid response elements
β^{A}	CT	V$HNF6	Onecut homeodomain factor HNF6
β^{A}	ST	O$INRE	Core promoter initiator elements
β^{A}	ST	**V$SORY***	SOX/SRY-sex/testis determinig and related HMG box factors
β^{A}	ST	**V$KLFS***	Krueppel like transcription factors
β^{A}	ST	**V$RXRF***	RXR heterodimer binding sites
β^{A}	ST	V$RUSH	SWI/SNF related nucleophosphoproteins with a RING finger DNA binding motif
β^{A}	ST	**V$NR2F***	Nuclear receptor subfamily 2 factors
β^{A}	ST	V$NOLF	Neuron-specific olfactory factor
β^{A}	ST	**V$AP1R***	MAF and AP1 related factors
β^{A}	YBP	**V$AP1R***	MAF and AP1 related factors
β^{A}	YBP	**V$GREF***	Glucocorticoid responsive and related elements
β^{A}	YBP	**V$HEAT***	Heat shock factors
β^{A}	YBP	**V$NR2F***	Nuclear receptor subfamily 2 factors
β^{A}	YBP	V$CARE	Calcium-response elements
β^{A}	YBP	O$INRE	Core promoter initiator elements
β^{A}	YBP	**V$NFAT***	Nuclear factor of activated T-cells
β^{A}	YBP	**V$P53F***	p53 tumor suppressor
β^{A}	YBP	**V$SORY***	SOX/SRY-sex/testis determinig and related HMG box factors
β^{A}	YBP	V$OCT1	Octamer binding protein
β^{A}	YBP	**V$GLIF***	GLI zinc finger family
β^{A}	YBP	**V$RXRF***	RXR heterodimer binding sites
β^{A}	YBP	V$PDX1	Pancreatic and intestinal homeodomain transcription factor
β^{A}	YBP	V$EVI1	EVI1-myleoid transforming protein

Table 3.8: Genetic variation associated with upstream regions of each cluster: locus - represents the base-pair location on the designated 1,000 bp area, TF overlap – whether the SNP variant occurs in the start, stop, or anchor position of a predicted TFBS (CT, cinnamon teal; ST, speckled teal; YBP, yellow-billed pintail)

Cluster	Gene	Species	Locus	F_{ST}	TF overlap?
α globin	α^A	ST	34	0.77778	yes
	α^A	ST	107	1	yes
	α^A	ST	471	0.7619	yes
	α^A	CT	572	0.66667	yes
	α^A	ST	585	0.88889	yes
	α^π	ST	375	0.55556	yes
	α^π	YBP	955	0.86111	yes
	α^D	CT	55	0.44444	yes
	α^D	CT	250	0.44444	yes
	α^D	CT	279	0.44444	yes
	α^D	YBP	366	0.44444	yes
	α^D	YBP	432	0.44444	yes
	α^D	ST	542	0.55556	yes
	α^D	ST	560	0.66667	yes
	α^D	ST	564	0.66667	yes
	α^D	CT	732	0.44444	no
	α^D	CT	776	0.44444	yes
	α^D	ST	908	0.39394	yes
β globin	β^A	YBP	43	0.86111	yes
	β^A	YBP	134	1	yes
	β^A	YBP	207	1	yes
	β^A	ST	258	0.66667	yes
	β^A	YBP	289	1	yes
	β^A	YBP	401	1	yes
	β^A	YBP	402	0.86111	yes
	β^A	YBP	406	0.86111	no
	β^A	ST	456	0.66667	yes
	β^A	YBP	456	0.86111	yes
	β^A	ST	517	0.88889	yes
	β^A	YBP	517	0.86111	yes
	β^A	YBP	537	0.86111	yes
	β^A	YBP	539	0.86111	yes
	β^A	ST	614	1	yes
	β^A	YBP	614	1	yes
	β^A	ST	618	1	yes
	β^A	YBP	618	1	yes
	β^A	ST	701	1	yes
	β^A	YBP	701	1	yes
	β^A	YBP	716	1	yes
	β^A	YBP	717	1	yes
	β^A	YBP	804	1	yes
	β^A	YBP	823	0.66667	yes
	β^A	YBP	866	1	yes
	β^A	YBP	926	1	yes
	β^A	YBP	927	1	yes
	β^ε	ST	62	0.77778	yes
	β^ε	YBP	62	0.74125	yes
	β^ε	ST	69	1	yes
	β^ε	YBP	69	0.40171	yes
	β^ε	ST	120	0.88889	yes
	β^ε	YBP	120	0.87654	yes
	β^ε	YBP	189	0.87654	yes
	β^ε	ST	189	0.66667	yes

	β^{ε}	YBP	232	0.72222	yes
	β^{ε}	ST	233	0.51807	yes
	β^{ε}	YBP	284	0.63889	yes
	β^{ε}	ST	285	0.42511	yes
	β^{ε}	YBP	289	0.77778	yes
	β^{ε}	ST	290	0.39394	yes
	β^{ε}	YBP	296	0.75309	yes
	β^{ε}	ST	297	0.39394	yes
	β^{ε}	YBP	320	1	yes
	β^{ε}	ST	321	0.51807	yes
	β^{ε}	YBP	333	0.55556	yes
	β^{ε}	YBP	401	1	no
	β^{ε}	ST	402	0.51807	no
	β^{ε}	YBP	456	0.72222	yes
	β^{ε}	ST	457	0.64072	yes
	β^{ε}	YBP	482	1	no
	β^{ε}	ST	483	0.64072	no
	β^{ε}	YBP	526	1	yes
	β^{ε}	ST	527	1	yes
	β^{ε}	YBP	558	0.55556	no
	β^{ε}	ST	675	0.88166	yes
	β^{ε}	ST	686	0.51807	yes
	β^{ε}	ST	694	1	yes
	β^{ε}	YBP	694	0.88889	yes
	β^{ε}	ST	715	1	yes
	β^{ε}	YBP	715	0.88889	yes
	β^{ε}	ST	730	1	yes
	β^{ε}	YBP	730	0.86111	yes
	β^{ε}	ST	747	0.7351	yes
	β^{ε}	YBP	757	0.44444	yes
	β^{ε}	ST	793	0.88889	yes
	β^{ε}	YBP	793	1	yes
	β^{ε}	ST	806	1	yes
	β^{ε}	YBP	806	0.86111	yes
	β^{ε}	ST	807	0.55556	yes
	β^{ε}	ST	821	0.55556	yes
	β^{ε}	YBP	821	0.52381	yes
	β^{ε}	ST	822	0.44444	yes
	β^{ε}	ST	899	1	yes
	β^{ε}	YBP	899	0.77496	yes
	β^{ε}	ST	911	1	yes
	β^{ε}	YBP	911	0.55939	yes

Table 3.9: Genetic variation associated with upstream regions of the three genes of the α-globin cluster: locus - represents the base-pair location on the designated 1,000 bp area (see Table 8), sequence – full predicted TFBS where the capitalized base-pairs represent the anchor area (CT, cinnamon teal; ST, speckled teal; YBP, yellow-billed pintail)

Species	Locus	Gene	Matrix family	Detailed family information	Start pos.	End pos.	Anchor pos.	Sequence
ST	34	α^A	V$CHRF	Cell cycle regulators: Cell cycle homology element	35	47	41	cagtTTGAagggc
ST	107	α^A	V$GREF	Glucocorticoid responsive and related elements	90	108	99	gcagggcagagTGTTttca
ST	107	α^A	V$SORY	SOX/SRY-sex/testis determinig and related HMG box factors	93	115	104	aAACActctgccctgctactctt
ST	471	α^A	V$HAND	Twist subfamily of class B bHLH transcription factors	456	476	466	gtacacaCAGCagggacaagg
ST	471	α^A	V$ZICF	Members of ZIC-family, zinc finger protein of the cerebellum	460	474	467	acacaCAGCagggac
CT	572	α^A	V$CP2F	CP2-erythrocyte Factor related to drosophila Elf1	556	574	565	agCTGGgttggggttacgg
CT	572	α^A	V$RREB	Ras-responsive element binding protein	562	576	569	cCCCAacccagctct
CT	572	α^A	V$AP4R	AP4 and related proteins	565	581	573	caaccCAGCtcttgtca
ST	585	α^A	V$CSEN	Calsenilin, presenilin binding protein, EF hand transcription factor	576	586	581	ttGTCAgggtt
ST	585	α^A	O$TF3C	General transcription factor IIIC, GTF3C	583	593	588	GGTTggaaccc
CT	55	α^D	V$SORY	SOX/SRY-sex/testis determinig and related HMG box factors	35	57	46	aagaAATTctccaaaattctcca
CT	55	α^D	V$HEAT	Heat shock factors	35	59	47	agaagaaattctccaaaaTTCTcca
CT	55	α^D	V$SORY	SOX/SRY-sex/testis determinig and related HMG box factors	36	58	47	ggagAATTttggagaatttcttc
CT	55	α^D	V$HEAT	Heat shock factors	36	60	48	ggagaattttggAGAAtttcttctc
CT	55	α^D	V$STAT	Signal transducer and activator of transcription	50	68	59	cactTTCTgagaagaaatt
CT	55	α^D	V$STAT	Signal transducer and activator of transcription	52	70	61	tttcttctcAGAAagtgta
CT	250	α^D	V$DUXF	Double homeobox factors	249	263	256	tatGATTccatgact
CT	250	α^D	V$PAX3	PAX-3 binding sites	247	265	256	aaagtCATGgaatcataga
CT	250	α^D	V$CLOX	CLOX and CLOX homology (CDP) factors	249	271	260	taTGATtccatgactttggaaag
CT	279	α^D	V$PRDF	Positive regulatory domain I binding factor	278	296	287	taccaatGAAAgaggttct
CT	279	α^D	V$SORY	SOX/SRY-sex/testis determinig and related HMG box factors	278	300	289	tacCAATgaaagaggttcttccc
CT	279	α^D	V$P53F	p53 tumor suppressor	347	371	359	aggtcCATGgcatgacctgctaggg
CT	279	α^D	V$RORA	v-ERB and RAR-related orphan receptor alpha	348	372	360	gccctagcaGGTCatgccatggacc
YBP	366	α^D	V$RXRF	RXR heterodimer binding sites	354	378	366	tggcatGACCtgctagggcttgccc
YBP	432	α^D	V$AP1R	MAF and AP1 related factors	422	444	433	cagaaagcaaaggcAGCAcgggg
YBP	432	α^D	V$EBOX	E-box binding factors	432	448	440	tgctcccCGTGctgcct
YBP	432	α^D	V$NRSF	Neuron-restrictive silencer factor	425	455	440	aaagcaaaggcaGCACggggagcagaggtgg
ST	542	α^D	V$BRAC	Brachyury gene, mesoderm developmental factor	520	548	534	acttgtgtcagggcagatcccACACctgc
ST	542	α^D	V$CSEN	Calsenilin, presenilin binding protein, EF hand transcription factor	534	544	539	gtGTCAgggca
ST	542	α^D	V$RORA	v-ERB and RAR-related orphan receptor alpha	528	552	540	tgtcacttgtGTCAgggcagatccc
ST	542	α^D	V$EBOX	E-box binding factors	538	554	546	tttgTCACttgtgtcag
ST	560	α^D	V$SORY	SOX/SRY-sex/testis determinig and related HMG box factors	545	567	556	aagtgACAAagaaccaagcaggg
ST	560	α^D	V$CAAT	CCAAT binding factors	554	568	561	agaaCCAAgcagggc

ST	560	α^D	V$MOKF	Mouse Krueppel like factor	555	575	565	gaaccaagcagggCCTTggca
ST	564	α^D	V$SF1F	Vertebrate steroidogenic factor	562	576	569	ttgcCAAGgccctgc
CT	776	α^D	V$AP4R	AP4 and related proteins	775	791	783	ccttccaGCTGagctgg
CT	776	α^D	V$ZICF	Members of ZIC-family, zinc finger protein of the cerebellum	776	790	783	cagctCAGCtggaag
ST	908	α^D	V$MOKF	Mouse Krueppel like factor	889	909	899	ggctgcagcagttCCTTgggg
ST	908	α^D	V$ZTRE	Zinc transcriptional regulatory element	907	923	915	cgggaccaGGGAgaggc
ST	375	α^π	V$STAT	Signal transducer and activator of transcription	361	379	370	gggcTTCCtaagacagtta
ST	375	α^π	V$MIZ1	Myc-interacting Zn finger protein 1	373	383	378	gaagcCCTCtg
ST	375	α^π	V$AP1R	MAF and AP1 related factors	375	397	386	agccctctggtcTCAGcactttg
ST	375	α^π	V$AP1R	MAF and AP1 related factors	375	397	386	caaagtGCTGagaccagagggct
YBP	955	α^π	V$ZBED	Zinc finger BED domain-containing protein	949	961	955	aTGTCtggacaca
YBP	955	α^π	V$ZBED	Zinc finger BED domain-containing protein	950	962	956	gTGTCcagacatt
YBP	955	α^π	V$SMAD	Vertebrate SMAD family of transcription factors	952	962	957	aatGTCTggac

Table 3.10: Genetic variation associated with upstream regions of the three genes of the β-globin cluster: locus - represents the base-pair location on the designated 1,000 bp area (see Table 8), sequence – full predicted TFBS where the capitalized base-pairs represent the anchor area (CT, cinnamon teal; ST, speckled teal; YBP, yellow-billed pintail)

Species	Locus	Gene	Matrix family	Detailed family information	Start pos	End pos	Anchor pos	Sequence
YBP	43	β^A	V$WHNF	Winged helix binding sites	42	52	47	gggACGCtgca
YBP	43	β^A	V$NOLF	Neuron-specific olfactory factor	43	65	54	gcagcgTCCCaggaggcactttt
YBP	43	β^A	V$NGRE	"Negative" glucocoticoid response elements	45	59	52	gcCTCCtgggacgct
YBP	134	β^A	V$STAF	Selenocysteine tRNA activating factor	108	138	123	cctatgggccttgCCCAggattcagaaatta
YBP	134	β^A	V$MOKF	Mouse Krueppel like factor	123	143	133	aggggcctatgggCCTTgccc
YBP	207	β^A	V$SORY	SOX/SRY-sex/testis determinig and related HMG box factors	204	226	215	agtgcagataAATGtatcataca
YBP	207	β^A	V$GATA	GATA binding factors	206	218	212	tgcaGATAaatgt
ST	258	β^A	V$RXRF	RXR heterodimer binding sites	240	264	252	gttctttaGGTCagaaactacagca
ST	258	β^A	V$RORA	v-ERB and RAR-related orphan receptor alpha	241	265	253	tgttctttaGGTCagaaactacagc
ST	258	β^A	V$CREB	cAMP-responsive element binding proteins	246	266	256	agtttcTGACctaaagaacag
ST	258	β^A	V$RXRF	RXR heterodimer binding sites	250	274	262	gactctgtctgttctttAGGTcaga
ST	258	β^A	V$GREF	Glucocorticoid responsive and related elements	258	276	267	gtgactctgtctGTTCttt
YBP	289	β^A	V$PAX3	PAX-3 binding sites	288	306	297	gTCGTcagatctcctcacg
YBP	401/ 402	β^A	V$MITF	Microphthalmia transcription factor	388	402	395	cacatCATGtgagtc
YBP/ST	456	β^A	V$AP4R	AP4 and related proteins	442	458	450	tcacccaGCTGcacgta
YBP/ST	456	β^A	V$MYOD	Myoblast determining factors	442	458	450	tacgtgCAGCtgggtga
YBP/ST	456	β^A	V$NEUR	NeuroD, Beta2, HLH domain	443	457	450	cacccaGCTGcacgt
YBP/ST	456	β^A	V$MYOD	Myoblast determining factors	443	459	451	ctcaccCAGCtgcacgt
YBP/ST	456	β^A	V$YY1F	Activator/repressor binding to transcription initiation site	455	477	466	gctaggCCATgtattgtcctcac
YBP/ST	456	β^A	V$SORY	SOX/SRY-sex/testis determinig and related HMG box factors	456	478	467	tgaggACAAtacatggcctagca
YBP/ST	517	β^A	V$SNAP	snRNA-activating protein complex	502	520	511	ctcatCCTAtagcctacat
YBP/ST	517	β^A	V$RUSH	SWI/SNF related nucleophosphoproteins with a RING finger DNA binding motif	515	525	520	ctACATttggt
YBP	573/ 539	β^A	V$NFAT	Nuclear factor of activated T-cells	520	538	529	tttggtGGAAaacccacgt
YBP	573/ 539	β^A	V$NFAT	Nuclear factor of activated T-cells	522	540	531	tggtGGAAaacccacgttg
YBP	573/ 539	β^A	V$NFKB	Nuclear factor kappa B/c-rel	523	537	530	cgtgggttTTCCacc
YBP	573/ 539	β^A	V$GLIF	GLI zinc finger family	525	541	533	tggaaaaccCACGttga
YBP	539	β^A	V$GATA	GATA binding factors	539	551	545	tgaaGATAaaatc
YBP/ST	614/ 618	β^A	V$IRFF	Interferon regulatory factors	600	624	612	acagagctGAAAatgagactcccat
YBP/ST	701	β^A	V$AP1R	MAF and AP1 related factors	679	701	690	ccctgagctttGTCAttctgctg
YBP/ST	701	β^A	V$PERO	Peroxisome proliferator-activated receptor	679	701	690	cagcagaatgacAAAGctcaggg
YBP/ST	701	β^A	V$NR2F	Nuclear receptor subfamily 2 factors	680	704	692	agcagaatgaCAAAgctcagggagg
YBP/ST	701	β^A	V$SORY	SOX/SRY-sex/testis determinig and related HMG box factors	684	706	695	gaatgACAAagctcagggagggc
YBP/ST	701	β^A	V$ZTRE	Zinc transcriptional regulatory element	689	705	697	ccCTCCctgagctttgt
YBP/ST	701	β^A	V$KLFS	Krueppel like transcription factors	694	712	703	gctcagGGAGggctggcag
YBP/ST	701	β^A	V$ZTRE	Zinc transcriptional regulatory element	697	713	705	cagGGAGggctggcaga
YBP/ST	701	β^A	V$MIZ1	Myc-interacting Zn finger protein 1	699	709	704	ccagcCCTCcc
YBP	716/ 717	β^A	V$NF1F	Nuclear factor 1	703	723	713	gggcTGGCagataggcatgtg
YBP	716/ 717	β^A	V$GATA	GATA binding factors	708	720	714	ggcaGATAggcat

YBP	716/717	β^A	V$STEM	Motif composed of binding sites for pluripotency or stem cell factors	710	728	719	cagatagGCATgtgaagct
YBP	716/717	β^A	V$MITF	Microphthalmia transcription factor	713	727	720	ataggCATGtgaagc
YBP	716/717	β^A	V$STAF	Selenocysteine tRNA activating factor	714	744	729	ttgttctgcttcTCCCagcttcacatgccta
YBP	804	β^A	V$HAML	Human acute myelogenous leukemia factors	801	815	808	atctGTGGtttaaaa
YBP	804	β^A	O$VTBP	Vertebrate TATA binding protein factor	801	817	809	attttTAAAccacagat
YBP	823	β^A	V$AP1R	MAF and AP1 related factors	808	830	819	acaaaaaatcAGCAtttttaaac
YBP	823	β^A	V$AP1R	MAF and AP1 related factors	812	834	823	aaccacaaaaaaTCAGcatttttt
YBP	823	β^A	V$HNF6	Onecut homeodomain factor HNF6	814	830	822	acaaaaAATCagcattt
YBP	823	β^A	V$RXRF	RXR heterodimer binding sites	815	839	827	aatgctgatttTTTGtggtttaaaa
YBP	866	β^A	V$NFAT	Nuclear factor of activated T-cells	850	868	859	gactccATGGaaaagcagg
YBP	866	β^A	V$NFAT	Nuclear factor of activated T-cells	854	872	863	ccatGGAAaagcaggcaga
YBP	866	β^A	V$BARB	Barbiturate-inducible element box from pro+eukaryotic genes	857	871	864	tggaAAAGcaggcag
YBP	926/927	β^A	V$FKHD	Fork head domain factors	916	932	924	cactcaaAACAgtccat
YBP	926/927	β^A	V$NKXH	NKX homeodomain factors	922	940	931	tgtttTGAGtgttttttga
YBP	926/927	β^A	V$FKHD	Fork head domain factors	925	941	933	ctcaaaaAACActcaaa
YBP/ST	62/69	β^E	V$HDBP	Huntington's disease gene regulatory region binding proteins	55	73	64	ccgctgCCGGcagcccctc
YBP/ST	62/69	β^E	V$ZTRE	Zinc transcriptional regulatory element	61	77	69	cggGGAGgggctgccgg
YBP/ST	62/69	β^E	V$EGRF	EGR/nerve growth factor induced protein C & related factors	61	79	70	ggcgGGGAggggctgccgg
YBP/ST	62/69	β^E	V$NDPK	Nucleoside diphosphate kinase	62	78	70	gcGGGGaggggctgccg
YBP/ST	62/69	β^E	V$KLFS	Krueppel like transcription factors	62	80	71	gggcggGGAGgggctgccg
YBP/ST	120	β^E	V$ZF02	C2H2 zinc finger transcription factors 2	98	120	109	cggacCCCCacccatccccagcc
YBP/ST	120	β^E	V$KLFS	Krueppel like transcription factors	103	121	112	gggatGGGTgggggtccgc
YBP/ST	120	β^E	V$GLIF	GLI zinc finger family	104	120	112	cggaCCCCcacccatcc
YBP/ST	189	β^E	V$DEAF	Homolog to deformed epidermal autoregulatory factor-1 from D. melanogaster	172	190	181	gccTCGGgttatccccagt
YBP/ST	189	β^E	V$NFKB	Nuclear factor kappa B/c-rel	174	188	181	tgGGGAtaacccgag
YBP/ST	189	β^E	V$GATA	GATA binding factors	176	188	182	tgggGATAacccg
YBP/ST	189	β^E	V$NOLF	Neuron-specific olfactory factor	177	199	188	gggttaTCCCcagtgcaagggct
YBP/ST	189	β^E	V$SORY	SOX/SRY-sex/testis determinig and related HMG box factors	186	208	197	aagggACAAagcccttgcactgg
YBP/ST	189	β^E	V$SF1F	Vertebrate steroidogenic factor	188	202	195	agtgCAAGggctttg
YBP/ST	189	β^E	V$NR2F	Nuclear receptor subfamily 2 factors	188	212	200	ggggaagggaCAAAgcccttgcact
YBP/ST	233/234	β^E	V$RORA	v-ERB and RAR-related orphan receptor alpha	217	241	229	ccgctgtcaGGTCactggttctcat
YBP/ST	233/234	β^E	V$GRHL	Grainyhead-like transcription factors	227	239	233	gtcactGGTTctc
YBP/ST	233/234	β^E	V$EVI1	EVI1-myleoid transforming protein	233	249	241	gatctggGATGagaacc
YBP/ST	284/285/289	β^E	V$HOMF	Homeodomain transcription factors	271	289	280	ggctgcacAAGTcgctgcg
YBP/ST	320/321	β^E	V$FKHD	Fork head domain factors	304	320	312	catttaaAACAcgtgca
YBP/ST	320/321	β^E	V$EBOX	E-box binding factors	307	323	315	ccatgCACGtgttttaa
YBP/ST	320/321	β^E	V$HIFF	Hypoxia inducible factor, bHLH/PAS protein family	307	323	315	ttaaaacACGTgcatgg
YBP/ST	320/321	β^E	V$HESF	Vertebrate homologues of enhancer of split complex	308	322	315	catgcaCGTGtttta
YBP/ST	320/321	β^E	V$EBOX	E-box binding factors	308	324	316	taaaaCACGtgcatggg
YBP/ST	320/321/333	β^E	V$HIFF	Hypoxia inducible factor, bHLH/PAS protein family	308	324	316	cccatgcaCGTGtttta
YBP/ST	320/321	β^E	V$HESF	Vertebrate homologues of enhancer of split complex	309	323	316	aaaacaCGTGcatgg
YBP/ST	320/321/333	β^E	V$SORY	SOX/SRY-sex/testis determinig and related HMG box factors	315	337	326	cgtgcatgggacTCATggggcac
YBP	333	β^E	V$HICF	Krueppel-like C2H2 zinc finger factors hypermethylated in cancer	327	339	333	tcgTGCCccatga
YBP	333	β^E	V$MOKF	Mouse Krueppel like factor	328	348	338	catggggcacgagCCTTgtgt

YBP	333	β$^{\varepsilon}$	V$HESF	Vertebrate homologues of enhancer of split complex	331	345	338	ggggCACGagccttg
YBP/ST	456	β$^{\varepsilon}$	V$ZF02	C2H2 zinc finger transcription factors 2	437	459	448	agcagccCCCCagcaggagtagg
YBP/ST	456	β$^{\varepsilon}$	V$HAND	Twist subfamily of class B bHLH transcription factors	440	460	450	agcccccCAGCaggagtaggg
YBP/ST	456	β$^{\varepsilon}$	V$ZICF	Members of ZIC-family, zinc finger protein of the cerebellum	442	456	449	cccccCAGCaggagt
YBP/ST	456	β$^{\varepsilon}$	V$GCMF	Chorion-specific transcription factors with a GCM DNA binding domain	452	466	459	aggcaCCCCtactcc
YBP/ST	456	β$^{\varepsilon}$	V$PLAG	Pleomorphic adenoma gene	454	476	465	agtaGGGGtgcctgcaccgcagg
YBP/ST	526/ 527	β$^{\varepsilon}$	V$NRSF	Neuron-restrictive silencer factor	499	529	514	aggagtgctggagctgctCGGAgggctcgg g
YBP/ST	526/ 527	β$^{\varepsilon}$	V$ZF02	C2H2 zinc finger transcription factors 2	515	537	526	cagcaCCCCccgagccctccgag
YBP/ST	526/ 527	β$^{\varepsilon}$	V$GCF2	Transcriptional repressor GC-binding factor 2	516	534	525	cacCCCCcgagccctccga
YBP/ST	526/ 527	β$^{\varepsilon}$	V$MIZ1	Myc-interacting Zn finger protein 1	517	527	522	cgagcCCTCcg
YBP/ST	526/ 527	β$^{\varepsilon}$	V$ZF02	C2H2 zinc finger transcription factors 2	518	540	529	ccccagcaCCCCccgagccctcc
YBP/ST	526/ 527	β$^{\varepsilon}$	V$KLFS	Krueppel like transcription factors	519	537	528	gagggctcgggGGGTgctg
YBP/ST	526/ 527	β$^{\varepsilon}$	V$GLIF	GLI zinc finger family	520	536	528	agcaCCCCccgagccct
YBP/ST	526/ 527	β$^{\varepsilon}$	V$BEDF	BED subclass of zinc-finger proteins	521	535	528	gggctcgGGGGgtgc
YBP/ST	526/ 527	β$^{\varepsilon}$	V$STAF	Selenocysteine tRNA activating factor	522	552	537	tcagctgcaaagcCCCAgcacccccccgag cc
YBP/ST	526/ 527	β$^{\varepsilon}$	V$GLIF	GLI zinc finger family	523	539	531	cccagcaCCCCccgagc
YBP/ST	526/ 527	β$^{\varepsilon}$	V$SP1F	GC-Box factors SP1/GC	523	539	531	gctcGGGGggtgctggg
YBP/ST	526/ 527	β$^{\varepsilon}$	V$ZF02	C2H2 zinc finger transcription factors 2	523	545	534	caaagcCCCAgcacccccccgagc
YBP/ST	526/ 527	β$^{\varepsilon}$	V$PLAG	Pleomorphic adenoma gene	525	547	536	tcggGGGGtgctggggctttgca
YBP/ST	526/ 527	β$^{\varepsilon}$	V$RREB	Ras-responsive element binding protein	526	540	533	cCCCAgcacccccg
ST	675/ 686	β$^{\varepsilon}$	V$HAND	Twist subfamily of class B bHLH transcription factors	662	682	672	cacatagCACCtggctggtca
ST	675/ 686	β$^{\varepsilon}$	V$HAND	Twist subfamily of class B bHLH transcription factors	663	683	673	gaccagcCAGGtgctatgtgg
ST	675/ 686	β$^{\varepsilon}$	V$CTCF	CTCF and BORIS gene family, transcriptional regulators with 11 highly conserved zinc finger domains	665	691	678	ccagccaggtgctatgtGGCAgtgcag
ST	694	β$^{\varepsilon}$	V$PAX2	PAX-2 binding sites	687	709	698	tgcagggagcagagatcAAACag
YBP/ST	715	β$^{\varepsilon}$	V$PAX3	PAX-3 binding sites	711	729	720	aTCGTcattttcagctgag
YBP/ST	715	β$^{\varepsilon}$	V$TCFF	TCF11 transcription factor	714	720	717	GTCAttt
ST	747	β$^{\varepsilon}$	V$CDXF	Vertebrate caudal related homeodomain protein	743	761	752	cctgaactTTATgggtgag
YBP	757	β$^{\varepsilon}$	V$MYT1	MYT1 C2HC zinc finger protein	750	762	756	ataAAGTtcaggt
YBP	757	β$^{\varepsilon}$	V$MEF3	MEF3 binding sites	754	766	760	agtTCAGgtgaaa
YBP	757	β$^{\varepsilon}$	V$ZFHX	Two-handed zinc finger homeodomain transcription factors	755	767	761	atttcACCTgaac
YBP/ST	793	β$^{\varepsilon}$	V$NF1F	Nuclear factor 1	775	795	785	gaacaccaggtttGCCAacaa
YBP/ST	793	β$^{\varepsilon}$	V$LEFF	LEF1/TCF	786	802	794	ttgccaaCAAAggatct
YBP/ST	793	β$^{\varepsilon}$	V$KLFS	Krueppel like transcription factors	786	804	795	ttgccaacAAAGgatctgc
YBP/ST	793	β$^{\varepsilon}$	V$SORY	SOX/SRY-sex/testis determinig and related HMG box factors	787	809	798	tgccaACAAaggatctgctgcag
YBP/ST	806/ 807	β$^{\varepsilon}$	V$MOKF	Mouse Krueppel like factor	791	811	801	gcctgcagcagatCCTTtgtt
YBP/ST	821/ 822	β$^{\varepsilon}$	V$RXRF	RXR heterodimer binding sites	813	837	825	tgagctgaggtTTTGccgccttta
YBP/ST	821/ 822	β$^{\varepsilon}$	V$E2FF	E2F-myc activator/cell cycle regulator	814	830	822	aaaaGGCGgcaaaacct
YBP/ST	821/ 822	β$^{\varepsilon}$	V$ZF57	KRAB domain zinc finger protein 57	815	827	821	tttTGCCgccttt
YBP/ST	899	β$^{\varepsilon}$	V$BRAC	Brachyury gene, mesoderm developmental factor	877	905	891	tgatgaggctggTGTGacccccctgagca
YBP/ST	899	β$^{\varepsilon}$	V$PLAG	Pleomorphic adenoma gene	879	901	890	caGGGGggtcacaccagcctcat
YBP/ST	899	β$^{\varepsilon}$	V$ESRR	Estrogen-related receptors	887	909	898	cttgtgctcaggggGGTCacacc
YBP/ST	899	β$^{\varepsilon}$	V$RXRF	RXR heterodimer binding sites	888	912	900	ccccttgtgctcaggggGGTCacac
YBP/ST	899	β$^{\varepsilon}$	V$ZF02	C2H2 zinc finger transcription factors 2	889	911	900	tgtgacCCCCctgagcacaaggg
YBP/ST	899	β$^{\varepsilon}$	V$GLIF	GLI zinc finger family	890	906	898	gtgaCCCCcctgagcac

YBP/ST	911	β^{ε}	V$SORY	SOX/SRY-sex/testis determinig and related HMG box factors	900	922	911	tgagcACAAggggtgatgcaccc
YBP/ST	911	β^{ε}	V$GLIF	GLI zinc finger family	903	919	911	tgcatcaCCCCttgtgc

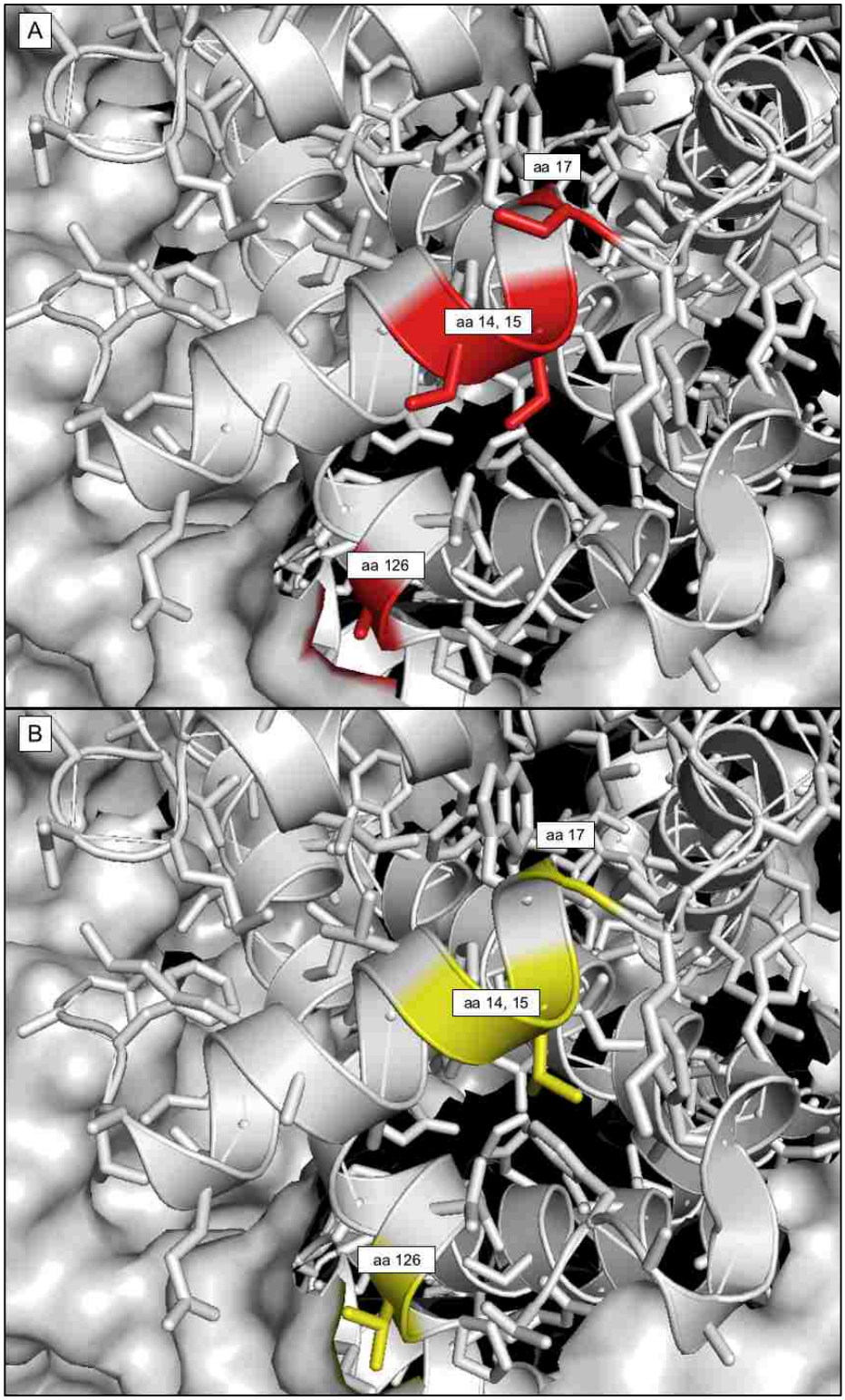

Figure 3.4: Location of the four nonsynonymous substitutions in fetal/embryonic Hb and their corresponding amino-acid identities on the human embryonic Hb protein structure (PDB 1a9w) for [A] low-altitude population (β^{ε}14Ser, β^{ε}15Ile, β^{ε}17Ser, and β^{ε}126Ala) [B] high-altitude population (β^{ε}14Gly, β^{ε}15Leu, β^{ε}17Gly, and β^{ε}126Thr).

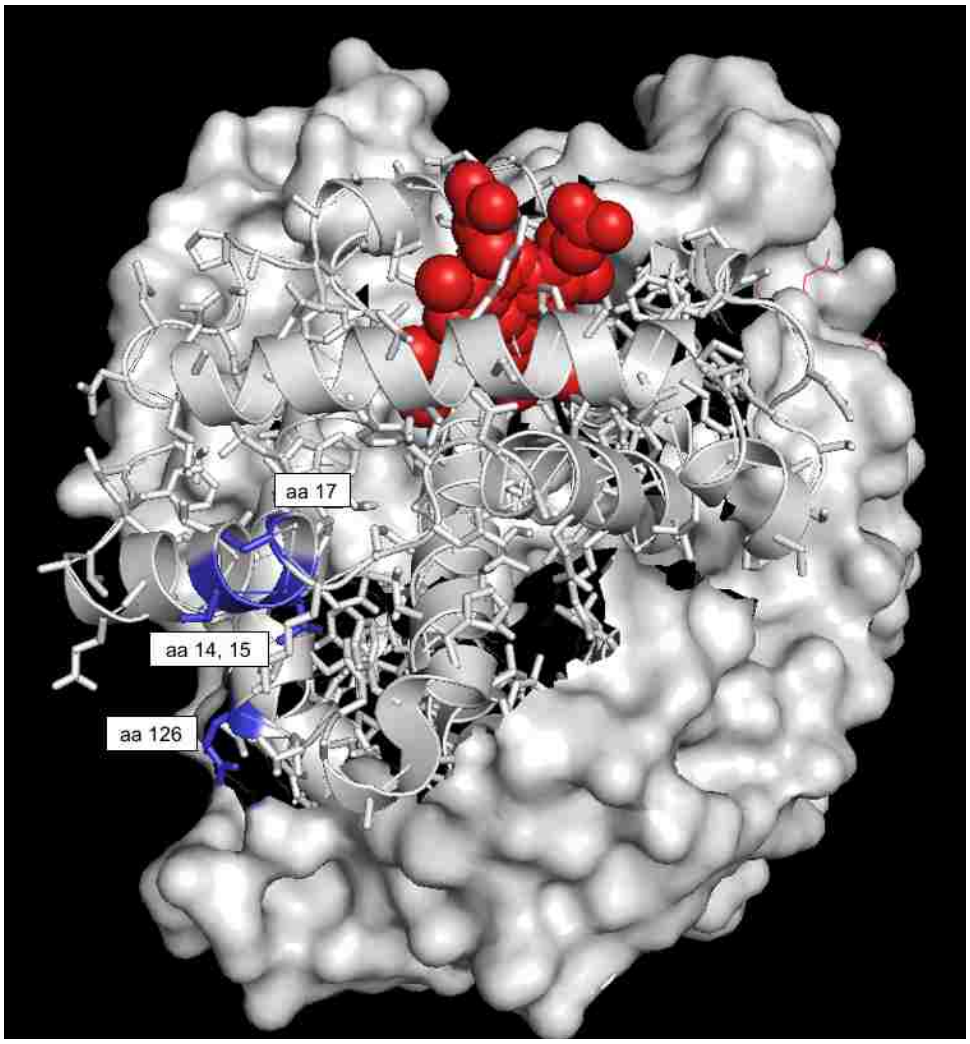

Figure 3.5: General location of the variants (blue) on one epsilon monomer (exposed) in relation to where heme binds O_2 (red).

Chapter 4: Hypoxia-inducible factor (HIF) pathway genes *EGLN1* and *EPAS1* evolved in parallel in high-altitude ducks and convergently between ducks and humans

Background

Convergent evolution of adaptive traits in distantly-related organisms inhabiting similar environments with the same selective pressures is common, both at the genomic and pathway levels (Losos 2011; Stern and Orgogozo 2009; Gompel and Prud'homme 2009; Storz 2016; Conte et al. 2012). Hypoxia (or low O_2) is one selective pressure that stimulates a similar physiological response across metazoans allowing organisms to match O_2 supply and demand (Semenza 2007b). During reduced O_2 supply, changes in gene expression are mediated by transcription factors known as Hypoxia-Inducible Factors (HIF), which have been shown to play a key role in cellular responses to low O_2 tension in a variety of organisms (Lisy and Peet 2008; Semenza 2007a; Webb et al. 2009). Specifically, exposure to low O_2 triggers a hypoxic response pathway regulated by the activity of the hypoxia-inducible transcription factors (HIFs). Gene members of this transcription factor family include *HIF1α*, *HIF2α* (also *EPAS1*), *HIF3α*, and the *ARNT* (i.e., HIF1β) gene family. Ultimately, the response of these transcription factors to hypoxia is the transcriptional regulation of gene expression that is mediated by HIF1A and HIF2A proteins, with the HIF1A heterodimer in particular considered the "master-regulator" of oxygen homeostasis (Semenza 2007a; Wang et al. 1995).

All HIFα proteins are characterized by the presence of an N-terminal bHLH DNA binding domain just upstream of two PAS domains. In addition, all vertebrate α-

subunits include an inhibitory domain called the oxygen-dependent degradation domain (ODDD), and an N-terminal transactivation domain (NTAD). HIF1A and EPAS1 are characterized by the presence of a C-terminal transactivation domain (CTAD) located at the C-terminal end of the protein; although neither HIF3A nor any of the β-subunits contain the CTAD (Lisy and Peet 2008). These domains are considered critical to the overall function of HIF proteins: the bHLH domain contacts the core nucleotides of HIF-responsive elements (Dinkel et al.), while bHLH and PAS domains together mediate both dimerization and sequence specific DNA binding (Ledent and Vervoort 2001; Crews and Fan 1999). The NTAD is thought to confer target specificity (Hu et al. 2007), while the CTAD is required for full HIF activity (Lando et al. 2002) and interactions with co-activators (Ema et al. 1999; Carrero et al. 2000).

In mammalian cells, low O_2 triggers a response regulated by HIF transcription activity, which is regulated post-translationally through degradation by ubiquitination under normoxic conditions. HIFA members are hydroxylated on specific prolyl residues near the NTAD (Pro405, Pro531), via prolyl hydroxylase domain (PHD) family (including EGLN1); this serves as a signal allowing for recognition by a multi-component ubiquitian ligase, comprised of the von Hippel-Lindau (VHL) tumor suppressor protein, elongin B/C and Cul2, which flags it further for proteasomal degradation (Kaelin and Ratcliffe 2008) (Figure 4.1). At the same time, FIH (factor-inhibiting HIF) hydoxylates a conserved asparaginyl residue (Asn847)within the HIF CTAD preventing the recruitment of coactivators (Kaelin Jr 2005; Schofield and Ratcliffe 2004).

However, under hypoxic conditions, ubiquination, and hydroxylation of the HIFA subunits do not occur, therefore allowing the HIF heterodimers to enter the nucleus where they recognize HIF-responsive elements (HREs) within the promoters of a large number of genes, effecting changes in transcriptional activity (Wenger et al. 2005). Many of these target genes increase O_2 transport to hypoxic tissues by promoting red blood cell maturation and angiogenesis/vasomotor control (Haase 2013; Majmundar et al. 2010). Therefore, it is not surprising that HIF transcription factors, and a number of their downstream targets, have been implicated in adaptation to high-altitude environments (Hanaoka et al. 2012; Qu et al. 2013; Qiu et al. 2012; Li et al. 2014; Wang et al. 2014), as well as in other O_2 depleted environments (Aggarwal et al. 2010; Peng et al. 2011; Ge et al. 2012; Xing et al. 2013; Rytkönen et al. 2007; Terova et al. 2008).

Most studies of convergent evolution in the HIF-pathway has been documented in high-altitude human populations (Simonson et al. 2012; Scheinfeldt and Tishkoff 2010). Several HIF-pathway candidate genes for adaptation to altitude have been identified. In Andean people, *VEGF*, *NOS2A*, and *EGLN1* are the best-supported candidates (Bigham et al. 2009; Bigham et al. 2010); and in Tibetans, *EPAS1* and *EGLN1* have been identified (Simonson et al. 2010; Beall et al. 2010; Yi et al. 2010; Xu et al. 2011; Peng et al. 2011; Jeong et al. 2014; Simonson et al. 2012; Lorenzo et al. 2014). Studies focused on Ethiopian highlanders suggest the involvement of a different combination of HIF genes, *BHLHE41*, *THRB*, and *ARNT2* (Scheinfeldt et al. 2012; Alkorta-Aranburu et al. 2012; Huerta-Sánchez et al. 2013). At least one study indicates that ancient admixture followed by a soft selective sweep on standing variation may be responsible for the high frequency of specific *EPAS1* variants in Tibetan people (Huerta-Sánchez et al. 2014).

Despite widespread evidence for convergence at the HIF pathway level, the pattern does not extend at the gene-level for most human population pairwise comparisons. For example, *EGLN1* represents the only overlapping gene between the studies cited above, having been identified in both Tibetan and Andean populations; however, both populations exhibit distinct haplotypes, with genotypic variants located in different exons (Bigham et al. 2010). Convergent evolution may perhaps be more widespread than the literature suggests, due to differences in analyses rather than differences in selective and demographic histories associated with each high-altitude population (Foll et al. 2014).

Here I test whether variation in 26 genes in the HIF signaling pathway are associated with high-altitude and corresponding O_2 availability in three duck species that colonized the Andes from ancestral low-altitude habitats in South America (McCracken et al. 2009a; McCracken et al. 2009b). Analysis of these three species, each of which became established at high altitude independently, allowed us to identify whether (a) polymorphism in HIF pathway genes is statistically associated with occupancy of high-altitude hypoxic environments, and (b) if so, were the same pathway genes converged upon in each duck lineage, or did novel gene usage patterns arise among the three species. Ultimately, I found strong support for both convergent and parallel evolution at different taxonomic levels. In the case of two of the three duck species, the same exonic regions in the same genes (*EGLN1*, *EPAS1*) exhibited sharply demarcated outliers with a high probability of directional selection in two high-altitude populations. This example of parallel evolution is striking given that these same two genes also have been identified in high-altitude human populations. However, in the case of humans, different regions of

these genes appear to be under directional selection, therefore illustrating an example of convergent evolution between ducks and humans. Not only does this study pinpoint the molecular mechanism for high-altitude adaptation in Andean duck species, but it also highlights the degree to which natural selection can be both predictable and unpredictable (ie. idiosyncratic), arriving at both parallel and convergent mechanisms of adaptation to similar environments.

Materials and Methods

Specimen Collection and DNA Extraction

A total of 60 individuals were used for this study from three different Andean duck species – cinnamon teal (*Anas cyanoptera*), speckled teal (*A. flavirostris*), and yellow-billed pintail (*A. georgica*). Sampling from each species was comprised of 20 individuals with 10 individuals from each of the high-altitude populations and 10 individuals from each of the low-altitude populations. For the cinnamon teal, individuals from low-altitude populations are the *A. c. cyanoptera* subspecies (n = 10; elevation range 8-23 m) and from high-altitude are the *A. c. orinomus* subspecies (n = 10; elevation range 3533-3,871 m) (Wilson et al. 2013). For the speckled teal, individuals from low-altitude populations are the *A. f. flavirostris* subspecies (n = 10; elevation range 77 - 860 m) and from high-altitude are the *A. f. oxyptera* subspecies (n = 10; elevation range 3,211 - 4,405). For the yellow-billed pintail, individuals from both populations are taxonomically identified as *Anas georgica spinicauda*. A total of 20 yellow-billed pintails were collected from low- (n = 10; elevation range 292 - 914 m) and high-altitude (n = 10;

elevation range 3,332 - 4,070 m). Genomic DNA was extracted from tissue using a DNeasy Tissue Kit (Qiagen, Valencia, California, USA) following manufacturers protocols.

HIF-Pathway Gene Rationale

HIF-pathway genes were chosen based on a combination of (1) being found to be a gene of interest in previous studies of high-altitude animals, using whole-genome scans looking to identify genetic adaptations to high-altitude or hypoxic environments, (2) being a part of the canonical HIF-pathway either through being a known downstream target, or a part of the repression machinery of the pathway, and/or (3) being a part of similar transcription factor families (i.e., bHLH-PAS containing proteins). Of the list created using these criteria, 31 were found to be present in the *Anas platyrhynchos* (mallard) genome (ENSAPLG00000004677; Table 4.1, Figure 4.1).

Target-enrichment Sequencing

I used in-solution target capture to selectively enrich libraries for regions of interest prior to NGS sequencing (Gnirke et al. 2009). All steps of the process were performed by MYcroarray (Ann Arbor, MI). A custom MYbaits® biotinylated ssRNA target capture baitset was designed for enriching target sequences. In total, 31 genes associated with the HIF-pathway from the duck nuclear genome were selected from the Ensembl *A. platyrhynchos* v1.0 genome, including all introns and exons and upstream promotor and downstream regions. These 31 sequences were first screened using the web server version of RepeatMasker (http://www.repeatmasker.org/cgi-

bin/WEBRepeatMasker) with default settings and selecting *A. platyrhynchos* as the species, soft-masking all repeats. Next, 120mer probes at 2x tiling density were designed from these soft-masked sequences. All candidate probes were then screened against the duck genome using MYcroarray's in silico bait analysis software pipeline, in order to filter out any probes that were non-specific in the genome. A final set of 12,062 filtered probes was chosen for 26 of the 31 genes (Table 4.1, Figure 4.1), which was comprised of all probes that (a) did not contain any soft-masked regions and (b) passed MYcroarray's most liberal filtering threshold. Following hybridization, target regions were purified on magnetic beads followed by post-hybridization amplification to ligate indexing sequences. Sequencing was performed on an Illumina HiSeq platform paired-end (100 bp) with a 250-300 bp insert size.

HIF-Pathway Sequence Assembly Pipeline

Sequences were received pre-parsed by individual, with adapters trimmed and quality filtered (Q < 30). Additional adapter trimming was performed utilizing *fastq-clipper* (AGATCGGAAGAGC) and remaining sequences were then filtered by length and quality using *fastq-quality filter* (reads < 20 bp, and Q < 30) from the FASTX-Toolkit v. 0.0.13 package (Gordon and Hannon 2010). A custom pipeline (https://github.com/amgraham07) was created to remove orphan sequences, and assemble sequences against the 26 reference genes using BWA v0.7.15 (Li and Durbin 2009). The Samtools package v1.3.1, including BCFtools v.1.3.1 (Li et al. 2009) was then used to create a VCF file, as well as provide assembly statistics (i.e., bp-by-bp coverage). These programs used in the pipeline called SNP variants that were different against the mallard

reference, including indels (insertion/deletion); however, the indel information was excluded in the final dataset since the softwares used do not accommodate indels.

RAD-Seq Reference Data

To generate a genome-wide reference data set to compare HIF-pathway genes to, I used RAD-Seq (Restriction Site Associated DNA Sequencing). Genomic DNA was normalized to around 2ng/uL using a Qubit Fluorometer (Invitrogen, Grand Island, New York, USA) and submitted to Floragenex (Eugene, Oregon, USA) for sequencing. In short, genomic DNA was first digested with the 8 base-pair SbfI restriction enzyme (CCTGCAGG) followed by barcode and adapter ligation. Individually barcoded samples were multiplexed and sequenced on the Illumina HiSeq 2000 with single-end 100 bp sequencing chemistry. Following the run, RAD sequences were de-multiplexed and trimmed to yield resulting RAD sequences of 90 bp. The methods used by Floragenex are described in (Hohenlohe et al. 2010; Miller et al. 2007; Baird et al. 2008), and are summarized below.

Genotypes at each nucleotide site were determined using the VCF_popgen v.4.0 pipeline to generate a customized VCF 4.1 (variant call format) database with parameters set as follows: minimum AF for genotyping = 0.075, minimum Phred score = 15, minimum depth of sequencing coverage = 10x, and allowing missing genotypes from up to 2 out of 20 individuals (10%) at each site. To filter out base calls that were not useful due to low quality scores or insufficient coverage, genotypes at each nucleotide site were inferred using the Bayesian maximum likelihood algorithm. The genotyping algorithm incorporates the site-specific sequencing error rate, and assigns the most likely diploid

genotype to each site using a likelihood ratio test and significance level of $p = 0.05$. Sites for which the test statistic is not significant are not assigned a genotype for that base in that individual, effectively removing data from sites for which there are too few high-quality sequencing reads. The sequencing coverage and quality scores were then summarized using the software VCFtools v.0.1.11 (Danecek et al. 2011). Custom perl scripts written by M. Campbell (University of Alaska Fairbanks) were used to retain bi-allelic sites only and converted to VCF file format.

All RAD clusters were then subject to a BLAST search (database – nr, e-value < e−10, annotation cut-off > 50) in Blast2GO (Conesa et al. 2005) using the taxonomy filter for birds (taxa: 8782, Aves) to determine gene identity. Ultimately, SNPs on clusters whose BLAST result was that of a protein-coding gene was then incorporated into the final dataset used for outlier analyses.

Identification of Outlier Loci

The dataset used for detection of outlier loci was comprised of (a) all SNPs associated with the HIF-pathway and (b) SNPs from RAD clusters whose BLAST results hit back to a protein coding gene (see previous section). The latter provides both a comparison and a genome-wide reference data set for the potential HIF outliers to be analyzed. I tested for signatures of directional selection using three different outlier detection techniques, which minimized the risk of detecting false positives. Outlier loci were considered to be those that were identified by at least two of the three methods as being significant outliers, with the most significant outliers classified as being identified by all three methods.

First, a genomic scan was performed by obtaining pairwise site-by-site F_{ST} calculated in Arlequin v. 3.5 (Excoffier and Lischer 2010); these values were then imported into JMP Pro 12 for distribution visualization, as well as for percentile calculations for each species. Candidate loci F_{ST} values that exceeded the 99th percentile of the F_{ST} values were therefore considered likely targets of selection. Second, MCHEZA, which implements the Dfdist function, was used to demarcate markers putatively under positive directional selection (Antao and Beaumont 2011). MCHEZA analyses were based on the Infinite Alleles Model with 50,000 simulations, a confidence interval of 0.95 and a false-discovery rate (FDR) of 0.01, using the neutral mean F_{ST} and forcing mean F_{ST} options. It is important to note here, that MCHEZA was used in lieu of LOSITAN (Antao et al. 2008) due to the size of the dataset, and current issues with Java updates (T. Antao, *pers. comm.*).

Third, I used a Bayesian approach as implemented in BayeScan v. 2.1 to again identify putatively selected loci. BayeScan uses a logistic regression model to separate locus-specific effects of selection from demographic differences (Foll and Gaggiotti 2008). For each SNP, BayeScan estimates the posterior distribution under neutrality $\alpha = 0$ and separately allowing for selection $\alpha \neq 0$ and computes the posterior odds ratio (PO) as a measure of support for the model of local adaptation relative to neutral demography. Foll (2012) proposed a logarithmic scale for the posterior odds defined as: Bayes Factor (BF) 3–10 substantial ($\log_{10}$PO > 0.5–1.0); BF 10–32 strong ($\log_{10}$PO > 1–1.5); BF 32–100 very strong ($\log_{10}$PO > 1.5–2.0); and BF 100–∞ decisive ($\log_{10}$PO > 2.0–∞) evidence for accepting a model. In the genome scans, a threshold for $\log_{10}$PO > 0.5 (substantial), representing a Bayes Factor of 3.0 and posterior probability of 0.76, was

used for a marker to be considered under selection. Therefore, loci identified as "substantial" probability of being under selection were classified as being significant outliers, under this approach.

Results

HIF Pathway Assembly Information and Population Divergence Estimates

The total number of bp for the 26 genes was 839,657 bp per individual, with a goal of ~50x coverage per gene in each individual. Across all individuals, the final coverage for each gene was between 73x (*P4HA3*) and 587x (*CUL2*), with an overall average across all genes of 415x. This coverage included genes whose total regions covered 81-100% of the reference mallard sequence, with an average of 98.1% (± 3.9%) (Table 4.2). Within each species, there were a total of (1) 16,339 HIF pathway SNPs for the cinnamon teal (2) 21,674 SNPs for speckled teal and (3) 26,484 SNPs for the yellow-billed pintail.

The RAD-Seq yielded a total of (1) 18,145 SNPs from cinnamon teal, (2) 47,732 SNPs from speckled teal, and (3) 49,670 SNPs for the yellow-billed pintail. Among these, (1) 2,762 SNPs for cinnamon teal, (2) 6,280 SNPs for speckled teal, and (3) 6,523 SNPs for yellow-billed pintail mapped to a gene region, representing 13-15% of the total RAD-Seq clusters.

F_{ST} for each SNP was calculated for three separate subsets of the data between each pair of low- and high- altitude populations: the HIF pathway only, RAD-Seq gene-only, and from the combined HIF pathway and RAD-Seq gene-only datasets. All species

showed similar average F_{ST} values, except for the HIF-pathway which had slightly greater overall divergence than the RAD-Seq dataset. These estimates mirrored previous calculations with nuclear loci (Table 4.3).

Identifying Genomic Regions with High F_{ST}

Genomic regions with high F_{ST} were examined using three different methods (percentile, Dfdist, Bayesian) for cross-comparison among species. There was general evidence of parallelism for the HIF pathway across high-altitude populations of all three duck species (Table 4.4 - 4.6). For cinnamon teal, the outliers from the Dfdist analyses showed seven outliers, with *EPAS1* (36) and NOS1 (10) having the most hits, followed by EGLN1 (4), *CLOCK* (2), and *PPARA1*, *MTOR*, and *P4HA3* (1 each); however, none were additionally significant in the corresponding Bayesian analysis, nor were any variants in exonic regions (Table 4.4). The strongest outliers in speckled teal and yellow-billed pintail were *EGLN1* and *EPAS1* and were highly significant in both the Bayesian and Dfdist analyses for both species (Figure 4.2; Table 4.5, 4.6). Although most of the SNPs associated with these two genes were located in intronic/noncoding regions, both speckled teal and yellow-billed pintail contained highly significant F_{ST} outliers in exon 6 (YBP) and exon 12 (ST, YBP) of *EPAS1* and in exon 2 (YBP) of *EGLN1*. All SNP variants in the low-altitude populations were identical to the ancestral alleles found in the mallard reference, whereas the SNP variants in high-altitude populations were derived.

Conservation of Exons in Human and Mallard EPAS1 and EGLN1

Further analyses of focused on those high F_{ST} variants found in exon 6, 12 in *EPAS1* and exon 2 in *EGLN1*. In order to gauge the conservation of exon sequence between ducks and humans, the full protein sequences from mallard and human were aligned to each other and exons 6 and 12 were annotated (Figure 4.3). Overall, the alignment showed sufficient conservation between sequences, in combination with orthology quality-control statistics from ENSEMBL (GOC – 100/100; WGA – 91.92/100). In speckled teal and yellow-billed pintail, four out of the five SNPs resulted in non-synonymous substitutions in the first and second codon positions (Figure 4.4, 4.5). Specifically, in EPAS1 there was a Cys → Tyr substitution at amino acid position 23 in exon 6 and then Arg → His and Tyr → His substitutions at amino acid positions 32 and 127, respectively, in speckled teal. In yellow-billed pintail, there were two nonsynonymous substitutions at two different positions in exon 12, Pro → Glu at position 71 and Ala → Thr at amino acid position 104. For EGLN1, the same general approach was taken, with the mallard and human protein sequence, and human and mallard nucleotide sequence for exon 2; however, the nucleotide change resulted in a synonymous substitution.

Biochemical Properties and Structural locations of Amino Acid Replacements in EPAS1 and EGLN1

The two nonsynonymous SNPs in *EPAS1* exon 12 of yellow-billed pintail resulted in substitutions of amino acids with very different biochemical properties. The Pro → Glu substitution at amino acid position 71 resulted in a change from a neutral,

polar amino acid, to a hydrophobic, nonpolar amino-acid in the high-altitude population (Figure 4.5). The Ala → Thr substitution at amino acid position 104 resulted in a change from a nonpolar amino acid with a hydrophobic side-chain, to a neutral, polar amino acid that has a non-aromatic hydroxyl group attached in the high-altitude population (Figure 4.5).

Of the three nonsynonymous SNPs in *EPAS1* exon 12 in speckled teal two resulted in conservative substitutions with similar biochemical properties, Cys → Tyr at position 23 and Arg → His at position 32. Tyr → His at amino acid position 127, however, resulted in substitution of a nonpolar aromatic amino acid with a hydrophobic side-chain, to a basic, polar amino acid with a positively charged side-chain in the high-altitude population. Locations of these variants were unable to be placed on existing PDB protein models, due to a general lack of models containing any of the ODD/NTAD/CTAD domains.

Discussion

The results of this study show a large degree of parallel evolution on the HIF-pathway in two Andean duck species (*EPAS1*, *EGLN1*), and to some degree in all three (*EPAS1*). For the speckled teal, and yellow-billed pintail, this parallelism has occurred at the level of the same genes (*EPAS1, EGLN1*) and the same gene regions (exon 12), but not at the same amino acid positions, as substitutions even if they are adjacent in the protein, have occurred at different sites in high-altitude lineages. The results also show significant levels of convergence with previously studied populations of other organisms living at high-altitudes, specifically humans. This convergence extends out to the

specific genes associated with adaptations to low-oxygen environments, including the specific molecular strategies associated with the genetic variation (ie. effecting changes in HIF activity through the ODD/NTAD domains).

Distinctions are traditionally made between what constitutes convergent evolution and parallelism, although such a distinction has typically been centered around phenotype – when a given phenotype evolves, it is assumed that in the case of convergence, the underlying genetic mechanisms are different in distantly related species, but in the case of parallelism, similar in closely related species (Arendt and Reznick 2008). In this study, I compared three species of Andean ducks, which are closely related, to two populations of humans. Thus, my results indicate that there has been parallel evolution between speckled teal and yellow-billed pintail (same gene, same exons, adjacent sites in the protein), but convergent evolution between ducks and high-altitude humans (same genes, different exons/introns).

It is thought that genes that form a hub in a regulatory network between a series of upstream activators and a battery of downstream effector genes have been proposed to be "hotspot" genes, and therefore more likely to become the targets of repeated evolution (Martin and Orgogozo 2013; Stern and Orgogozo 2009; Rosenblum et al. 2014). It is also thought that such variation would be more likely to occur in domains associated with protein-protein interactions rather than DNA-binding activity (Wagner and Lynch 2008), they play an essential role in coordinating the expression of target genes in response to multiple input signals in a way that cis-regulatory elements cannot (Wagner and Lynch 2010). Ultimately, these results confirm that the HIF-pathway, and

perhaps even *EPAS1* and *EGLN* are "hotspots", due to their specific position in the regulatory network response to hypoxia (Martin and Orgogozo 2013; Beall 2007).

Linkage Disequilibrium Between EPAS1 and EGLN1

Finally, the different HIF-pathway genes targeted in this study were largely distributed on different chromosomes. However, *EGLN1* and *EPAS1* are both located on chromosome 3 in the chicken genome. This may suggest a role for linkage disequilibrium in driving the pattern seen in these results; however, there are several reasons this seems an unlikely scenario. First, although they are on the same chromosome, *EGLN1* and *EPAS1* are 1.3 million bp distance from each other, which is a large distance for strong LD to persist especially in birds whose genomes have shown rapid decay of LD over relatively short distances (Balakrishnan and Edwards 2009; Backström et al. 2006; Groenen et al. 2000; Megens et al. 2009). Second, two other genes sampled (*VEGF*, *ANGPT2*) are also on chromosome 3, with *VEGF* being in between *EGLN1* and *EPAS1*, and closer to *EPAS1* (4kb). Yet, neither show significant outliers due to genetic hitchhiking, or from physical linkage. Thus, it seems unlikely that LD is playing a large role in the variation seen between the two genes in speckled teal and yellow-billed pintail.

The Role of the HIF-Pathway in Adaptations to High-altitude Environments

Much of what we know about how the HIF-pathway is associated with adaptation to high altitude environments comes from early genome-wide association studies in humans; these genome-wide scans identified many candidate genes that may

contribute to adaptive evolution, including two genes (*EPAS1* and *EGLN1*) that are involved in the HIF-pathway, which showed the strongest signals of selective sweeps in Tibetan and Andean human (Beall et al. 2010; Bigham et al. 2010; Peng et al. 2011; Simonson et al. 2012; Xu et al. 2011), as well as in other high-altitude human populations such as Caucasian/Russians (Pagani et al. 2012) and Mongolians (Xing et al. 2013). Additionally, animals residing in similar high-altitude environments have also had the same HIF-pathway members identified as major outliers, including ungulates (Song et al. 2016), birds (Qu et al. 2013), pigs (Ai et al. 2014; Li et al. 2013a), dogs (Li et al. 2014; Gou et al. 2014).

Under normoxia, EGLN1 can perform its O_2-dependent hydroxylase function on both HIF1A and EPAS1 proteins, which triggers degradation via additional enzymatic complexes (VHL, elonginB/C, CUL2). However, under hypoxia, the hydroxylase activity of EGLN1 is suppressed, resulting in the accumulation of HIFA's that can then transactivate hundreds of downstream target genes, thus inducing numerous physiological responses, including changes in red-blood cell (RBC) production (Hu et al. 2003; Haase 2013).

It was reported that the Tibetan version of the *EGLN1* haplotype carries two Tibetan-specific missense mutations, which are thought to increase HIF-degradation under hypoxia, cancelling HIF-mediated erythropoiesis, effectively protecting Tibetans from RBC overproduction (i.e., erythrocytosis/polycythemia) at high altitudes (Lorenzo et al. 2014; Song et al. 2014; Xiang et al. 2013). However, exactly how the Tibetan

version of *EPAS1* works is still unclear - unlike *EGLN1* variants, many *EPAS1* variants are located in the noncoding regions, suggesting that they could affect the regulation of *EPAS1* at the transcriptional level (Peng et al. 2017; Peng et al. 2011).

EPAS1and EGLN1 in High-altitude Adaptation in Andean ducks

In the results of this study, most EPAS1 outliers were distributed in intronic regions, although there were five that were located within exons, three in speckled teal and two in yellow-billed pintail, respectively (Table 4.5, 4.6). In the latter case, these are located in two exons, which are within critical domains of the EPAS1 protein that define HIF function, the PAS domain (exon 6) and ODD/NTAD domain (exon 12; Figure 4.6). Even though humans and birds are separated by millions of years, both have retained very similar exon structure, (Figure 4.3). It thus seems appropriate to extrapolate results from what is known about similar variation in human, and how it might relate back to these Andean duck species.

In humans, variation in exon 12 has been implicated with both loss-of-function and gain-of-function mutations associated with red-blood cell (RBC) production. Specifically, gain-of-function mutations are known to cause erythrocytosis, resulting in an increased number of RBCs through its regulation of erythropoietan (EPO), in combination with pulmonary arterial hypertension (Percy et al. 2008b; Percy et al. 2008a; Gale et al. 2008). Besides human studies, similar gain-of-function variants in exon 12 have been associated with high-altitude pulmonary hypertension in cattle (Newman et al. 2015). Speckled teal had additional outliers in *EPAS1* located in exon 6, which correspond to the same exon in humans (Figure 4.3). However, unlike exon 12,

variation in/near exon 6 has been identified in human high-altitude populations (Yi et al. 2010), and is associated with changes in heart rate and hypertension (Buroker et al. 2012). Although, in both the human and cattle studies, those variants were largely considered deleterious, they demonstrate the ability for genetic variation in exon 12 of *EPAS1* to have direct physiological effects frequently targeted by selection in hypoxic environment. Ultimately, my results mirror variation in *EPAS1* in other human populations, but largely do not overlap with the types of changes associated with high-altitude adaptation in Tibetan and Andean populations.

Unlike *EPAS1*, high-altitude human populations have shown both loss-of-function, and gain-of-function mutations associated with *EGLN1*, through variation in the proteins' zinc-finger domains (Song et al. 2014; Xiang et al. 2013). It is thought that under hypoxic conditions, hydroxylation through EGLN1 is inhibited through variants, thereby maintaining a stable EPAS1 protein, activating a broad range of effects orchestrating acclimatization to hypoxia, potentially leading to a blunted hypoxic response (Song et al. 2014). Another avenue is that the missense variants in *EGLN1* promote increased HIF degradation, with *EPAS1* effectively downregulated, thereby protecting the individual from damaging physiological compensatory mechanisms associated with chronic hypoxia (erythrocytosis/polycythemia, hypertension etc.) due to a blunted hypoxic acclimatization response (Lorenzo et al. 2014; Xiang et al. 2013). Lastly, it has been suggested that variants associated with *EPAS1* cause weaker activity at the downstream genes with HRE promoters, thus also potentially causing a blunted set of compensatory mechanisms (Peng et al. 2017). However, the results identified no

variation in coding regions of EGLN1, but rather F_{ST} outliers were associated with intronic regions, suggesting a role for transcriptional regulation of *EGLN1* in high-altitude adaptation in these Andean duck species.

Despite *EGLN1* being at the focus of adaptations associated with high-altitude human populations, it has been suggested that high-altitude human variants in *EPAS1* have a direct effect on erythropoietin (EPO) production, thus changing RBC production and blunting the hypoxic response (Peng et al. 2011). In combination with known variants in other human populations in *EPAS1*, there seems to be evidence of this gene effecting various hemoglobin traits (i.e., Hb concentration, hematocrit/Hct levels), perhaps by directly altering transcription of EPO. However, another possibility could be that these variants are having an indirect effect by generally enhancing O_2 perfusion to tissues, including the kidneys, which are the site of EPO production. This would result in a similar response in changes to RBC/Hb/Hct levels, without EPAS1 ever encountering the HRE associated with the upstream regulatory elements of *EPO*. It is important to note, however, that EPO and hemoglobin are not the only changes in response to EPAS1 activity. The Tibetan *EPAS1* variants have been associated with low pulmonary arterial pressure that is considered an adaptive trait (Wu and Kayser 2006), through a down-regulation of *ACE* and *VEGF* (Peng et al. 2017).

Exactly which mechanistic avenues these Andean duck species are using is unclear, given the current data; however, it is interesting to note that speckled teal, like Tibetan humans in particular, are characterized by Hb levels matching that of their low-altitude counterparts, unlike other some Andean duck populations, which show more elevated Hb like Andean people (Beall 2007). As discussed above, this could suggest a

connection between *EPAS1*, *EGLN1* and Hb production, especially based on what we know about Tibetan and Andean populations, whose genetic variation in both genes has been implicated in protecting individuals from erythrocytosis/polycythemia resulting from living at high-altitude. With potentially causative exonic variation present in *EPAS1* and intronic variation in *EGLN1*, the relationship between these two genes and potential avenues for high-altitude adaptation mimics human populations, although in reverse, since its intronic variation in *EPAS1* and exonic variation in *EGLN1*. A combination of gain-of-function or loss-of-function mutations in *EPAS1*, and transcriptional regulation differences between *EGLN1* haplotype variants, may be causing changes in downstream target transactivation, resulting in a blunted hypoxic response.

Conclusions

Hypoxia as a stressor in high-altitude environments has shown to have facilitated a great degree of parallelism and convergence across many animals, but especially in the case of human and Andean duck populations, as demonstrated by the results of this research. Ultimately, I was able to show not only a high degree of convergence between high-altitude human and duck populations on the HIF-pathway itself, but that convergence included two gene members, *EPAS1* and *EGLN1*. In addition, between the high-altitude populations of speckled teal and yellow-billed pintail, there was evidence of strong molecular parallelism at the location of specific exons (exon 12) resulting in nonsynonymous changes in close proximity to protein domains associated with oxygen-driven protein stability and transactivation. Although the specific molecular mechanisms

associated with these variants are currently unknown, previous work suggests that these variants are likely resulting in a blunted hypoxic response, potentially through hemoglobin and other downstream targets. Not only was I able to identify a potential molecular mechanism for high-altitude adaptation in Andean duck species, through the HIF-pathway, but the results also emphasize how frequently natural selection can select predictable genes and pathways, while also selecting on idiosynchratic molecular underpinnings, ultimately reaching both parallel and convergent mechanisms of adaptation to similar environments.

Table 4.1: List of HIF signaling pathway members, their location in the mallard genome (ENSEMBL), including the number of probes/baits designed for each gene (MYcroarray).

Gene Name	Location in Duck Genome (ENSEMBL)	begin	end	total length	# Probes
ACE	Scaffold KB744725.1: 386,729-403,257	386729	403257	16,528	236
ANGPT1	Scaffold KB743875.1: 238,671-419,023	238671	419023	180,352	2,251
ANGPT2	Scaffold KB742406.1: 1,290,414-1,334,615	1290414	1334615	44,201	627
ARNT	Scaffold KB744049.1: 20,467-38,811	20467	38811	18,344	284
ARNT2	Scaffold KB744113.1: 278,897-358,152	278897	358152	79,255	1201
ARNTL	Scaffold KB744206.1: 57,975-84,046	57975	84046	26,071	347
CLOCK	Scaffold KB742619.1: 827,048-854,490	827048	854490	27,442	414
CUL2	Scaffold KB742539.1: 2,267,987-2,311,340	2267987	2311340	43,353	669
EGLN1	Scaffold KB743594.1: 39,865-71,284	39865	71284	31,419	454
EGLN3	Scaffold KB743383.1: 66,183-90,875	66183	90875	24,692	377
EPAS1 (HIF2α)	Scaffold KB742444.1: 734,232-772,478	734232	772478	38,246	553
HIF1α	Scaffold KB745078.1: 277,272-293,388	277272	293388	16,116	257
HSP90	Scaffold KB743197.1: 687,101-694,368	687101	694368	7,267	112
MTOR	Scaffold KB743246.1: 258,381-322,791	258381	322791	64,410	999
NOS1	Scaffold KB744108.1: 141,687-174,300	141687	174300	32,613	504
NOS2	Scaffold KB743217.1: 377,684-396,078	377684	396078	18,394	288
P4HA1	Scaffold KB744430.1: 583,757-609,851	583757	609851	26,094	378
P4HA2	Scaffold KB743591.1: 85,973-106,217	85973	106217	20,244	316
P4HA3	Scaffold KB744216.1: 262,001-267,897	262001	267897	5,896	60
PDHA1	Scaffold KB742701.1: 132,835-143,350	132835	143350	10,515	172
PPARA (alpha)	Scaffold KB742459.1: 937,145-969,765	937145	969765	32,620	463
PPARD (delta)	Scaffold KB743510.1: 216,148-220,291	216148	220291	4,143	63
PPARG (gamma)	Scaffold KB742679.1: 37,634-61,994	37634	61994	24,360	363
TCEB3 (elongin A)	Scaffold KB742637.1: 387,311-394,427	387311	394427	7,116	105
THRB	Scaffold KB744720.1: 282,100-308,664	282100	308664	26,564	385
VEGF	Scaffold KB742809.1: 508,129-521,531	508129	521531	13,402	184

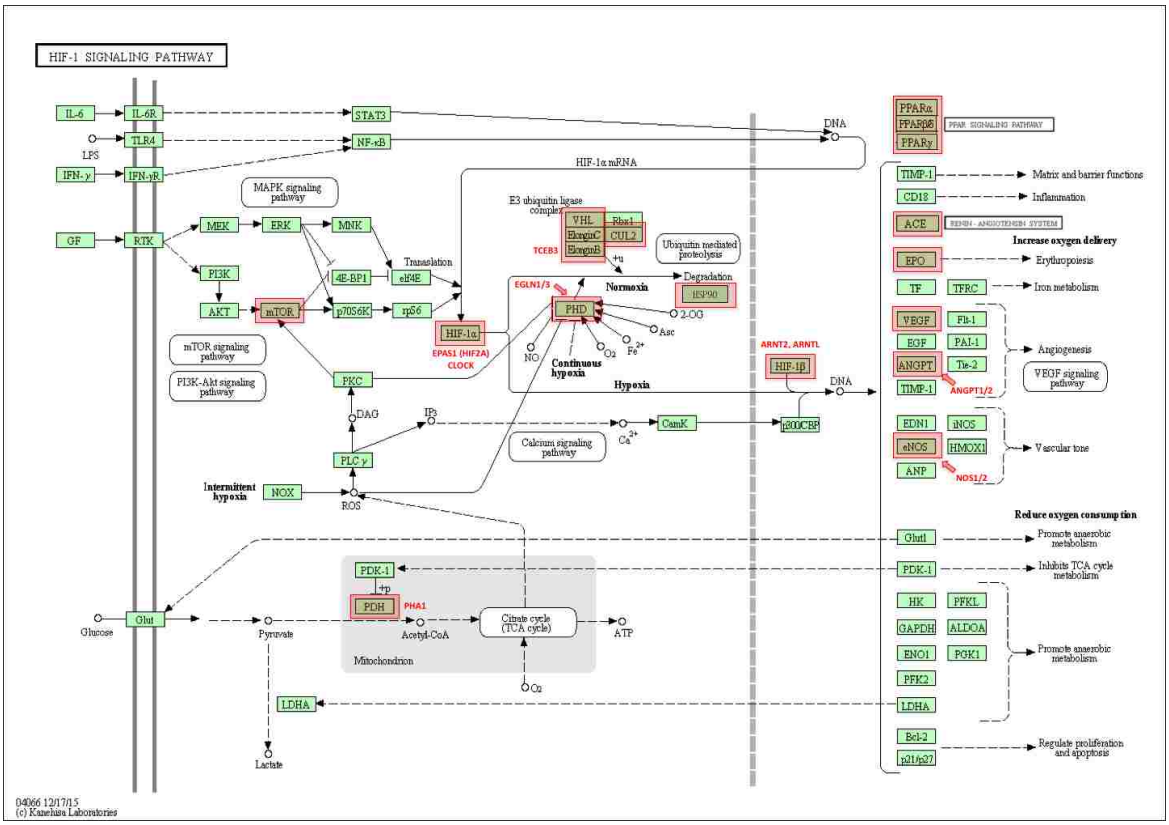

Figure 4.1: The HIF signaling pathway, and downstream targets. Genes highlighted in red are those which were sequenced through target-enrichment; some members sequenced are paralogs of existing members of this pathway or of downstream targets (in red text). Modified from KEGG database (Kanehisa and Goto 2000) - image used with written permission from Kanehisa Laboratories.

Table 4.2: Assembly statistics for the 26 genes sequenced through target-enrichment – "depth of coverage" represents the number of reads that included in the reconstructed sequence against the mallard reference; "gene coveraged" refers to the total amount of the mallard reference that included any amount of coverage.

	Speckled teal		Yellow-billed pintail		Cinnamon teal	
Gene Name	**Cov. Depth**	**Gene Covered**	**Cov. Depth**	**Gene Covered**	**Cov. Depth**	**Gene Covered**
ACE	147	97.7%	157	99.2%	165	99.2%
ANGPT1	456	95.5%	463	96.1%	492	94.6%
ANGPT2	514	99.4%	518	99.6%	570	99.6%
ARNT	411	99.5%	428	99.8%	437	99.8%
ARNT2	463	99.8%	478	99.9%	520	99.9%
ARNTL	558	91.7%	567	92.1%	601	90.9%
CLOCK	463	99.8%	471	100.0%	497	100.0%
CUL2	589	99.9%	592	100.0%	572	98.0%
EGLN1	388	99.3%	404	99.7%	398	99.6%
EGLN3	421	99.3%	436	99.6%	460	99.7%
EPAS1	456	99.7%	457	100.0%	485	100.0%
HIF1α	512	99.9%	523	100.0%	546	100.0%
HSP90AA1	476	99.9%	474	100.0%	405	100.0%
MTOR	476	99.8%	485	100.0%	510	100.0%
NOS1	140	95.2%	151	96.6%	152	96.1%
NOS2	540	100.0%	549	100.0%	602	99.9%
P4HA1	406	98.0%	413	98.5%	455	98.4%
P4HA2	509	99.7%	514	99.9%	569	99.8%
P4HA3	72	79.8%	78	81.8%	72	81.5%
PHD1A	507	100.0%	506	99.9%	500	99.6%
PPARA	532	98.4%	557	98.6%	596	97.7%
PPARD	119	98.2%	132	100.0%	121	99.3%
PPARG	516	99.7%	523	99.7%	565	99.8%
TCEB3	127	95.7%	143	97.8%	128	96.6%
THRB	430	99.5%	439	99.9%	492	99.8%
VEGF	219	99.0%	229	99.8%	234	99.8%

Table 4.3: Estimates of divergence ***(F_{ST})*** for the three species across the different data-sets, including [1] the variants associated with the HIF-pathway (ie. 26 genes), [2] variants associated with the RAD-seq dataset whose sequence had a significant BLAST hit to any protein-coding gene, and [3] the combined variant dataset including both [1-2].

	F_{ST}		
Species	**HIF Pathway Enriched**	**RAD-seq (Gene)**	**Combined**
Cinnamon teal	0.051	0.042	0.04
Yellow-billed pintail	0.045	0.013	0.02
Speckled teal	0.124	0.058	0.06

Table 4.4: Outlier list from MCHEZA in Cinnamon teal, including the SNP variants associated with the HIF-pathway (and corresponding position on the gene) who met the FDR > 0.99, and who also were in the top 1% of F_{ST} values; het – heterozygosity.

Gene	*POS*	*Het*	F_{ST}	*P (Simul F_{ST} <sample F_{ST})*	Exon Overlap
CLOCK	17325	0.255142	0.614556	1	-
CLOCK	17366	0.28188	0.564621	1	-
EGLN1	5005	0.509129	0.567272	1	-
EGLN1	24029	0.507958	0.524956	1	-
EGLN1	24100	0.507958	0.524956	1	-
EGLN1	29894	0.486069	0.535239	1	-
EPAS1	3914	0.483233	0.518324	1	-
EPAS1	4012	0.483233	0.518324	1	-
EPAS1	4018	0.483233	0.518324	1	-
EPAS1	4038	0.483233	0.518324	1	-
EPAS1	9066	0.491823	0.718793	1	-
EPAS1	9111	0.491823	0.718793	1	-
EPAS1	9432	0.491823	0.718793	1	-
EPAS1	9873	0.491823	0.718793	1	-
EPAS1	12027	0.491823	0.718793	1	-
EPAS1	12059	0.483233	0.753212	1	-
EPAS1	12684	0.507958	0.524956	1	-
EPAS1	12698	0.507958	0.524956	1	-
EPAS1	12779	0.491823	0.718793	1	-
EPAS1	12964	0.491823	0.718793	1	-
EPAS1	13046	0.507958	0.524956	1	-
EPAS1	13785	0.507958	0.524956	1	-
EPAS1	14303	0.483233	0.753212	1	-
EPAS1	14414	0.507798	0.607577	1	-
EPAS1	14502	0.483233	0.753212	1	-
EPAS1	14672	0.498905	0.683147	1	-
EPAS1	15292	0.509129	0.567272	1	-
EPAS1	15422	0.507958	0.524956	1	-
EPAS1	15502	0.356756	0.526242	1	-
EPAS1	15887	0.507958	0.524956	1	-
EPAS1	17001	0.507958	0.524956	1	-
EPAS1	17083	0.507958	0.524956	1	-
EPAS1	20362	0.504302	0.646132	1	-
EPAS1	20799	0.507958	0.524956	1	-
EPAS1	21693	0.507798	0.607577	1	-

EPAS1	23022	0.507958	0.524956	1	-
EPAS1	24169	0.483233	0.518324	1	-
EPAS1	24204	0.462104	0.599203	1	-
EPAS1	24566	0.462104	0.599203	1	-
EPAS1	24950	0.512892	0.610632	1	-
EPAS1	25792	0.509129	0.567272	1	-
EPAS1	26691	0.509129	0.567272	1	-
MTOR	29375	0.507958	0.524956	1	-
NOS1	9350	0.266267	0.679482	1	-
NOS1	19393	0.509129	0.567272	1	-
NOS1	19394	0.509129	0.567272	1	-
NOS1	20356	0.45071	0.727606	1	-
NOS1	23680	0.45139	0.757868	1	-
NOS1	23724	0.496403	0.799555	1	-
NOS1	23747	0.496403	0.799555	1	-
NOS1	23772	0.515465	0.823984	1	-
NOS1	23835	0.496403	0.799555	1	-
NOS1	24359	0.507958	0.524956	1	-
P4HA3	3770	0.507958	0.524956	1	-
PPARA1	4399	0.345688	0.663169	1	-

Table 4.5: Outlier list from MCHEZA in yellow-billed pintail, including the SNP variants associated with the HIF-pathway (and corresponding position on the gene) who [1] were in the top 1% of of F_{ST} values, [2] met the FDR > 0.99 in MCHEZA, and [3] whose Log10(PO)>0.5 (ie."substantial") in BayeScan.

MCHEZA + 99th percentile (>0.710)					*BayeScan (Log10(PO)>0.5 = "substantial")*							
Gene	*POS*	*Het*	*F_{ST}*	*P (Simul F_{ST} <sample F_{ST})*	*Gene*	*POS*	*prob*	*log10(PO)*	*qval*	*alpha*	*Simul F_{ST}*	*Exon Overlap*
EGLN1	881	0.462104	0.818823	1	EGLN1	881	0.9956	2.3545	0.0028542	2.0586	0.32283	-
EGLN1	2779	0.473285	0.786521	1	EGLN1	2779	0.96439	1.4327	0.008721	1.9034	0.29609	-
EGLN1	2914	0.473285	0.786521	1	EGLN1	2914	0.96619	1.4561	0.0053532	1.917	0.29872	-
EGLN1	3559	0.483233	0.753212	1	EGLN1	3559	0.83417	0.70158	0.035227	1.601	0.2541	-
EGLN1	5997	0.462104	0.818823	1	EGLN1	5997	0.998	2.698	0.0016003	2.0389	0.31886	Exon2
EGLN1	9377	0.483233	0.753212	1	EGLN1	9377	0.83357	0.6997	0.036653	1.5882	0.25194	-
EGLN1	27855	0.473285	0.786521	1	EGLN1	27855	0.90678	0.988	0.023966	1.7225	0.26854	-
EPAS1	2599	0.422162	0.71501	1	EPAS1	2599	0.9974	2.5838	0.0021583	1.9724	0.30429	-
EPAS1	2621	0.422162	0.71501	1	EPAS1	2621	0.9978	2.6565	0.0018004	1.9712	0.30403	-
EPAS1	2631	0.422162	0.71501	1	EPAS1	2631	0.9972	2.5515	0.0023125	1.9562	0.30137	-
EPAS1	6305	0.422162	0.71501	1	EPAS1	6305	0.9966	2.467	0.0024779	1.9689	0.3039	-
EPAS1	25376	0.422162	0.71501	1	EPAS1	25376	0.9962	2.4185	0.0026177	1.967	0.30361	-
EPAS1	26021	0.422162	0.71501	1	EPAS1	26021	0.997	2.5215	0.0023862	1.97	0.30383	-
EPAS1	26519	0.422162	0.71501	1	EPAS1	26519	0.9984	2.7951	0.0010002	1.9638	0.30268	-
EPAS1	27470	0.422162	0.71501	1	EPAS1	27470	0.9958	2.3748	0.0027749	1.9537	0.30089	-
EPAS1	27546	0.422162	0.71501	1	EPAS1	27546	0.9976	2.6187	0.0020404	1.9745	0.30523	-
EPAS1	27862	0.483233	0.753212	1	EPAS1	27862	0.83637	0.70853	0.030786	1.6061	0.25489	-
EPAS1	28384	0.422162	0.71501	1	EPAS1	28384	0.9968	2.4934	0.0024143	1.97	0.30406	-
EPAS1	28474	0.422162	0.71501	1	EPAS1	28474	0.9966	2.467	0.0024779	1.9787	0.30543	-
EPAS1	28583	0.422162	0.71501	1	EPAS1	28583	0.9972	2.5515	0.0023125	1.9497	0.30017	-
EPAS1	29121	0.422162	0.71501	1	EPAS1	29121	0.9976	2.6187	0.0020404	1.9659	0.30307	-
EPAS1	30636	0.422162	0.71501	1	EPAS1	30636	0.9962	2.4185	0.0026177	1.9692	0.30389	-
EPAS1	30642	0.422162	0.71501	1	EPAS1	30642	0.998	2.698	0.0016003	1.9788	0.3058	-
EPAS1	30855	0.422162	0.71501	1	EPAS1	30855	0.9956	2.3545	0.0028542	1.9578	0.30169	-
EPAS1	30872	0.422162	0.71501	1	EPAS1	30872	0.9978	2.6565	0.0018004	1.9581	0.30135	-
EPAS1	31342	0.422162	0.71501	1	EPAS1	31342	0.9972	2.5515	0.0023125	1.9658	0.30341	-
EPAS1	31362	0.422162	0.71501	1	EPAS1	31362	0.9976	2.6187	0.0020404	1.9625	0.30187	-
EPAS1	31413	0.422162	0.71501	1	EPAS1	31413	0.9974	2.5838	0.0021583	1.9683	0.30328	-
EPAS1	31491	0.422162	0.71501	1	EPAS1	31491	0.997	2.5215	0.0023862	1.9641	0.30264	-
EPAS1	31559	0.422162	0.71501	1	EPAS1	31559	0.9972	2.5515	0.0023125	1.9612	0.30191	-
EPAS1	32499	0.422162	0.71501	1	EPAS1	32499	0.9974	2.5838	0.0021583	1.9655	0.30336	-
EPAS1	32739	0.422162	0.71501	1	EPAS1	32739	0.9972	2.5515	0.0023125	1.9528	0.30062	Exon 12
EPAS1	32837	0.422162	0.71501	1	EPAS1	32837	0.9972	2.5515	0.0023125	1.9552	0.30085	Exon 12

EPAS1	33197	0.422162	0.71501	1	EPAS1	33197	0.9976	2.6187	0.0020404	1.9575	0.30142	-
EPAS1	33971	0.422162	0.71501	1	EPAS1	33971	0.997	2.5215	0.0023862	1.9591	0.30137	-
EPAS1	34510	0.422162	0.71501	1	EPAS1	34510	0.996	2.3961	0.0026561	1.987	0.30728	-
EPAS1	34514	0.422162	0.71501	1	EPAS1	34514	0.9982	2.7439	0.0014003	1.9677	0.30342	-
EPAS1	34516	0.422162	0.71501	1	EPAS1	34516	0.9964	2.442	0.002546	1.9738	0.30496	-
EPAS1	34519	0.422162	0.71501	1	EPAS1	34519	0.9978	2.6565	0.0018004	1.9739	0.30495	-
EPAS1	34619	0.422162	0.71501	1	EPAS1	34619	0.9976	2.6187	0.0020404	1.9642	0.30255	-
EPAS1	34677	0.422162	0.71501	1	EPAS1	34677	0.9974	2.5838	0.0021583	1.9639	0.30276	-

Table 4.6: Outlier list from MCHEZA in speckled teal, including the SNP variants associated with the HIF-pathway (and corresponding position on the gene) who [1] were in the top 1% of of F_{ST} values, [2] met the FDR > 0.99 in MCHEZA, and [3] whose Log10(PO)>0.5 (ie."substantial") in BayeScan.

MCHEZA + 99th percentile Fst (>0.889)					*BayeScan (Log10(PO)>0.5 = "substantial")*							
Gene	***POS***	***Het***	***F_{ST}***	***P (Simul F_{ST} <sample F_{ST})***	***Gene***	***POS***	***prob***	***log10(PO)***	***qval***	***alpha***	***Simul F_{ST}***	***Exon Overlap***
EGLN1	16425	0.422162	0.910485	1	EGLN1	16425	0.86977	0.82471	0.060736	1.491	0.51713	-
EGLN1	16427	0.422162	0.910485	1	EGLN1	16427	0.94759	1.2572	0.050477	1.7731	0.57711	-
EPAS1	13775	0.422162	0.910485	1	EPAS1	13775	0.93439	1.1535	0.05595	1.7743	0.57747	-
EPAS1	14848	0.422162	0.910485	1	EPAS1	14848	0.94779	1.259	0.04951	1.7806	0.57862	Exon 6
EPAS1	15026	0.422162	0.910485	1	EPAS1	15026	0.95379	1.3147	0.046209	1.7915	0.58056	-
EPAS1	16831	0.422162	0.910485	1	EPAS1	16831	0.94319	1.2202	0.053068	1.7815	0.5786	-
EPAS1	16832	0.422162	0.910485	1	EPAS1	16832	0.94299	1.2185	0.053331	1.7792	0.57807	-
EPAS1	26494	0.422162	0.910485	1	EPAS1	26494	0.94279	1.2169	0.053573	1.77	0.5762	-
EPAS1	28232	0.422162	0.910485	1	EPAS1	28232	0.94419	1.2283	0.052511	1.7804	0.57853	-
EPAS1	29126	0.422162	0.910485	1	EPAS1	29126	0.94399	1.2267	0.05278	1.7757	0.57733	-
EPAS1	29716	0.422162	0.910485	1	EPAS1	29716	0.94099	1.2026	0.054521	1.7959	0.58165	-
EPAS1	30099	0.422162	0.910485	1	EPAS1	30099	0.94219	1.2121	0.054044	1.7845	0.57909	-
EPAS1	30102	0.422162	0.910485	1	EPAS1	30102	0.94219	1.2121	0.054044	1.784	0.57919	-
EPAS1	30332	0.422162	0.910485	1	EPAS1	30332	0.94499	1.235	0.05221	1.7813	0.57881	-
EPAS1	31922	0.423029	0.904802	1	EPAS1	31922	0.89318	0.92228	0.058255	1.6688	0.5548	-
EPAS1	32283	0.437598	0.909661	1	EPAS1	32283	0.93039	1.126	0.056456	1.7548	0.57298	-
EPAS1	32384	0.422162	0.910485	1	EPAS1	32384	0.94139	1.2058	0.054285	1.7719	0.57699	-
EPAS1	32388	0.422162	0.910485	1	EPAS1	32388	0.93979	1.1933	0.055563	1.7786	0.57794	-
EPAS1	32410	0.422162	0.910485	1	EPAS1	32410	0.94019	1.1964	0.055369	1.7745	0.57717	-
EPAS1	32411	0.422162	0.910485	1	EPAS1	32411	0.94539	1.2383	0.05193	1.7724	0.57668	-
EPAS1	32412	0.422162	0.910485	1	EPAS1	32412	0.94059	1.1995	0.054754	1.7586	0.57385	-
EPAS1	32449	0.422162	0.910485	1	EPAS1	32449	0.94039	1.198	0.055176	1.7724	0.57662	-
EPAS1	32453	0.422162	0.910485	1	EPAS1	32453	0.94739	1.2555	0.050782	1.7792	0.57848	-
EPAS1	32497	0.422162	0.910485	1	EPAS1	32497	0.94539	1.2383	0.05193	1.7725	0.57647	-
EPAS1	32501	0.422162	0.910485	1	EPAS1	32501	0.94899	1.2696	0.04861	1.7892	0.58042	-
EPAS1	32622	0.422162	0.910485	1	EPAS1	32622	0.94539	1.2383	0.05193	1.7877	0.57993	Exon 12
EPAS1	32764	0.422162	0.910485	1	EPAS1	32764	0.94759	1.2572	0.050477	1.7875	0.5804	Exon 12
EPAS1	32906	0.422162	0.910485	1	EPAS1	32906	0.95139	1.2916	0.047409	1.7906	0.58065	Exon 12

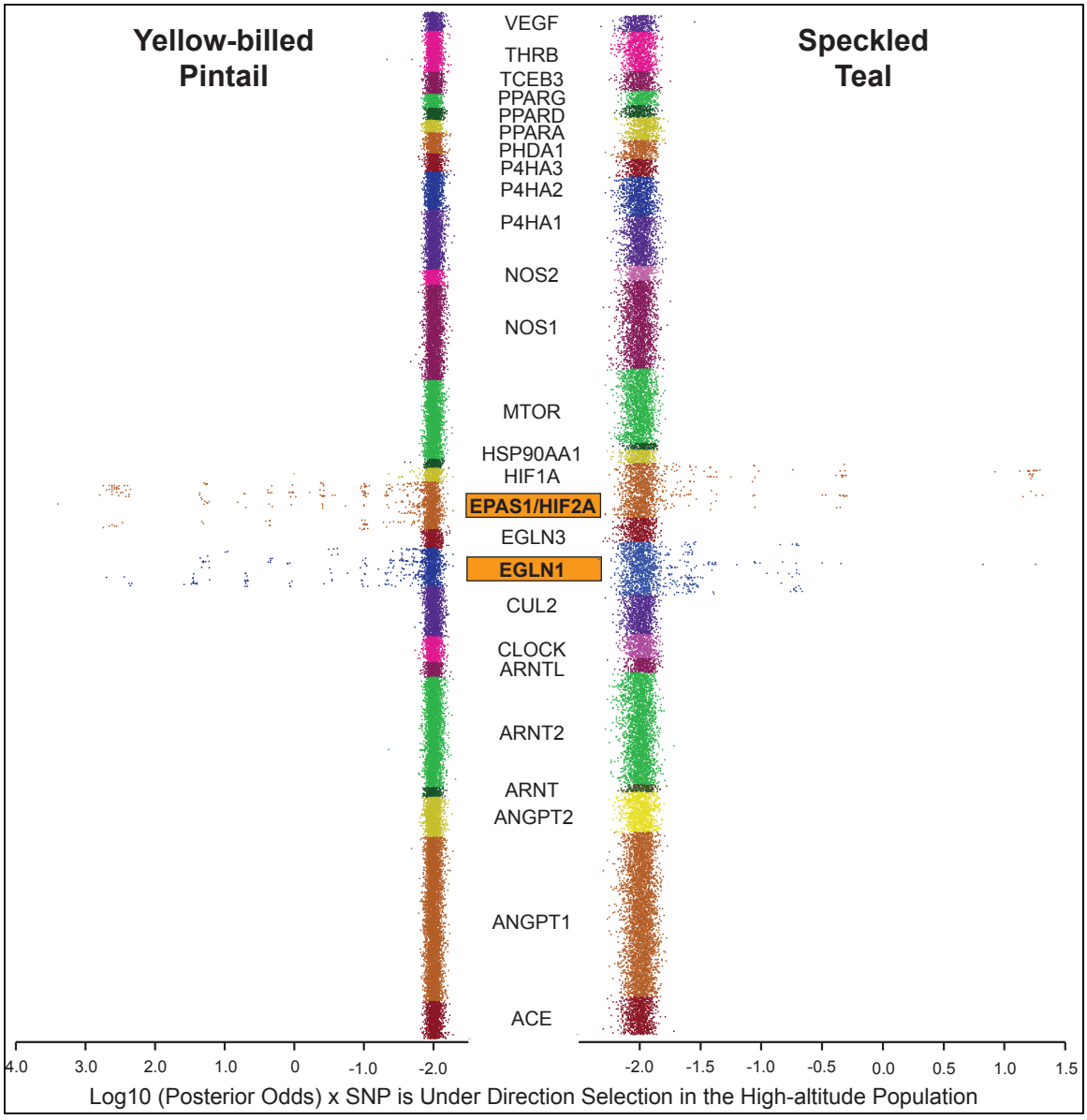

Figure 4.2: Manhattan scatterplots of each of the 26 HIF-pathway gene members listed in alphabetical order, and with their positions in numerical order for both yellow-billed pintail and speckled teal.

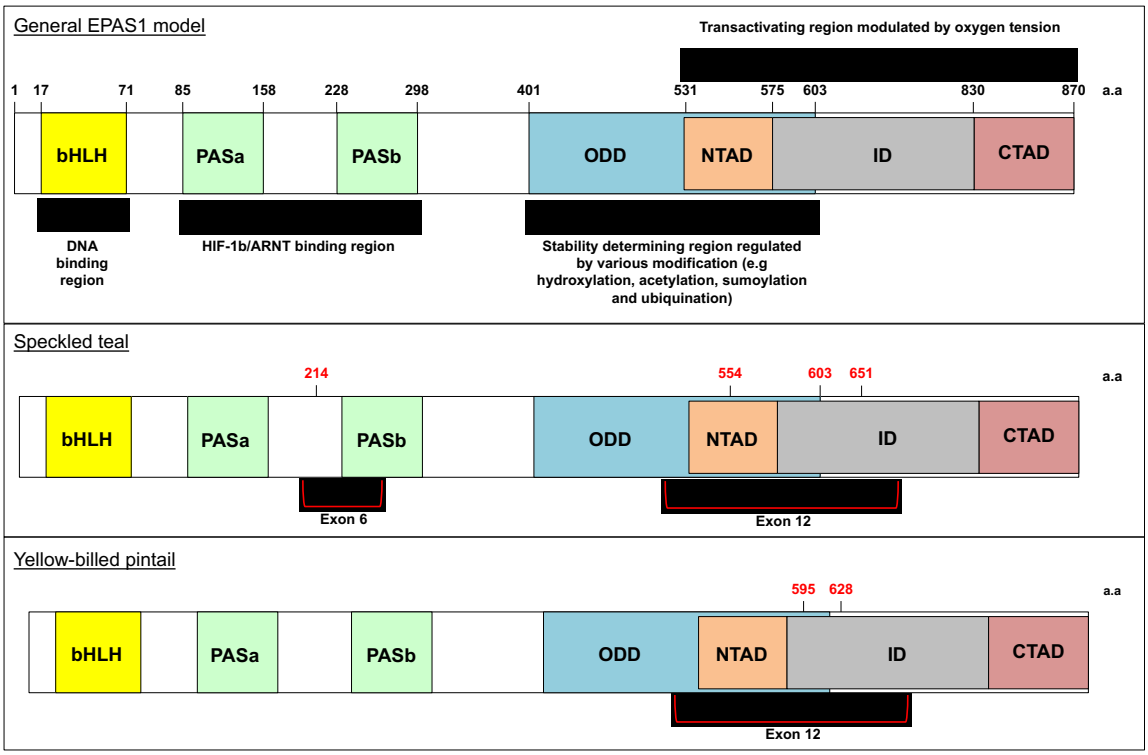

Figure 4.3: General protein model of EPAS1 (HIF2α) Modified from Hong et al. (2004) Dengler et al. (2014) including the canonical domains of the HIF gene family. The variants associated with exonic regions in the gene for both yellow-billed pintail and speckled teal are shown with their associated location/proximity to those domains.

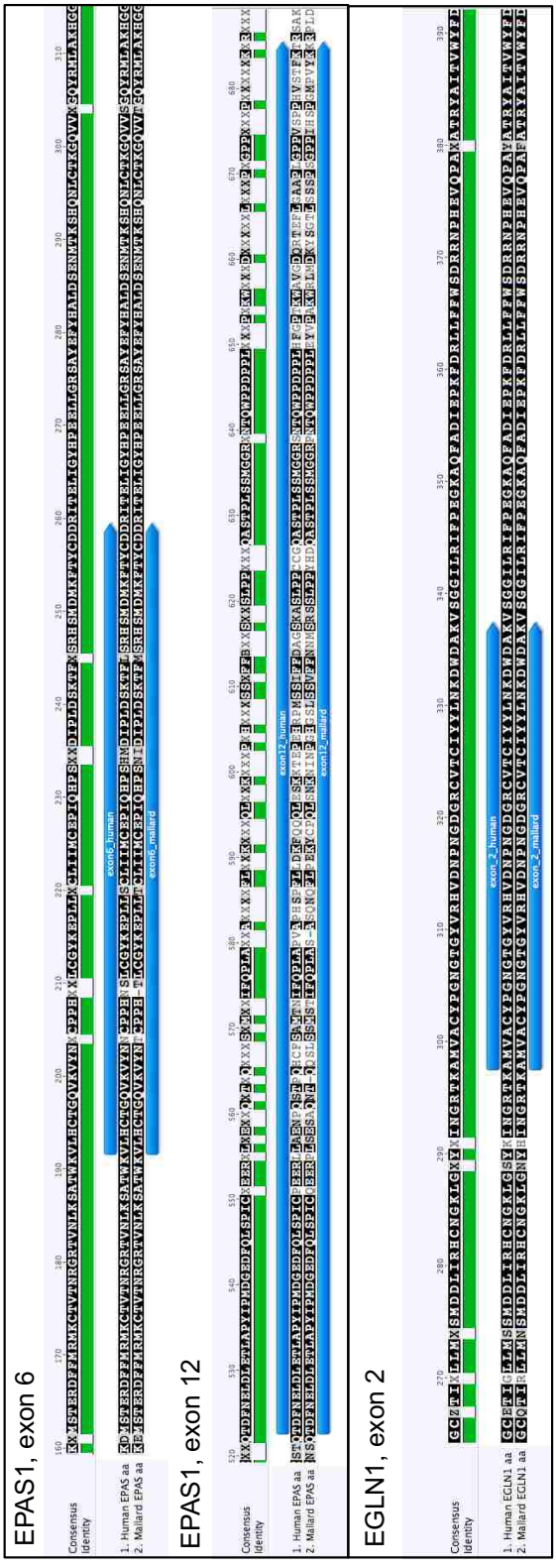

Figure 4.4: Protein alignments of human and mallard reference sequences for exons 6 and 12 in EPAS1, and exon 2 in EGLN1.

Figure 4.5: Nucleotide and protein translation for both speckled teal (above), and yellow-billed pintail (below) in exon 12 of EPAS1, with their corresponding location of significant nucleotide variants labeled in black.

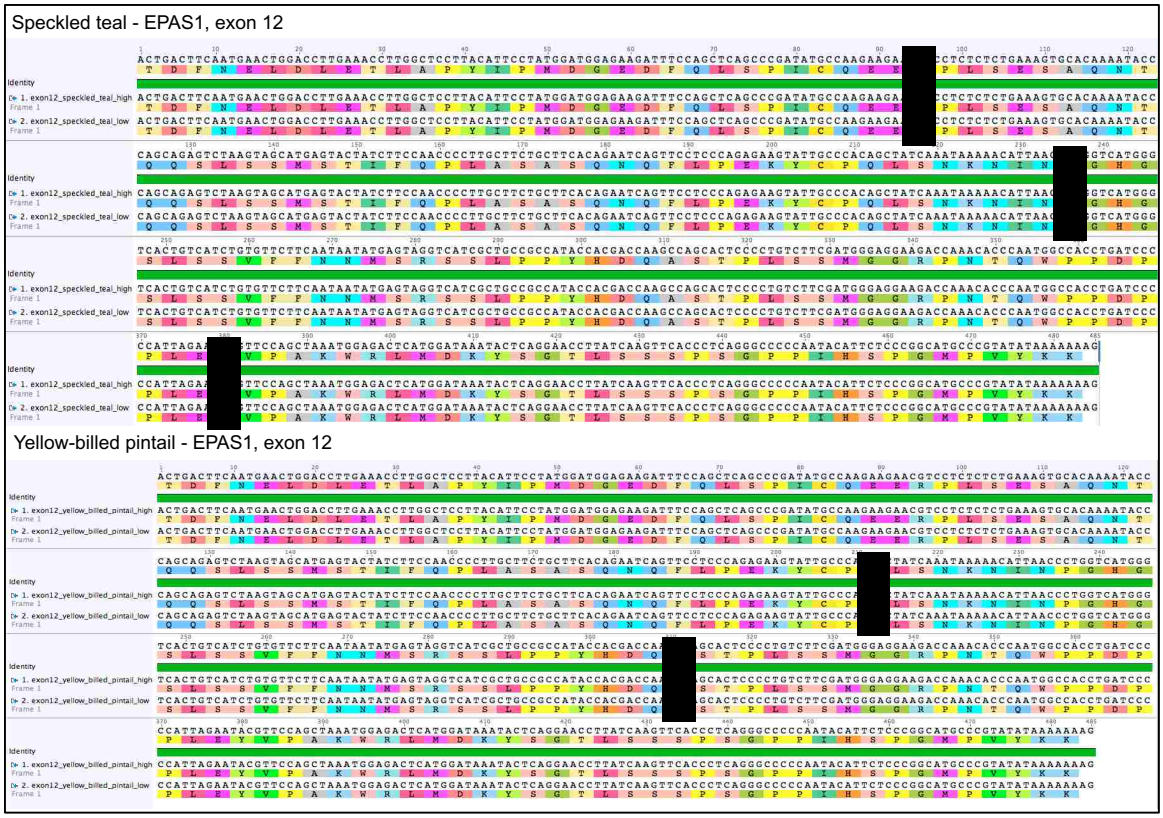

Figure 4.6: Nucleotide alignments and protein alignments for the high- and low-altitude populations of both yellow-billed pintail (above) and speckled teal (below) in exon 12 of EPAS1, with their corresponding location of significant nucleotide variants labeled in black.

Chapter 5: Migration-selection balance drives genetic differentiation in genes associated with high-altitude function in the speckled teal (*Anas flavirostris*)

Background

Heterogeneous landscapes provide venues in which populations experiencing divergent selective pressures can differentiate into locally adapted subpopulations (Nosil et al. 2009). However, the probability for local adaptation and continued divergence depends on the strength of selection balanced by gene flow among populations; i.e., migration-selection balance. Specifically, if the strength of selection is weak relative to gene flow, any local genetic variation will tend to be homogenized among populations. Alternatively, if selection is stronger than the amalgamating force of gene flow, genetic differentiation can accumulate and be maintained (Yeaman and Whitlock 2011). In essence, in the latter scenario, selection against maladapted immigrants limits the effects of gene flow and provides a means for local adaptation (Yeaman and Whitlock 2011; Feder et al. 2012). In general, studies continue to find that much of a genome is free to introgress during early stages of divergence (Via 2012; Via and West 2008; Feder and Nosil 2010; Nosil et al. 2009). For example, specific genetic regions continue to be identified as being important in early stages of divergence and maintaining species boundaries even in the face of gene flow, including sex-linked traits (Martin et al. 2013; Ellegren 2009; Lavretsky et al. 2015; Elgvin et al. 2011; Carling and Brumfield 2008; Storchová et al. 2010), chromosomal inversions or "supergenes"(Kirkpatrick and Barton 2006; Thompson and Jiggins 2014), and a variety of other genes associated with adaptation to extreme environments (Wit and Palumbi 2013; Soria-Carrasco et al. 2014;

DeFaveri et al. 2013; Chapman et al. 2013). Thus, the probability of becoming locally adapted is influenced by this interplay between gene flow and divergent selection (Savolainen et al. 2013; Le Corre and Kremer 2012).

Species that have recently invaded extreme environments are ideal systems to study the effects of selection and gene flow on the adaptation process (Feder *et al.* 2012). Among these, high-altitude habitats offer an excellent opportunity to investigate the genetic basis of local adaptation (Cheviron and Brumfield 2011; Storz et al. 2010) as the characteristics of high-altitude habitats (e.g., low temperatures, ultraviolet radiation, increased water-loss, and low oxygen or hypoxic conditions) are typically debilitating for low-altitude individuals (Hopkins and Powell 2001). Therefore, high-altitude species represent an unparalleled system for studying adaptation in animals, not only because organisms have to overcome a clear challenge to survival and reproduction, but also because the physiological mechanisms of oxygen transport are well understood (Powell 2003; Beall 2001), and selective pressures relatively well defined (Cheviron and Brumfield 2011).

In general, high-altitude species have dynamic respiratory and circulatory systems capable of responding to changes in oxygen (O_2) supply and demand (Beall 2001). Adaptation to hypoxic environments has been shown to be refereed in part through the activation of hypoxia-inducible transcription factors (HIF), which start a signaling cascade of genes involved in biological processes such as angiogenesis and erythropoiesis, and assist in promoting and increasing O_2 delivery to hypoxic tissues (Hoogewijs et al. 2007; Gorr et al. 2006; Rytkönen et al. 2011; Semenza 2011, 2007a). Moreover, for aerobic energy production, genes involved in mitochondrial function and

energy metabolism, O_2 binding and delivery, and hematopoiesis are activated (Hoogewijs et al. 2007; Hopkins and Powell 2001). Ultimately, high-altitude adaptation involves coordinated changes in the expression of many genes that are involved in interacting biochemical pathways (Cheviron and Brumfield 2011; Cork and Purugganan 2004; Cheviron et al. 2012). Moreover, with different biochemical pathways, organisms can mechanistically differ in their response to hypoxic environments (Storz et al. 2010), making it important to understand how organisms adapt in different ways to these environments. At the genomic level, determining the balance between gene flow and selection is especially important when attempting to demarcate genomic regions associated with how populations have adapted to new environments. I expect established high-altitude populations to be genetically differentiated from their low-altitude counterparts at those genes/genetic regions associated with their new environment. Additionally, if gene flow is maintained between two populations, then I expect selection against maladapted genotypes to result in signatures of divergent selection at specific genes, with the remaining genome largely undifferentiated (Via 2012; Feder et al. 2012).

In this study, I tested for genomic differentiation, and specifically for regions putatively under divergent selection between low- (<1,500 m) and high- (up to 5,000m) altitude populations of the speckled teal (*Anas flavirostris*), a widespread South American waterfowl. Although high-altitude speckled teal (*A. f. oxyptera*) are believed to be locally adapted to the Andean environment and largely allopatric from low-altitude birds (*A. f. flavirostris*), speckled teal are known to disperse long-distances, which appears to result in recurrent gene flow events across altitudinal gradients (McCracken et al. 2009a; McCracken et al. 2009b). Previous research suggested a degree of asymmetry in gene

flow in the speckled teal, with more pronounced levels of gene flow going from low to high than high to low (McCracken et al. 2009a). Additionally, genes shown to be under positive selection such as hemoglobin (Natarajan *et al.* 2015) had lower levels of gene flow compared to neutral loci, attributed to countervailing selection at loci associated with high-altitude adaptation (McCracken et al. 2009a; McCracken et al. 2009c). However, studies to date have been restricted in genomic coverage, and thus the extent and influence of gene flow in regards to local, high-altitude adaption in speckled teal remains still unexplored.

Here, I used Restriction Site Associated DNA Sequencing (RAD-Seq) to subsample the genomes of low- and high-altitude speckled teal to examine whether highly differentiated loci are due to selection to high-altitude environments, or are consistent with neutral (allopatric) divergence. In addition, I also sequenced the the α- and β-hemoglobin genes and the mitochondrial (mtDNA) control region. To identify highly differentiated loci, I calculated F_{ST} and performed genomic scans and outlier analyses using Bayesian and other methods at three marker-types with different patterns of inheritance, autosomal, Z- chromosome sex-linked, and mitochondrial. If the two populations are diverged due to selective processes, I expected outliers to be associated with pathways related to hypoxia, and/or those involved with growth and development, metabolism, O_2 transport (i.e., hemoglobin), energy production, and oxidative damage (Cheviron and Brumfield 2011; Storz and Cheviron 2016). Moreover, if there has been local adaptation in the face of continuous gene flow, then I expect a small number of genes/genetic regions to be under strong selection within high-altitude populations, whereas the rest of the genome to be largely undifferentiated, with low levels of allelic

differentiation. Conversely, if high-altitude populations have been largely allopatric during their divergence from lower altitude populations, then I expect to find a larger portion of loci across the genome to be differentiated, primarily through stochastic processes such as genetic drift, characteristic of later stages of divergence (Feder et al. 2012).

Methods

Specimen Collection and DNA extraction

A total of 20 speckled teal were collected from low- (n = 10; elevation range 77-860 meters) and high-altitude (n = 10; elevation range 3,211-4,405) sites (Figure 5.1). Genomic DNA was extracted from muscle or blood using a DNeasy Tissue Kit (Qiagen, Valencia, California, USA) and following manufacturers protocols.

Hemoglobin and Mitochondrial Sequences

For low-altitude speckled teal (*A. f. flavirostris*), the α_A-hemoglobin subunit (HBA), β_A-hemoglobin subunit (HBB), and mitochondrial sequences containing part of the control region and tRNA-Phe gene (McCracken & Wilson 2011) were obtained from Genbank, for the same 20 individuals that were used in this study (McCracken et al. 2009a). For high-altitude speckled teal (*A. f. oxyptera*), the same markers were sequenced, using PCR and DNA sequencing protocols as described in McCracken et al. (2009a). For NCBI accession numbers refer to Table 5.1.

RAD sequencing and Bioinformatics

Genomic DNA was normalized to ~2ng/uL using a Qubit Fluorometer (Invitrogen, Grand Island, New York, USA) and submitted to Floragenex (Eugene, Oregon, USA) for RAD-Seq. In short, genomic DNA was first digested with the 8 base-pair SbfI restriction enzyme (CCTGCAGG) followed by barcode and adapter ligation. Individually barcoded samples were multiplexed and sequenced on the Illumina HiSeq 2000 with single-end 100 bp sequencing chemistry. Following the run, RAD sequences were de-multiplexed and trimmed to yield resulting RAD sequences of 90 bp. The methods used by Floragenex are described by (Hohenlohe et al. 2010; Miller et al. 2007; Baird et al. 2008), and are summarized below.

Genotypes at each variable site were determined using Floragenex's VCF_popgen v.4.0 pipeline to generate a customized VCF 4.1 (variant call format) database with parameters set as follows: minimum AF for genotyping = 0.075, minimum Phred score = 15, minimum depth of sequencing coverage = 10x, and allowing missing genotypes from up to 2 out of 20 individuals (10%) at each site. To filter out base calls that were not useful due to low quality scores or insufficient coverage, genotypes at each nucleotide site were inferred using the Bayesian maximum likelihood algorithm. The genotyping algorithm incorporates the site-specific sequencing error rate, and assigns the most likely diploid genotype to each site using a likelihood ratio test and significance level of p = 0.05. Sites for which the test statistic is not significant are not assigned a genotype for that base in that individual, effectively removing data from sites for which there are too few high-quality sequencing reads. The sequencing coverage and quality scores were

then summarized using the software VCFtools v.0.1.11 (Danecek et al. 2011). Custom perl scripts written by M. Campbell (University of Alaska Fairbanks) were used to retain bi-allelic sites only and converted to VCF file format.

Population Structure

Nucleotide diversity (π) and pairwise F_{ST} (Weir and Cockerham (1984) for each SNP between low- and high-altitude populations for RAD-seq, hemoglobin, and the mtDNA were calculated in Arlequin v. 3.5 (Excoffier and Lischer 2010).

A haplotype network was used to visualize structure at mtDNA using TCS (Clement et al. 2000) and as implemented and visualized in the program Population Analysis with Reticulate Trees (PopART) (software available at: http://popart.otago.ac.nz). Analyses were done under default settings.

For nuclear markers, population structure was first visualized using a principal component analysis (PCA) as implemented in the software package PLINK v1.9 (Purcell et al. 2007). The PCA used in PLINK uses a two dimension reduction routine based on the variance-standardization relationship matrix. The top principle components are used as co-variants in association analysis regressions to help correct for population stratification, while multi-dimensional scaling (MDS) coordinates helps visualize genetic distances (Compagnon and Green 1994).

Next, assignment probabilities were calculated using ADMIXTURE (Alexander and Lange 2011; Alexander et al. 2009). ADMIXTURE assumes a specific number of hypothetical populations (*K*) and provides a maximum likelihood estimate (i.e., Q estimate) of allele frequencies for each population and admixture proportion for each

individual. I analyzed values of K = 1 to 5 using the block-relaxation method algorithm for point estimation and terminating them when the estimates increased by <0.0001.

Identifying Outlier Loci: Hemoglobin Sequences and RAD-Seq Data

A genomic scan was performed by obtaining pairwise F_{ST} estimates calculated in Arlequin v. 3.5 (Excoffier and Lischer 2010) between low- and high-altitude populations of each species. Instances of speciation-with-gene-flow are typically characterized by a vast majority of the genome being homogenized via gene-flow with low F_{ST}, while a few regions containing genes under strong divergent selection have restricted gene-flow and high F_{ST}. Therefore, empirical examples are expected to produce a characteristic L-shaped distribution of differentiation across loci in the genome, with most loci having low F_{ST} values. Whereas, populations experiencing allopatric speciation with established reproductive isolation are characterized by less gene-flow and greater divergence across much more of the genome; characterized by a distribution with less skew, more density in the center, and a more pronounced tail of high F_{ST} values (Feder et al. 2012; Savolainen et al. 2006; Via 2001)

Next, I tested for signatures of selection using two programs, which minimized the probability of detecting false positives. First, Lositan, which implements the FDIST2 function, was used to demarcate markers putatively under positive selection (Antao et al. 2008; Beaumont and Nichols 1996): analyses were based on the Infinite Alleles Model with 50,000 simulations, a confidence interval of 0.95 and a false-discovery rate (FDR) of 0.01, as well as using the neutral mean F_{ST} and forcing mean F_{ST} options. In addition, I used a Bayesian approach as implemented in BayeScan v. 2.1 to again identify putatively

selected loci, using default parameters suggested for large datasets (prods = 100). BayeScan uses a logistic regression model to separate locus-specific effects of selection from demographic differences (Foll and Gaggiotti 2008). Previous tests have shown that outliers detected by BayeScan are likely to be "better" candidate regions of the genome, because of its more conservative approach (Pérez‐Figueroa et al. 2010). Foll (2012) proposed a logarithmic scale for model choice defined as: >3 substantial ($\log_{10}$PO > 0.5); > 10 strong ($\log_{10}$PO > 1.0); >32 very strong ($\log_{10}$PO > 1.5); and >100 decisive evidence for accepting a model ($\log_{10}$PO > 2.0). In the genome scans, a threshold for PO > 0.5 (substantial) was used for a marker to be considered under selection. Thus, top outlier loci were considered to be those that were identified by both methods, as previously stated. General outliers were considered those that were identified by both methods, including markers which were significant in Lositan analyses, and were in the top1% of Log10(PO) values in BayeScan; this data was used for gene-flow analyses (see below).

Any markers identified as putatively being under divergent selection by both Lositan and BayeScan were next subjected to a BLAST search (database – nr, e-value < e−10, annotation cut-off > 50) in Blast2GO (Conesa et al. 2005) using the taxonomy filter for birds (taxa: 8782, Aves) to determine gene identity. Additionally, they were categorized by chromosomal location, either as autosomal or Z-linked based on a BLAST search to the Chicken genome (*Gallus gallus* 5.0 reference) through NCBI, using stringent criteria (blastn, >75% identity, e-value < 10^{-3}, max-score >40).

Gene Flow (∂a∂i) and TMRCA (IM) estimates

Previous multi-locus work in this species suggested asymmetric gene flow from low- to high-altitude populations (McCracken et al. 2009a). I tested for and estimated rates, as well as the directionality of gene flow, with the program *∂a∂i* (Gutenkunst et al. 2009, 2010), which implements a diffusion-based approach to test against specified evolutionary models with the best-fit model determined using a log likelihood-based multinomial approach. *∂a∂i* analyses were done with RAD-seq data, which were ultimately parsed into three categories that were analyzed separately, [1] All SNPs, [2] outliers-only as determined by Lositan/BayeScan analyses (see above), and [3] non-outliers only. Given an interest in identifying differences between migration rates between outlier and nonoutlier loci, autosomal and Z-chromosome linked SNPs were analyzed together. To estimate bi-directional migration rates, the data were tested against the Isolation-with-Migration evolutionary model. Parameters in *∂a∂i* were optimized prior to calculating the likelihood, which is the product of Poisson likelihoods for each parameter given an expected value from the model frequency spectrum.

Results

A total of 159 million RAD-Seq sequence reads were obtained with an average depth of 38.6 (±8.0 SD) million reads per sample. After filtering, an average of 19,434 RAD clusters were recovered per sample, which represents ~1% of a typical 1.1 gigabase avian genome (Zhang et al. 2014). Finally, a total of 47,731 polymorphic SNPs were identified. Overall, F_{ST} estimates between low- and high-altitude speckled teal

populations were identical for Lositan (F_{ST} = 0.065) and Arlequin (F_{ST} = 0.065). The mtDNA with F_{ST} = 0.77 exhibited a 14-fold difference in population differentiation compared to nuclear DNA (Table 5.2).

Population Structure, and Outlier Detection

For nuclear sequences, ADMIXTURE results corresponded with PCA analyses in which low- and high-altitude individuals were highly differentiated into two clusters with 99% assignment probabilities (Figure 5.2b-c), despite overall low F_{ST}. The mitochondrial sequences corroborate the nuclear data showing two distinct populations, however, suggesting a much deeper level of divergence with F_{ST} = 0.77 (Figure 5.2a).

BayeScan identified ~1% of markers (457 SNPs associated with 420 RAD clusters), as compared to Lositan's ~6% (3,008 SNPs, associated with 2,664 RAD clusters) as being putatively under positive selection (Figure 5.3). Thus, the RAD-Seq data show a F_{ST} distribution with a pronounced right-tailed L-shape, with most variants showing little-to-no population differentiation and a smaller percentage of the genome comprised of highly differentiated loci (Figure 4). Loci were classified as outliers if they were identified by both methods (Lositan and BayeScan) as being significant. Ultimately, although 356 RAD-seq loci overlapped both programs, only 34 of them returned a significant BLAST hit, and 29 of those were to a gene or transcript with an identifier. For the RAD-Seq data, of the top hits with the greatest prior odds were TOPORS, and NNT, whereas F8, IGF-1, and BMP-2 had slightly lower prior odds (Table 5.2).

Outlier detection within the hemoglobin sequences was performed by calculating pairwise F_{ST} for each SNP via a locus-by-locus AMOVA. This identified four SNPs on

the α-chain subunit and four SNPs on the β-chain of the major hemoglobin. All four of these variants have previously been identified as having a putatively beneficial role in hemoglobin function via increased O_2-binding affinity: α77 (pos 379, F_{ST} = 0.95), β 116 (pos 346, F_{ST} = 1.00), β 133 (pos 397, F_{ST} = 1.00), and β13 (pos 37, F_{ST} = 0.42) (Natarajan et al. 2015).

Chromosomal Location

Of the 19,434 RAD clusters, 889 clusters were identified as putatively Z-linked (i.e. having a significant BLAST hit to the Chicken Z-chromosome); therefore, the rest of the 18,549 RAD-seq clusters were classified as putatively autosomal, although there is no way of knowing for sure given the short length of the RAD clusters (ie. 100 bp). Compared to the autosomal clusters, the Z-linked clusters had higher F_{ST} by a factor of two; Z-linked F_{ST} = 0.123 (1,797 SNPs from 889 seqs) and autosomal F_{ST} = 0.061 (45,944 SNPs from 18,549 seqs;).This Z:Autosomal ratio of 2.01 is higher than expected under a neutral model of evolution (Whitlock and McCauley 1999).

The chromosomal locations of the outliers were also identified: of the total outliers (356), 1.67% were mapped to the autosome (310/18,549), compared to 5.17% outliers on the Z-chromosome (46/889), approximately a 3-fold difference. The chicken Z-chromosome is 82,363,669 bp representing ~7.68% of the total chicken genome (1,072,544,763 bp*, Gallus gallus*, NCBI Annotation Release 103). Thus, given what we know about the size of the chicken genome, out of the 356 the overlapping outliers, I classified 49 as belonging on the Z-chromosome (12.9%), which is a significantly larger

portion of outliers than expected by chance ($X^2 = 5.5102$, p-value = 0.019). The two top hits in the outlier analyses were located on the Z-chromosome (NNT, TOPORS).

Gene Flow and TMRCA Estimates

Although low- and high-altitude populations are differentiated, *∂a∂i* results under an isolation-with-migration model supports extensive gene flow between populations, specifically asymmetric gene flow from low to high altitude (Table 5.3), corroborating the previous coalescent study (McCracken et al. 2009a). These asymmetric gene flow estimates correspond with the same directionality obtained from mtDNA in which 10% of high-altitude individuals were recovered with low-altitude haplotypes, whereas no low-altitude individuals were recovered with high-altitude haplotypes (Fig. 5.2a). It is important to note that all of the introgressed mitochondrial haplotype were collected from the far north-west Argentina, which is close to the boundary between low- and high-altitude subspecies distributions.

The analyses also suggest a substantial reduction in gene flow in outlier loci compared to non-outlier loci, showing a two-fold decrease from high- to low-altitude estimates (0.525 → 0.234 migrants per generation), and a four-fold decrease from low- to high-altitude (7.76→1.97 migrants per generation; Table 5.4). These outlier regions may represent those portions of the genome, which are under selective constraint due to their being associated with adaptations to O_2 deprivation and/or cold temperatures at high-altitude.

Discussion

Our data reveal that despite low levels of genomic divergence a (F_{ST} = 0.065) and high gene flow, the low- and high-altitude populations of speckled teal are genetically distinct. Evidence of this can be seen in the population structure associated with both mitochondrial and nuclear markers (Table 1; Figure 2a-c). This pattern is due, in part, to a few regions (i.e., outliers) associated with genes known to be involved in responses to ROS production/oxidative damage (TOPORS, NNT), response to hypoxia through the HIF pathway (TOPORS, IGF, BMP), and response to hypoxia through heme-factors/blood (F8, HBA, HBB, BMP-2). These genes were found to involve autosomal (IGF, BMP, HBA, HBB, F8), Z-linked (NNT, TOPORS) and mitochondrial elements (NNT = nuclear encoded, mitochondrial embedded) (Figure 4, Table 2). These outliers also showed depressed gene-flow between the populations, compared to the rest of the genome, suggesting an important role for migration-selection balance in the evolution of these loci in these two populations (Table 5.3). Lastly, the results also hint a dominant role for genes the Z-chromosome (Figure 5.4, Table 5.2).

Migration-selection balance explains high-altitude adaptation in Speckled Teal

The interplay between selection and gene flow is critically important for population divergence in the presence of gene flow (Feder et al. 2012). Under such a scenario, the magnitude of selection dictates whether populations continue to diverge or evolve as a single group. Here, we identified a relatively small portion of markers to be under positive/divergent selection with restricted levels of gene flow between low- and high-altitude speckled teal populations.

Under neutrality, where genetic drift acts on markers with different effective population sizes (N_e) (Whitlock and McCauley 1999), there are different expectations for F_{ST} estimates associated with certain marker types (i.e., mitochondrial, nuclear, Z chromosome-autosome). The expectations associated with the effective population sizes of these markers are useful in illuminating whether selection has shaped the evolution through certain pressures on specific chromosome types (Charlesworth 2009). Previous work on waterfowl support that ducks fit the ¼ rule for mtDNA and ¾ rule for Z-linked markers (Lavretsky et al. 2015; Lavretsky et al. 2016). Specifically, under a scenario of equal reproductive success between males and females, the mitochondria will have ¼ the effective population size of the nuclear genome, resulting in a four-fold difference in F_{ST} between mtDNA and nuDNA (Caballero 1995; Whitlock and McCauley 1999; Dean et al. 2015). There is a similar expectation for Z chromosome-autosomal comparisons. Under equal reproductive success between males and females, the Z chromosome will have ¾ the effective population size of autosomes, resulting in an expected Z:Autosomal F_{ST} ratio of ≤1.33 (Caballero 1995; Whitlock and McCauley 1999; Dean et al. 2015).

However, the results for both sets of F_{ST} ratios show significant deviations from neutrality, with mtDNA:nuDNA resulting in a 14-fold difference, and a 2-fold difference for Z:Autosomal (Table 5.2), suggesting a substantial role for evolutionary forces such as selection, drift, or gene flow. There are several factors that may partly explain these patterns. First, the observed difference is most extreme for the mtDNA, which could be because of the faster sorting rate of mtDNA allowing it to accumulate differences faster than nuDNA (Zink and Barrowclough 2008), or because many waterfowl including speckled teal have very large population sizes (Wilson et al. 2011; Frankham 2007),

although this explanation is less likely. Second, it is more likely that male-biased dispersal and female philopatry common in ducks underlies the substantial structuring of mtDNA in the speckled teal (Flint et al. 2009; Peters et al. 2007; Avise et al. 1992; Zink and Barrowclough 2008; Peters et al. 2012). Lastly, the extreme divergence in the mitochondria, despite similar estimates of gene flow, may suggest that this pattern has resulted from selective processes, since the mitochondrial genome is frequently a source for selective sweeps and strong purifying selection on genetic variation associated with adaptations to hypoxic high-altitude environments (Li et al. 2013b; Zhang et al. 2013; Yu et al. 2011; Xu et al. 2007; Scott et al. 2011). However, it is difficult to dismiss this possibility without more of the mitochondrial genome to assess.

The Z: Autosome ratio also is larger (2.01) than predicted under neutrality (1.33), although not as large as the mtDNA:nuDNA ratio, suggesting a more subtle role for selective pressures for smaller portions of the genome. For example, significant gene flow, predominantly from low- to high-altitude, was detected by *∂a∂i* analyses regardless of marker set; however, there was a substantial decrease in gene flow across outlier loci as compared to non-outliers (Table 5.4). Similar decreases in gene flow have been detected in Hb sequences compared to other nuclear loci in speckled teal and other Andean waterfowl (McCracken et al. 2009a). We also found that significant outliers were on the Z-chromosome than expected; such loci have been implicated in population divergence and speciation in many bird species (Sæther et al. 2007; Carling and Brumfield 2009; Pryke 2010; Ellegren et al. 2012; Stölting et al. 2013; Elgvin et al. 2011; Storchová et al. 2010; Lavretsky et al. 2015). A higher density of Z-linked genes in these types of analysis is thought to be the result of a higher rate of evolution relative to

autosomes (Mank et al. 2007; Ellegren 2009), as well as through its role in sexual selection, and reproductive isolation (Ritchie 2007). However, it could also be because of a general reduction in recombination rates across the Z chromosome, which can cause blocks of linkage disequilibrium, serving to offset the disrupting force of gene flow (Qvarnström and Bailey 2009; Sæther et al. 2007).

Given that the two speckled teal populations are phenotypically distinct in both plumage and body size, it seems likely that the outlier analyses have pinpointed genes underlying mechanisms involved in speciation through adaptive divergence. The two genes with the highest divergence were Z-linked, and potentially have a direct link to adaptations to high-altitude functions (see next section). Two possibilities are that [1] hypoxia and the high-altitude environment are driving divergence directly in these genes or closely linked genes, or [2] that reproductive isolation is driving phenotypic divergence and these outliers are effectively hitchhiking on that selective pressure on the sex-chromosomes. However, with the current data, I was not able to clarify further which of the two possibilities is responsible, and thus potentially warrants future research. Nonetheless, in the case of the speckled teal, a significant role for non-neutral processes between Z-linked, and Autosomal markers is shown - a result consistent with recent findings in other birds (Dhami et al. 2016; Dean et al. 2015) as well as other duck species (Lavretsky et al. 2015; Lavretsky et al. 2016).

Overall, the results suggest selection is acting to prevent admixture at loci of adaptive importance, likely through selection against alleles originated from maladapted migrants (i.e. low-altitude individuals). This appears to be occurring despite these speckled teal populations being relatively divergent at an estimated between 0.5 - 1.0

mya, with gene flow likely playing a recurrent role throughout their recent history. Thus, at the genomic level, these populations do not seem to be suffering from an erosion of locally adaptive alleles through swamping. Ultimately, this matches predictions of divergent selection generating extrinsic reproductive isolation (Yeaman & Whitlock 2011; Feder et al. 2012), due to poor reproductive success of low-altitude migrants at high altitude compared to high-altitude residents with high-altitude genotypes. There are extra-limital examples of low-altitude species (e.g., *A. discors*, *A. bahamensis*) occasionally observed at high altitude (Monge and Leon-Velarde 1991); however, low egg hatchability due to embryonic susceptibility to hypoxia appears to present a serious selective pressure potentially driving the differences in reproductive success, that may ultimately limit a species' ability to colonize the high-altitude environment (Visschedijk et al. 1980; Monge and Leon-Velarde 1991; Leon-Velarde et al. 1984; Carey et al. 1994). Therefore, the results support strong roles for both selection and gene flow in shaping the genomic architecture of speckled teal populations in response to adaptations to the high-altitude Andean environment.

High altitude adaptation and Z-linked Genes: TOPORS and NNT

It is well known that exposure to low O_2 leads to a cascade of metabolic and physiological changes. At high altitudes, oxygen utilization has crucial consequences for cellular function, especially through the production of reactive oxygen species (ROS) typically produced through inefficiencies during mitochondrial respiration (Turrens 2003; Palmer and Clegg 2014). The imbalance of ROS and antioxidant capacity is driven by hypoxic stress and increases at high altitudes; therefore, mechanisms to protect against

oxidative damage or reduce ROS production would be advantageous (Storz et al. 2010). The data present two gene regions with functions associated with a response to oxidative damage: TOPORS and NNT.

The region that I identified as having the highest level of divergence between the two populations was TOPORS (F_{ST} = 1.0), which is an E3 ubiquitin-protein ligase. TOPORS is known regulate, p53 (Rajendra et al. 2004; Weger et al. 2005) and DNA Topoisomerase I (Haluska et al. 1999; Weger et al. 2005). Both of these genes/pathways are utilized as part of an intracellular non-immune surveillance mechanism that controls cellular response to a variety of stress conditions, including DNA damage and hypoxia, among others, leading to cell growth arrest and apoptosis (Sermeus and Michiels 2011; Levine 1997; Bertozzi et al. 2014). Although TOPORS has not previously been identified as candidate gene involved in adaptation to high-altitude environments, p53 has been implicated in response in low-oxygen O_2 including, hypoxic microenvironments created by tumors (Royds et al. 1998), underground burrows of rodents (Ashur-Fabian et al. 2004; Quina et al. 2015) and even high-altitude environments (Zhao et al. 2013b). In addition to roles previously stated, TOPORS has also been implicated in playing a key role in regulating primary cilia-dependent development and function in the retina (Chakarova et al. 2011), potentially in response to increased levels of ultraviolet radiation exposure, as well as oxidative stress, at higher altitudes (Paul and Gwynn-Jones 2003). Therefore, the role of TOPORS in hypoxia and DNA damage presents a potentially pleiotropic response to both low oxygen and DNA repair from oxidative damage.

The other gene of interest is the nuclear-encoded, but mitochondrially-embedded NAD (P) transhydrogenase (NNT), which is integral for oxidative phosphorylation

(OXPHOS). Under normal conditions, NNT uses energy from the mitochondrial proton gradient to produce high concentrations of NADPH; the resulting NADPH is used for biosynthesis, as well as in reactions inside the mitochondria required to remove reactive oxygen species (Figueira et al. 2013). To date, this is the first time NNT has been identified as an outlier/gene of interest in relation to high-altitude adaptation, or low-oxygen environments. Previously, mutations in NNT have been shown to increase the amount of oxidative damage due to its inability to regulate ROS within the mitochondria (Huang et al. 2006; Freeman et al. 2006). Although, other mechanisms involved in DNA damage have been implicated in other high-altitude species (Zhang et al. 2016; Yang et al. 2015; Qiao et al. 2016; Subramani et al. 2015), it has been in response to ultraviolet radiation.

Autosomal genes associated with high-altitude adaptation: IGF, BMP, and HBA and HBB

One of the biggest selective pressures for species invading high-altitude environments is the low levels of O_2 in the atmosphere causing hypoxia (Storz et al. 2010). Hypoxia triggers a conservation of energy through a global reduction in protein expression, as well as a switch to anaerobic metabolism via a family of transcription factors involved in the hypoxia-inducible factor (HIF) pathway (Semenza 2007a). This pathway is widely considered the "master regulator" of oxygen sensing, and because of its central role in mediating a system-wide response to low O_2, is frequently found to be of importance in high-altitude species (Beall et al. 2010; Hanaoka et al. 2012; Bigham et al. 2013; Wang et al. 2014; Li et al. 2014; Bigham et al. 2009).

Although this study did not identify members of the HIF pathway specifically, I did find that the insulin-like signaling pathway (IGF-1, IGFBP-5, MAPK) and skeletal morphogenesis (BMP2) are likely under positive selection between these two populations. Not only have both the IGF and BMP pathways been shown to interact with each other (Nakae et al. 2001; Guntur and Rosen 2013), but they are also involved in the system-wide response to hypoxia mediated via the HIF-pathway (Wang et al. 2007; Feldser et al. 1999; Fukuda et al. 2002). Both IGF signaling (Berg et al. 2015; Li et al. 2013a; Welch et al. 2014) and bone morphogenesis (Qu et al. 2013; Yang et al. 2016) have been implicated in the acclimation to various environmental changes, including adaptations to low O_2. Additionally, both the IGF and BMP pathways are known to interact with the p53 circuit (Harris and Levine 2005), suggesting the potential for multiple outliers in interconnected pathways having an effect on high-altitude adaptation in the speckled teal.

Hemoglobin has frequently been of interest in relationship to adaptive responses to low oxygen (Cheviron et al. 2014; Tufts et al. 2014; Projecto-Garcia et al. 2013; Revsbech et al. 2013; Grispo et al. 2012; Beall et al. 2010; Storz et al. 2009; Storz and Moriyama 2008; Weber 2007), and has been shown to be under selection in a variety of organisms, including Andean ducks (McCracken et al. 2010; McCracken et al. 2009a; McCracken et al. 2009b; McCracken et al. 2009c). This study was able to identify eight significant SNP variants between low- and high-altitude populations in the HBA and HBB subunits; of those genotypic variants identified, four have been previously identified (α^A77, β^A13, β^A116, and β^A133) as causing an increase in the HbA isoform's O_2-binding affinity (McCracken et al. 2009a; McCracken et al. 2009c), and functional

tests of variants have recently shown that β^A116-Ser and β^A133-Met were responsible for increased O_2-binding affinity for high-altitude speckled teal (Natarajan et al. 2015).

In addition to selective pressures on the hemoglobin subunits, another candidate for adaptation to high altitude is the blood coagulation factor VIII or F8, as found in the outlier analyses. This gene encodes for a large plasma glycoprotein, most notably responsible for hemophilia in humans (Gouw et al. 2012). However, it has not been identified as a candidate in other high-altitude organism studies, even though it is well established that plasma concentration of F8, and other clotting factors, are elevated in humans encountering hypoxic situations (Chohan 2014; O'Brodovich et al. 1984).

Finally, it is important to note that the patterns seen in autosomal loci might also result through genetic hitchhiking, or genetic draft, arising from linkage to genes or regions on the same chromosome that are the targets of selection. However, in birds, linkage disequilibrium (LD) decays quickly even over short distances, reflecting high recombination rates (Stapley et al. 2010; Wong et al. 2004; Balakrishnan and Edwards 2009; Backström et al. 2006); therefore, this possibility seems less likely.

Conclusions

The results suggest that adaptations to their high-altitude environment are resulting in genomic divergence, despite longstanding and recurrent gene flow between populations of speckled teal. This has created a genome wide pattern in which migration-selection balance is prevalent across various portions of the genome. I have identified a set of loci putatively under selection with allele frequencies strongly correlated with high- and low-altitude habitats – most notably those involved in the insulin-like signaling

pathway, bone morphogenesis, metabolic processes through the mitochondria (oxidative phosphorylation), responders to hypoxia-induced DNA damage, and feedback loops to the HIF pathway. Although gene flow was found on all linkage groups (mitochondrial, autosomal, and Z chromosome), many outlier loci in the nuclear genome were found to have depressed gene flow estimates, compared to other loci. I also discussed Z-linked loci and their role in the population differentiation of incipient diverging species; the data suggest that Z-linked loci may be simultaneously under selection due to their mechanistic role in high-altitude adaptation as well as phenotypic divergence. However, I note that the dataset represents a small percentage (~1%) of the speckled teal genome, and that there are likely many other candidate gene regions for high-altitude adaptation for this species, including genes involved in the HIF pathway, as well as mitochondrial elements (Storz et al. 2010; Scheinfeldt and Tishkoff 2010; Beall et al. 2012).

Together the data identify a proportion of the genome, known to be linked to high-altitude adaptions, that are likely under positive directional selection in the high-altitude speckled teal population. Overall, the results suggest a multi-factorial response to life at high altitudes through an array of interconnected pathways, that are not only under positive selection but, whose genetic components seem to be providing at least a partial genomic barrier to to gene flow and continued interbreeding, functioning as an avenue for population divergence and speciation, even if the speciation process has stalled short of completion (e.g., Peters et al. 2012). Ultimately, this study illustrates another example of how populations are able to invade novel, and sometimes adapt to extreme habitats, and provides the most comprehensive genomic study of this Andean species.

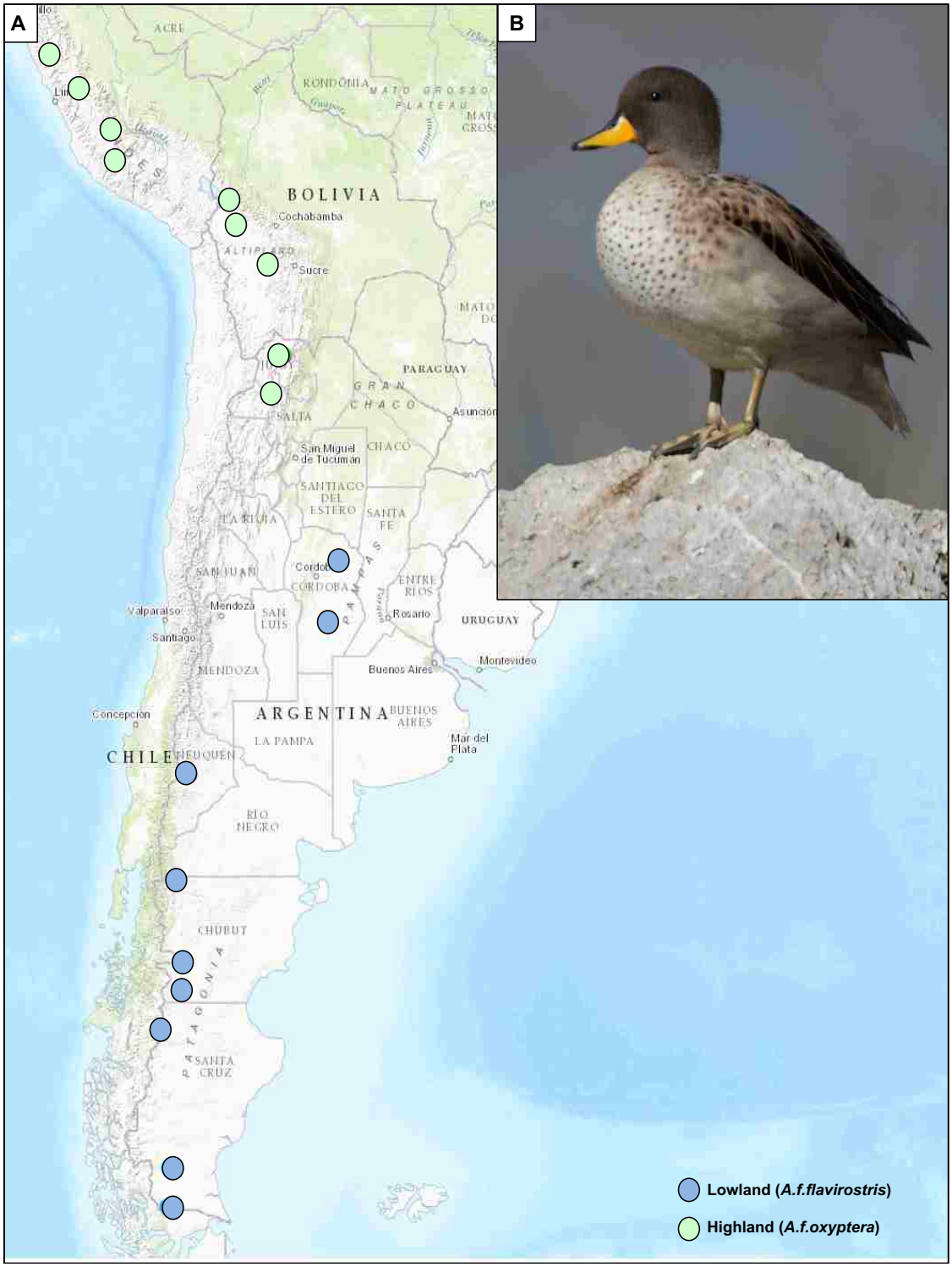

Figure 5.1: Specimen collection locations of the speckled teal, *A. flavirostris* (A) 10 individuals collected from high-altitude populations, *A. f. oxyptera*, shown in gray, and 10 individuals from low-altitude populations, *A. f. flavirostris*, shown in white (B) Representative photograph.

Table 5.1: NCBI accession number information for alpha- and beta-globin samples from previously published samples of *A.f.flavirostris*

Population	Specimen ID	mtDNA	mtDNA PopSet	alpha	alpha PopSet	beta	beta PopSet
A.f.flavirostris	KGM 267	JN223309	343479578	GQ271329	254768871	GQ272067	254770347
A.f.flavirostris	KGM 275	JN223310	343479578	GQ271330	254768871	GQ272068	254770347
A.f.flavirostris	KGM 285	JN223312	343479578	GQ271332	254768871	GQ272070	254770347
A.f.flavirostris	KGM 319	JN223313	343479578	GQ271333	254768871	GQ272071	254770347
A.f.flavirostris	KGM 699	JN223339	343479578	GQ271359	254768871	GQ272097	254770347
A.f.flavirostris	KGM 727	JN223345	343479578	GQ271365	254768871	GQ272103	254770347
A.f.flavirostris	KGM 735	JN223346	343479578	GQ271366	254768871	GQ272104	254770347
A.f.flavirostris	KGM 747	JN223348	343479578	GQ271368	254768871	GQ272106	254770347
A.f.flavirostris	KGM 778	JN223352	343479578	GQ271372	254768871	GQ272109	254770347
A.f.flavirostris	KGM 790	JN223357	343479578	GQ271377	254768871	GQ272114	254770347

Table 5.2: Divergence measurements (F_{ST}) associated with different subsets of markers, and their chromosomal location

Chromosomal Location	Marker Type	F_{st}
Mitochondria	-	0.770
All (RAD-seq)	All	0.065
Z-linked	All	0.123
Z-linked	Outlier	0.577
Z-linked	Non-outlier	0.064
Autosomal	All	0.061
Autosomal	Outlier	0.442
Autosomal	Non-outlier	0.049

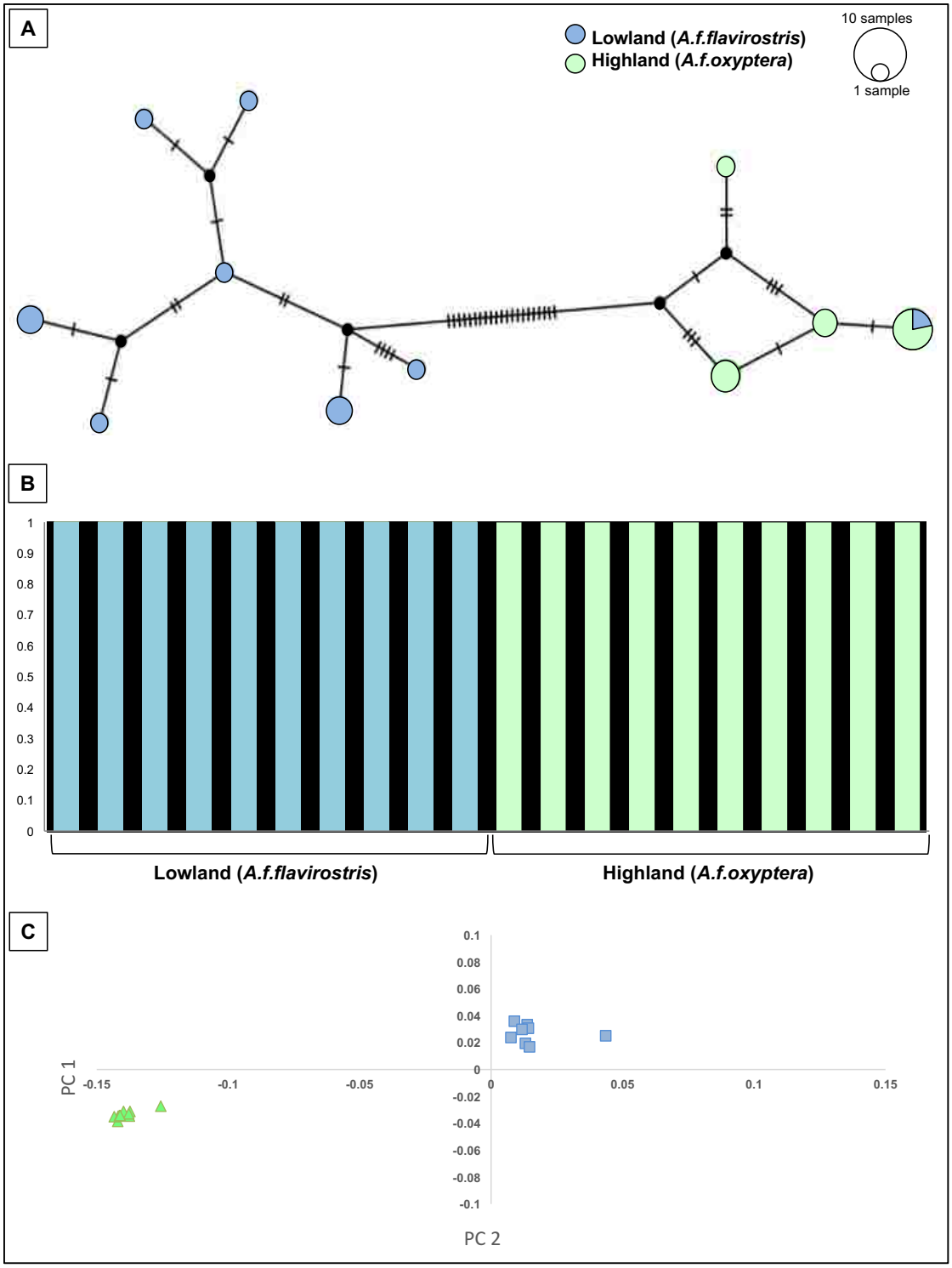

Figure 5.2: Population structure between the high- and low-altitude speckled teal populations (A) mitochondrial haplotype (B) ADMIXTURE results (C) principle component analysis.

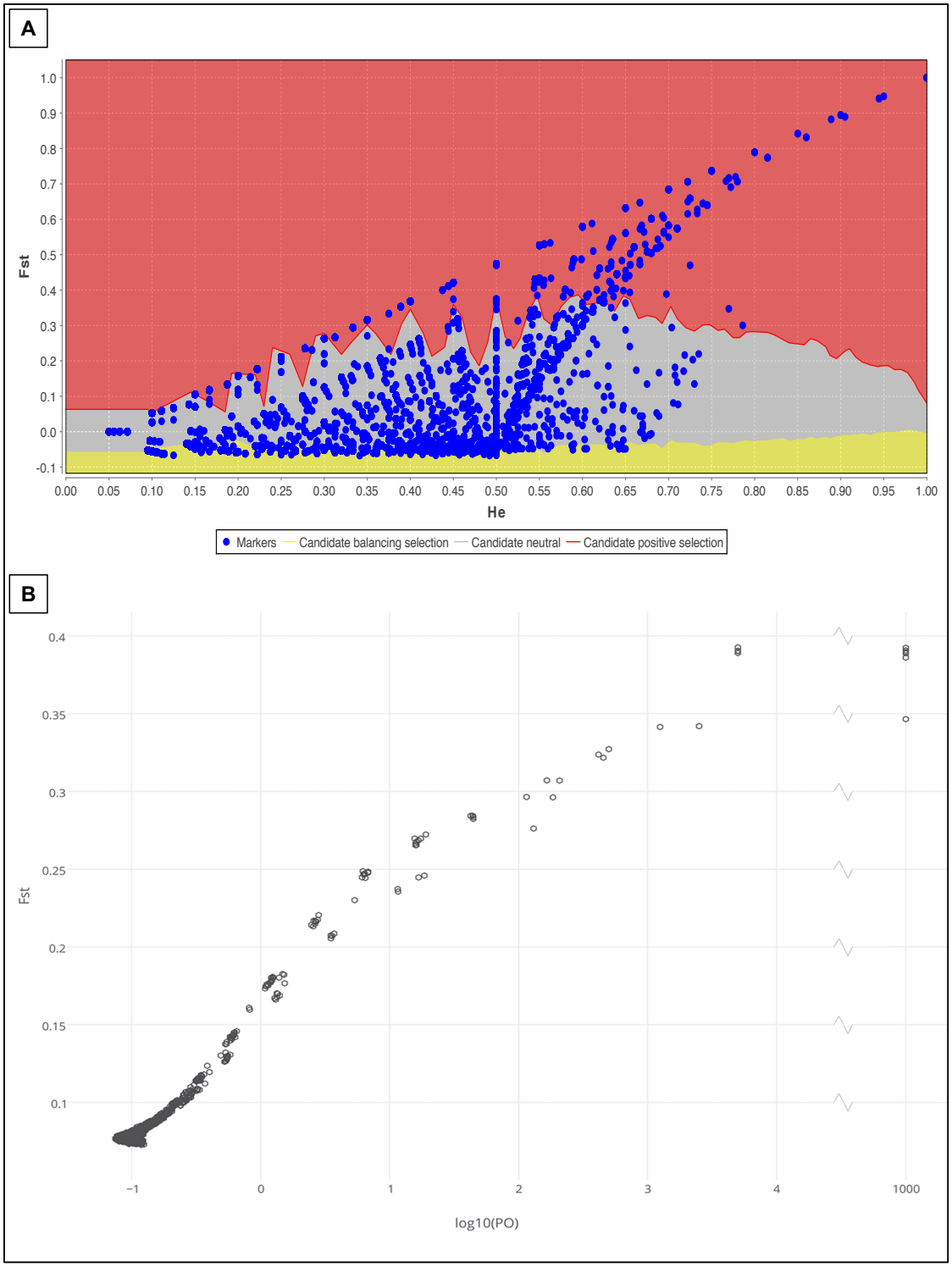

Figure 5.3: Outlier analyses for the RAD clusters (A) LOSITAN, each line represents F_{ST} thresholds to designate significance – any dot above the red line designates markers associated with positive selection, dots below the yellow line designates markers associated with balancing selection, while dots in between designates markers under neutrality (B) BayeScan, Log10(PO) represents the posterior probability.

Table 5.3: The gene identification information for RAD-seq markers that were outliers from both LOSITAN and BayeScan analyses, with chromosomal location, gene sequence description, min e-value and mean similarity (blast-n). One asterisks (*) denote candidates of medium interest, while two asterisks (**) denotes candidates of top interest.

Gene ID	Seq. Name	Chromosome	Seq. Description
METRN	RADid_0003361_depth_116	Autosome	anas platyrhynchos meteorin-like transcript variant mrna
ARHGAP44	RADid_0006028_depth_98	Autosome	gallus gallus rho gtpase activating protein 44 transcript variant mrna
TOPORS**	RADid_0016436_depth_109	Z-linked	anas platyrhynchos e3 ubiquitin-protein ligase topors-like partial mrna
ZNF469	RADid_0021978_depth_146	Autosome	anas platyrhynchos zinc finger protein 469 mrna
IL20	RADid_0022434_depth_89	Autosome	anas platyrhynchos interleukin 20 mrna
-	RADid_0027544_depth_59	Z-linked	gallus gallus bac clone tam33-29c6 from chromosome complete sequence
-	RADid_0027947_depth_120	Z-linked	opisthocomus hoazin bio-material lwl<deu
CAMKK1	RADid_0032309_depth_138	Autosome	anas platyrhynchos calcium calmodulin-dependent protein kinase kinase alpha transcript
TRIOBP	RADid_0035703_depth_125	Autosome	mus musculus trio and f-actin binding protein transcript variant mrna
F8*	RADid_0042196_depth_153	Autosome	anas platyrhynchos coagulation factor procoagulant component mrna
SYDE2	RADid_0043458_depth_294	Autosome	anas platyrhynchos synapse defective rho homolog 2 (elegans) transcript variant mrna
LIPA	RADid_0044346_depth_103	Autosome	anas platyrhynchos lysosomal acid lipase cholesteryl ester hydrolase-like transcript
IGF-1*	RADid_0050513_depth_248	Autosome	anser anser insulin-like growth factor i (igf-i) intron 3
TRIM71	RADid_0053903_depth_95	Autosome	anas platyrhynchos e3 ubiquitin-protein ligase trim71-like mrna
-	RADid_0061725_depth_138	Z-linked	gallus gallus bac clone ch261-26d19 from chromosome complete sequence
USP24	RADid_0064189_depth_48	Autosome	anas platyrhynchos ubiquitin specific peptidase 24 mrna
CLEC2L	RADid_0066143_depth_105	Z-linked	anas platyrhynchos c-type lectin domain family 2 member l-like transcript variant misc_rna
GEF	RADid_0068177_depth_21	Autosome	chlorocebus sabaeus rho guanine nucleotide exchange factor 10-like transcript variant
VIPR1	RADid_0069283_depth_65	Z-linked	ficedula albicollis vasoactive intestinal polypeptide receptor-like mrna
MID51	RADid_0071374_depth_108	Autosome	anas platyrhynchos smith-magenis syndrome chromosome candidate 7-like transcript
-	RADid_0073508_depth_100	Z-linked	anser cygnoides clone zaas082 microsatellite sequence
APEH	RADid_0075723_depth_129	Autosome	ficedula albicollis vasoactive intestinal polypeptide receptor-like mrna
TSPAN19	RADid_0075864_depth_49	Autosome	anas platyrhynchos tetraspanin 19 transcript variant mrna
TTLL11	RADid_0081068_depth_86	Autosome	anas platyrhynchos tubulin tyrosine ligase-like member 11 transcript variant mrna
DCIR	RADid_0081629_depth_106	Z-linked	anas platyrhynchos dendritic cell immunoreceptor complete cds dcar*null complete sequence and dendritic cell immunoactivating receptor complete cds
MRPL54	RADid_0084279_depth_80	Autosome	anas platyrhynchos mitochondrial ribosomal protein l54 partial mrna
ARHGEF18	RADid_0091284_depth_213	Autosome	anas platyrhynchos rho rac guanine nucleotide exchange factor 18 transcript
MED10	RADid_0098634_depth_34	Autosome	anas platyrhynchos mediator complex subunit 10 mrna
BMP2*	RADid_0105984_depth_195	Autosome	anas platyrhynchos bone morphogenetic protein 2-like mrna
DCIR	RADid_0107007_depth_106	Autosome	anas platyrhynchos dendritic cell immunoreceptor complete cds dcar*null complete sequence and dendritic cell immunoactivating receptor complete cds
NNT**	RADid_0107938_depth_49	Z-linked	chaetura pelagica nicotinamide nucleotide transhydrogenase mrna
-	RADid_0115589_depth_45	Z-linked	gallus gallus bac clone ch261-78c3 from chromosome complete sequence
TLE2	RADid_0118842_depth_161	Autosome	aquila chrysaetos canadensis transducin-like enhancer of split 2 transcript variant mrna
DCIR	RADid_0124734_depth_48	Z-linked	anas platyrhynchos dendritic cell immunoreceptor complete cds dcar*null complete sequence and dendritic cell immunoactivating receptor complete cds

Table 5.4: Gene flow estimates for both non-outlier and outlier loci with isolation-with-migration (IM) model in *∂a∂i* showing two independent runs between low- and high-altitude populations: Theta = Waterson's θ or nucleotide diversity, Ne = effective population size, Migration rate = individual migrates per generation. Averaged migration rates between the marker types for the two populations are shown in bold text.

						Ne		Migration Rate	
		Theta	Likelihood	S	time	Low	High	High -> Low	Low -> High
Non outliers	IM – run 1	9598.944863	-6375.446714	0.07592961	4.99593192	1.24293272	0.46102421	0.60743074	5.5158448
	IM – run 2	9598.944863	-6375.446714	0.06175858	1.68199377	1.10955537	0.37120589	0.44283411	9.99918778
	IM average	9598.944863	-	0.068844095	3.338962845	1.176244045	0.41611505	**0.525132425**	**7.75751629**
Outliers	IM – run 1	87.12585977	-996.6523572	0.80475161	9.99245969	0.90984298	0.17140578	0.24171386	2.34512902
	IM – run 2	87.12585977	-996.6523572	0.124112	9.99098621	2.9884493	0.23662533	0.22651885	1.58623989
	IM average	87.12585977	-	0.464431805	9.99172295	1.94914614	0.204015555	**0.234116355**	**1.965684455**

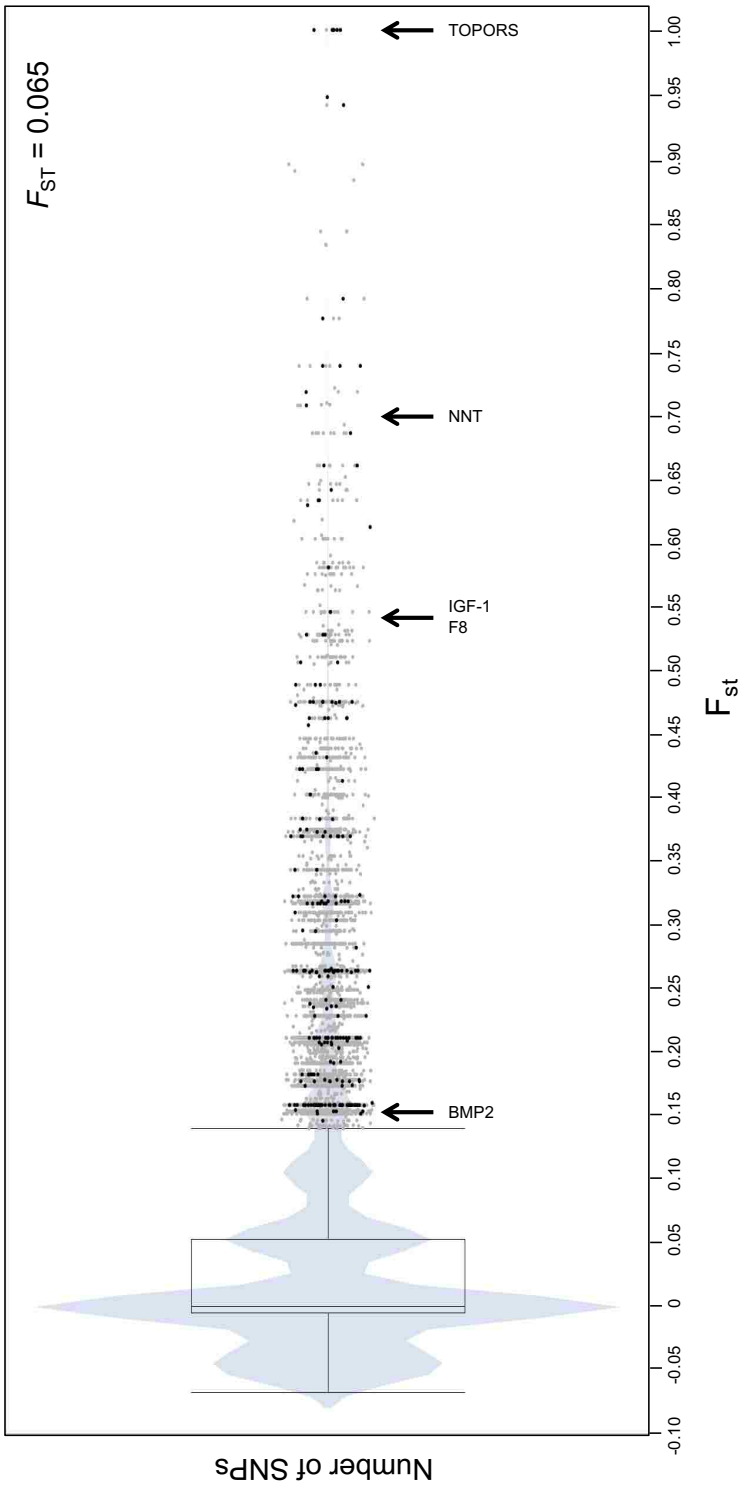

Figure 5.4: RAD cluster distribution against measurement of population divergence (F_{ST}): markers with black arrows signify RAD-seq markers and their respective gene identifier that were labeled as outliers via LOSITAN and BayeScan. Grey dots – putatively autosomal; Black dots – putatively Z-linked; high interest outliers are labeled.

Chapter 6: Synthesis

Given that oxygen homeostasis is considered a critical organismal biological constraint (Weir et al. 2005; Semenza 2007b, 2011), high-altitude environments provide valuable locations for investigating mechanisms for physiological adaptation, and for also illuminating the underlying genetic/genomic mechanisms associated with such a constraint (Cheviron and Brumfield 2011; Bigham and Lee 2014). The hallmark environmental constraint associated high-altitude environments is a drastic reduction of available O_2, with a 10% drop in O_2every 1,000 m. Organisms who invade these environments are subject to both immediate and long-term physiological acclimatization effects – increased breathing rate, blood-pressure, hematocrit and red-blood cell production, among others – and the requirement to adapt over the long term (Storz et al. 2010). Initially, there are risks of pulmonary and/or cerebral edema potentially resulting in death, but long-term acclimatization can result in pulmonary hypertension and erythrocytosis/polycythemia, known for increasing risks of heart-attack or stroke (Bärtsch and Swenson 2013). Therefore, organisms at high-altitude are under an extreme selective pressure to evolve adaptive mechanisms to mitigate these detrimental physiological effects. In the case of the Andean waterfowl, that were the subject of this dissertation, these same mechanisms represent agents for local adaptation, and population divergence.

Given the abundance of research on this same constraint in a multitude of organisms (Zhang et al. 2013; Gracey et al. 2004; Li et al. 2013b; Rissanen et al. 2006; Terova et al. 2008; Cheviron et al. 2008; Projecto-Garcia et al. 2013; Qiu et al. 2012; Qu et al. 2013; Scott et al. 2011; Wang et al. 2015; Wang et al. 2014; Yi et al. 2010; Xu et al. 2007; Xu et al. 2011; Kang et al. 2013; Alkorta-Aranburu et al. 2012; Xing et al. 2013;

Simonson et al. 2010; Simonson et al. 2012; Ge et al. 2012), I can also ask questions about the predictability of natural selection to converge on similar mechanisms, and the ability of parallelism to occur at different hierarchical levels (pathways → genes → exons/introns → amino acid/nucleotide). It is thought that mutations do not occur purely at random, since their distribution and location seems to be influenced by gene function, gene structure, and the roles of genes and gene products in functional networks (Stern and Orgogozo 2009). Therefore, a fundamental question in evolutionary genetics concerns the extent to which adaptive convergence in phenotype is caused by convergent or parallel changes at the molecular sequence level. Such a question has important implications for understanding the inherent repeatability (and, hence, predictability) of molecular adaptation (Stern 2013; Stern and Orgogozo 2008, 2009; Rosenblum et al. 2014).

Therefore, my dissertation asks, and seeks to answer, several questions by both assay of a priori candidate gene regions, and genome-wide scans: What are the genetic mechanisms (if any) associated with high-altitude/local adaptation in three independent lineages of Andean waterfowl species? Is there evidence for convergent evolution via parallel evolution, or collateral evolution through hybridization? And, to what extent is the convergent/parallel evolution occurring at the levels of pathway, gene, and individual codon or non-coding region?

Chapter Summaries

In chapter 2, I explored the role of genetic variation in the mitochondrial genomes of both high- and low-altitude populations of the three species. Mitochondria serve a

critical function in the production of cellular energy, due to their role in the electron transport chain and oxidative phosphorylation (OXPHOS). Evidence from a variety of organisms points to the possibility of convergent evolution on the mitochondria associated with physiological response to O_2 deprivation or temperature stress, including mechanisms for high-altitude adaptation. However, demographic fluctuations, such as population expansion or contraction, have the potential to mimic signatures of selection, producing a similar site-frequency spectrum. This study aimed to test for signatures of selection and/or demography with respect to potentially adaptive variation in the mitochondrial genomes of three Andean waterfowl species: yellow-billed pintail (*Anas georgica*), cinnamon teal (*Anas cyanoptera*), and speckled teal (*Anas flavirostris*). Here, I investigated the possible role of different types of selection, as associated with a major selective pressure (i.e., hypoxia or low O_2), as well as the potential effect of demographic forces as the neutral null hypothesis, by analyzing 60 full mitochondrial genomes for three different species sampled across the same altitudinal gradients. While each species that colonized high altitude appears to have undergone a founder effect and experienced a population expansion, our results show evidence for a pronounced role of purifying selection across the mitochondria for each population. I found more significant instances of purifying selection across high-altitude populations compared to their low-altitude counterparts across all three species. These results are based on a combination of analyses, which all point to a role for a specific selective pressure rather than due to patterns generated exclusively by demographic forces. In conclusion, these three Andean waterfowl species have become established in the high-altitude environment, at least in part due to a repeated pattern of purifying selection acting to maintain the OXPHOS

unit's ability to operate optimally through increased purifying (negative) selection in high-altitude populations.

In chapter 3, I investigated how variation in both the α- and β-globin gene clusters, with a focus on embryonic paralogs (α^{π} and β^{ε}), along with upstream cis-regulatory regions, were associated high-altitude adaptation. Hemoglobin is one of the leading, and best studied, molecular markers in studies of high-altitude adaptation, due to its' primary role of oxygen transportation. As a group, waterfowl have been featured prominently in studies of Hb function and evolution in response to high-altitude adaptation. Amino acid substitutions have been found that increase O_2-binding affinity of the major adult isoforms by affecting α- and β- subunit interactions and sensitivity to allosteric cofactors (Weber et al. 1993; Perutz 1983). Although the importance of Hb in regard to adapting to low- O_2 environments is well known, most studies have relied on sequencing only the major adult Hb genes, wholly excluding other features of the clusters - fetal/embryonic genes in the two clusters, as well as the up-stream regulatory elements containing transcription factor binding sites (TFBS). First, I found significant variation associated with the coding regions of the β^{ε} subunit in both Yellow-billed pintail (*Anas georgica*), and Speckled teal (*Anas flavirostris*). This variation was located at the same position in the β^{ε} subunit, resulting in identical amino-acid changes, ultimately suggesting either parallel evolution at these sites or gene-flow between these two high-altitude populations resulting in hybrid introgression. The same pattern of parallelism has been documented in these same species in their adult β^{A} subunit (McCracken et al. 2009a; McCracken et al. 2009b), resulting in amino-acid replacements that increase Hb-O_2 affinity (Natarajan et al. 2015). Previous work has established that these two species

periodically hybridize (McCracken and Wilson 2011), therefore, it seems likely that similar patterns across this entire gene cluster became established through gene-flow and introgression. Second, I also found significant genetic variation associated with putative TFBS in the upstream cis-regulatory regions of each gene, potentially altering TFBS identity in these species, in turn changing the binding behavior of the TFs that are directly involved in Hb expression. Ultimately, I concluded that exploring these additional avenues involved in control of gene expression has the potential to expand our understanding of the variety of ways organisms can adapt to high-altitude environments.

In chapter 4, I examined whether there is evidence for parallel evolution the HIF-pathway, not only across the three waterfowl species, but also convergence, in relation to high-altitude human populations. During periods of reduced O_2 supply, the most profound changes in gene expression are mediated by HIF-transcription factors who play a key role in cellular responses to low O_2 tension in a variety of organisms (Lisy and Peet 2008; Semenza 2007a; Webb et al. 2009), by targeting/upregulating genes involved in promoting red blood cell maturation and angiogenesis/vasomotor control (Haase 2013; Majmundar et al. 2010). Ultimately, I found strong support for both convergent and parallel evolution at different taxonomic levels. In the case of two of the three duck species, the same exonic regions in the same genes (*EGLN1*, *EPAS1*) exhibited sharply demarcated outliers with a high probability of directional selection in the high-altitude populations. My results mimic patterns of adaptation seen in human populations, with a combination of gain-of-function, loss-of-function mutations in *EPAS1*, and transcriptional regulation differences between *EGLN1*, causing changes in downstream target transactivation, resulting in a blunted hypoxic response. This example of parallel

evolution is striking given that these same two genes also have been identified in high-altitude human populations. However, in the case of humans, different regions of these genes appear to be under directional selection, therefore illustrating an example of predictable (same gene) yet unprectictable (different exons in a different part of the protein) convergent evolution between ducks and humans. Not only does this study pinpoint the molecular mechanism for high-altitude adaptation in Andean duck species, but it also highlights the degree to which natural selection frequently arrives at both parallel and convergent mechanisms of adaptation to similar environments. These results confirm that *EPAS1* and *EGLN* are "hotspots" for adaptations to life at high-altitudes, due to their centralized role in oxygen sensing.

Finally, in chapter 5, I used genome-wide variation from the speckled teal (*Anas flavirostris*) to explore other avenues for adaptation to the same environment through different mechanisms targeting similar physiological systems. Local adaptation frequently occurs across populations as a result of a balance between divergent selective pressures and gene flow associated with life in heterogeneous landscapes. Studying the effects of migration-selection balance on the adaptation process can be achieved in systems that have recently colonized extreme environments. This study utilizes an endemic South American duck species, the speckled teal (*Anas flavirostris*), which has both high- and low-altitude populations. High-altitude speckled teal (*A. f. oxyptera*) are locally adapted to the Andean environment and mostly allopatric from low-altitude birds (*A. f. flavirostris*); however, there is occasional gene flow across altitudinal gradients. In this study, I used next-generation sequencing (RAD-seq) to explore genetic patterns associated with high-altitude adaptation in speckled teal populations, as well as the extent

to which selection and migration have affected genetic architecture. I identified a set of loci with allele frequencies strongly correlated with altitude, including those involved in the insulin-like signaling pathway, bone morphogenesis, oxidative phosphorylation, responders to hypoxia-induced DNA damage, and feedback loops to the hypoxia-inducible factor (HIF)-pathway. These same outlier loci were found to have depressed gene flow estimates, as well as being highly concentrated on the Z-chromosome. These results suggest a multi-factorial response to life at high altitudes through an array of interconnected pathways that are likely under positive selection and whose genetic components seem to be providing an effective genomic barrier to interbreeding, potentially functioning as an avenue for population divergence and speciation.

Future Directions

The results of my dissertation have opened numerous avenues for potential research into elucidating further the mechanisms associated with adaptations to high-altitude environments.

First, although I did not find sufficient evidence for the role of genetic variation in the mitochondrial genome to be associated with high-altitude living (chapter 2), this fact does not preclude the mitochondria from being utilized in other ways that also facilitate adaptation to hypoxic environments. These include differential mitochondrial gene expression and, increased density of mitochondria in skeletal muscle, increased metabolic capacity and catalytic efficiency, all of which has been identified in other organisms (Hoppeler et al. 2003; Scott et al. 2009; Zhang et al. 2013; Gracey et al. 2004; Whitehead and Crawford 2006; Scott et al. 2015).

As for the other chapters in my dissertation, I have been careful to suggest that the variation I find in Hb (chapter 3), *EPAS1*, and *EGLN1* (chapter 4) has high potential to have resulted in functional change in these organisms; however, additional assays would need to be performed to more conclusively cement their role in high-altitude adaptation.

For example, site-directed mutagenesis could be used to assay the variation found in the embryonic hemoglobin β^{ε}. Previous functional properties of variants in adult globins have been successfully measured in this way by measure the functional effects of repeated substitutions that were implicated in convergent increases in Hb-O_2 affinity in high-altitude taxa. Although not possible with *in vivo* bird systems (unless through chicken CRISPR knock-ins), these experiments instead take place via recombinant *E.coli* vectors. Measurements of Hb-O_2 affinity could be taken of both low-altitude and high-altitude variants, and compared against one another.

In addition, general measurements of Hb concentrations could be taken in developing duck eggs of each of the species to get at differences at other Hb elements – relative concentrations of $\alpha^{\pi}/\beta^{\varepsilon}$ globins compared to adult, RBC count and hematocrit levels. Another result of from chapter 3, was potential evidence for adaptive introgression in the β-globin variants. Analyses utilizing estimations of gene-flow parameters from the β-globin cluster, as well as control loci would illuminate the degree to which the pattern seen is extreme parallelism or adaptive introgression.

Furthermore, significant variation was also shown to be associated with TFBS across the cluster. The methods used for TFBS identity were purely *in silico* predicted, and although it is likely not possible to assay changes in variation affect function (as suggested for β^{ε}), other analyses could be performed to assess additional aspects of

transcription factor binding activity (ie. ChIP-seq). This method is frequently used to map global binding sites precisely for any protein of interest, though largely does so across the entirety of the genome, rather than a specific region. Real-time PCR could be utilized to assess more general analyses of gene expression on a targeted region, however, this process would not be able to assess the TFBS specifically.

Lastly, the direct link between *EPAS1*, *EGLN1*, and Hb concentrations could also be probed further (chapter 5). Since the interaction between these genes are likely at the level of transcription, RNA-seq experiments could be utilized in order to assess transcription levels between the two populations, and perhaps also between tissue types in order to investigate whether *EPAS1* is directly or indirectly effecting Hb expression through EPO production.

Dissertation Summary

Overall, my dissertation shows a high degree of molecular convergence/parallelism on a number of previously identified genetic mechanisms, more specifically Hb and the HIF-pathway. In the Andean waterfowl, the level of parallelism was seen not only at the pathway level (chapter 4, 5), but at the gene (chapter 3, 4, 5), exon (chapter 4), and even nucleotide/amino-acid (chapter 4). To an extent, my dissertation suggests adaptive molecular evolution is indeed highly predictable, especially for adaptations to hypoxia in high-altitude environments, which presents a strong selective pressure. This is especially true in the case of Hb (chapter 3), *EPAS1*, and *EGLN1* (chapter 4), all of which occupy central roles in general response to physiological O_2-levels, with variation in these genes effecting a targeted response. However, in the

case of speckled teal (*Anas flavirostris*) and yellow-billed pintail (*Anas georgica*), the predictability of convergence was not always due to *de novo* mutations, but potentially due a *de novo* mutation in one species high-altitude population becoming adaptively introgressed into another species high-altitude population (chapter 3). It is also important to note, that such predictability did not extend to all three species assayed in this dissertation – specifically, the cinnamon teal (*Anas cyanoptera*) showed no amount of significant variation in any of the genes chosen for target enrichment. Although this may have more to do with a lack of time since divergence between its high- and low-altitude populations to allow for accumulation of adaptive variation, and a potential general reliance on phenotypic plasticity for physiological response. Altogether, my dissertation highlights the complexity to which adaptation occurs across the genome, due to the role of natural selection in response to an extreme environmental pressure.

References

Aggarwal, S., S. Negi, P. Jha, P. K. Singh, T. Stobdan, M. Q. Pasha, S. Ghosh, A. Agrawal, B. Prasher, and M. Mukerji (2010). EGLN1 involvement in high-altitude adaptation revealed through genetic analysis of extreme constitution types defined in Ayurveda. Proceedings of the National Academy of Sciences 107:18961-18966.

Ai, H., B. Yang, J. Li, X. Xie, H. Chen, and J. Ren (2014). Population history and genomic signatures for high-altitude adaptation in Tibetan pigs. BMC Genomics 15:834.

Alexander, D. H., and K. Lange (2011). Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics 12:246.

Alexander, D. H., J. Novembre, and K. Lange (2009). Fast model-based estimation of ancestry in unrelated individuals. Genome Research 19:1655-1664.

Alkorta-Aranburu, G., C. M. Beall, D. B. Witonsky, A. Gebremedhin, J. K. Pritchard, and A. Di Rienzo (2012). The genetic architecture of adaptations to high altitude in Ethiopia. PLoS Genetics 8:e1003110.

Anderson, M. G., J. M. Rhymer, and F. C. Rohwer (1992). Philopatry, dispersal, and the genetic structure of waterfowl populations. Ecology and Management of Breeding Waterfowl:365-395.

Antao, T., and M. A. Beaumont (2011). Mcheza: a workbench to detect selection using dominant markers. Bioinformatics 27:1717-1718.

Antao, T., A. Lopes, R. J. Lopes, A. Beja-Pereira, and G. Luikart (2008). LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. BMC Bioinformatics 9:323.

Arendt, J., and D. Reznick (2008). Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? Trends in Ecology & Evolution 23:26-32.

Arnqvist, G., D. K. Dowling, P. Eady, L. Gay, T. Tregenza, M. Tuda, and D. J. Hosken (2010). Genetic architecture of metabolic rate: environment specific epistasis between mitochondrial and nuclear genes in an insect. Evolution 64:3354-3363.

Ashur-Fabian, O., A. Avivi, L. Trakhtenbrot, K. Adamsky, M. Cohen, G. Kajakaro, A. Joel, N. Amariglio, E. Nevo, and G. Rechavi (2004). Evolution of p53 in hypoxia-stressed Spalax mimics human tumor mutation. Proceedings of the National Academy of Sciences of the United States of America 101:12236-12241.

Avise, J. C., R. T. Alisauskas, W. S. Nelson, and C. D. Ankney (1992). Matriarchal population genetic structure in an avian species with female natal philopatry. Evolution:1084-1096.

Backström, N., N. Karaiskou, E. H. Leder, L. Gustafsson, C. R. Primmer, A. Qvarnström, and H. Ellegren (2008). A gene-based genetic linkage map of the collared flycatcher (*Ficedula albicollis*) reveals extensive synteny and gene-order conservation during 100 million years of avian evolution. Genetics 179:1479-1495.

Backström, N., A. Qvarnström, L. Gustafsson, and H. Ellegren (2006). Levels of linkage disequilibrium in a wild bird population. Biology Letters 2:435-438.

Baird, N. A., P. D. Etter, T. S. Atwood, M. C. Currey, A. L. Shiver, Z. A. Lewis, E. U. Selker, W. A. Cresko, and E. A. Johnson (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. PloS ONE 3.

Bajic, V. B., M. R. Brent, R. H. Brown, A. Frankish, J. Harrow, U. Ohler, V. V. Solovyev, and S. L. Tan (2006). Performance assessment of promoter predictions on ENCODE regions in the EGASP experiment. Genome Biology 7:S3.

Balakrishnan, C. N., and S. V. Edwards (2009). Nucleotide variation, linkage disequilibrium and founder-facilitated speciation in wild populations of the zebra finch (*Taeniopygia guttata*). Genetics 181:645-660.

Ballard, J. W. O., and N. Pichaud (2014). Mitochondrial DNA: more than an evolutionary bystander. Functional Ecology 28:218-231.

Ballard, J. W. O., and M. C. Whitlock (2004). The incomplete natural history of mitochondria. Molecular Ecology 13:729-744.

Bamshad, M., and S. P. Wooding (2003). Signatures of natural selection in the human genome. Nature Reviews Genetics 4:99-111.

Bärtsch, P., and E. R. Swenson (2013). Acute high-altitude illnesses. New England Journal of Medicine 368:2294-2302.

Bazinet, A. L., D. J. Zwickl, and M. P. Cummings (2014). A gateway for phylogenetic analysis powered by grid computing featuring GARLI 2.0. Systematic Biology 63:812-818.

Beall, C. M. (2001). Adaptations to altitude: a current assessment. Annual Review of Anthropology:423-456.

Beall, C. M. (2007). Two routes to functional adaptation: Tibetan and Andean high-altitude natives. Proceedings of the National Academy of Sciences 104:8655-8660.

Beall, C. M., G. L. Cavalleri, L. Deng, R. C. Elston, Y. Gao, J. Knight, C. Li, J. C. Li, Y. Liang, and M. McCormack (2010). Natural selection on EPAS1 (HIF2 α) associated with low hemoglobin concentration in Tibetan highlanders. Proceedings of the National Academy of Sciences 107:11459-11464.

Beall, C. M., N. G. Jablonski, and A. Steegmann (2012). Human adaptation to climate: temperature, ultraviolet radiation, and altitude. Human biology: an evolutionary and biocultural perspective:163-224.

Beaumont, M. A., and R. A. Nichols (1996). Evaluating loci for use in the genetic analysis of population structure. Proceedings of the Royal Society of London. Series B: Biological Sciences 263:1619-1626.

Berg, P. R., S. Jentoft, B. Star, K. H. Ring, H. Knutsen, S. Lien, K. S. Jakobsen, and C. André (2015). Adaptation to low salinity promotes genomic divergence in Atlantic cod (*Gadus morhua L.*). Genome Biology and Evolution:evv093.

Bertozzi, D., J. Marinello, S. G. Manzo, F. Fornari, L. Gramantieri, and G. Capranico (2014). The natural inhibitor of DNA topoisomerase I, camptothecin, modulates HIF-1 α activity by changing miR expression patterns in human cancer cells. Molecular Cancer Therapeutics 13:239-248.

Bicudo, J. E. P., C. R. Vianna, and J. G. Chaui-Berlinck (2001). Thermogenesis in birds. Bioscience Reports 21:181-188.

Bigham, A., M. Bauchet, D. Pinto, X. Mao, J. M. Akey, R. Mei, S. W. Scherer, C. G. Julian, M. J. Wilson, and D. L. Herráez (2010). Identifying signatures of natural selection in Tibetan and Andean populations using dense genome scan data. PLoS Genetics 6:e1001116.

Bigham, A. W., and F. S. Lee (2014). Human high-altitude adaptation: forward genetics meets the HIF pathway. Genes & Development 28:2189-2204.

Bigham, A. W., X. Mao, R. Mei, T. Brutsaert, M. J. Wilson, C. G. Julian, E. J. Parra, J. M. Akey, L. G. Moore, and M. D. Shriver (2009). Identifying positive selection candidate loci for high-altitude adaptation in Andean populations. Human Genomics 4:79.

Bigham, A. W., M. J. Wilson, C. G. Julian, M. Kiyamu, E. Vargas, F. Leon‐Velarde, M. Rivera‐Chira, C. Rodriquez, V. A. Browne, and E. Parra (2013). Andean and Tibetan patterns of adaptation to high altitude. American Journal of Human Biology 25:190-197.

Blier, P. U., F. Dufresne, and R. S. Burton (2001). Natural selection and the evolution of mtDNA-encoded peptides: evidence for intergenomic co-adaptation. Trends in Genetics 17:400-406.

Bolnick, D. I., M. Turelli, H. Lopez-Fernández, P. C. Wainwright, and T. J. Near (2008). Accelerated mitochondrial evolution and "Darwin's corollary": asymmetric viability of reciprocal F1 hybrids in Centrarchid fishes. Genetics 178:1037-1048.

Borneman, A. R., T. A. Gianoulis, Z. D. Zhang, H. Yu, J. Rozowsky, M. R. Seringhaus, L. Y. Wang, M. Gerstein, and M. Snyder (2007). Divergence of transcription factor binding sites across related yeast species. Science 317:815-819.

Boyle, A. P., C. L. Araya, C. Brdlik, P. Cayting, C. Cheng, Y. Cheng, K. Gardner, L. W. Hillier, J. Janette, and L. Jiang (2014). Comparative analysis of regulatory information and circuits across distant species. Nature 512:453-456.

Breton, S., L. Milani, F. Ghiselli, D. Guerra, D. T. Stewart, and M. Passamonti (2014). A resourceful genome: updating the functional repertoire and evolutionary role of animal mitochondrial DNAs. Trends in Genetics 30:555-564.

Buroker, N. E., X.-H. Ning, Z.-N. Zhou, K. Li, W.-J. Cen, X.-F. Wu, W.-Z. Zhu, C. R. Scott, and S.-H. Chen (2012). EPAS1 and EGLN1 associations with high altitude sickness in Han and Tibetan Chinese at the Qinghai–Tibetan Plateau. Blood Cells, Molecules, and Diseases 49:67-73.

Caballero, A. (1995). On the effective size of populations with separate sexes, with particular reference to sex-linked genes. Genetics 139:1007-1011.

Cai, Q., X. Qian, Y. Lang, Y. Luo, J. Xu, S. Pan, Y. Hui, C. Gou, Y. Cai, and M. Hao (2013). Genome sequence of ground tit (*Pseudopodoces humilis*) and its adaptation to high altitude. Genome Biology 14:1.

Cao, A., and P. Moi (2002). Regulation of the globin genes. Pediatric Research 51:415-421.

Carey, C., O. Dunin-Borkowski, F. León-Velarde, D. Espinoza, and C. Monge (1994). Gas exchange and blood gases of Puna teal (Anas versicolor puna) embryos in the Peruvian Andes. Journal of Comparative Physiology B 163:649-656.

Carling, M. D., and R. T. Brumfield (2008). Haldane's rule in an avian system: Using cline theory and divergence population genetics to test for differential introgression of mitochondrial, autosomal, and sex-linked loci across the *Passerina* bunting hybrid zone. Evolution 62:2600-2615.

Carling, M. D., and R. T. Brumfield (2009). Speciation in Passerina buntings: introgression patterns of sex‐linked loci identify a candidate gene region for reproductive isolation. Molecular Ecology 18:834-847.

Carrero, P., K. Okamoto, P. Coumailleau, S. O'Brien, H. Tanaka, and L. Poellinger (2000). Redox-regulated recruitment of the transcriptional coactivators CREB-binding protein and SRC-1 to hypoxia-inducible factor 1 α. Molecular and Cellular Biology 20:402-415.

Cartharius, K., K. Frech, K. Grote, B. Klocke, M. Haltmeier, A. Klingenhoff, M. Frisch, M. Bayerlein, and T. Werner (2005). MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinformatics 21:2933-2942.

Castellana, S., S. Vicario, and C. Saccone (2011). Evolutionary patterns of the mitochondrial genome in Metazoa: exploring the role of mutation and selection in mitochondrial protein–coding Genes. Genome Biology and Evolution 3:1067-1079.

Caziani, S. M., E. J. Derlindati, A. Tálamo, A. L. Sureda, C. E. Trucco, and G. Nicolossi (2001). Waterbird richness in Altiplano wetlands of northwestern Argentina. Waterbirds:103-117.

Chakarova, C. F., H. Khanna, A. Z. Shah, S. B. Patil, T. Sedmak, C. A. Murga-Zamalloa, M. G. Papaioannou, K. Nagel-Wolfrum, I. Lopez, and P. Munro (2011). TOPORS, implicated in retinal degeneration, is a cilia-centrosomal protein. Human Molecular Genetics 20:975-987.

Chapman, B. S., L. E. Hood, and A. J. Tobin (1982). Minor early embryonic chick hemoglobin M. Amino acid sequences of the epsilon and alpha D chains. Journal of Biological Chemistry 257:651-658.

Chapman, B. S., A. J. Tobin, and L. E. Hood (1980). Complete amino acid sequences of the major early embryonic alpha-like globins of the chicken. Journal of Biological Chemistry 255:9051-9059.

Chapman, B. S., A. J. Tobin, and L. E. Hood (1981). Complete amino acid sequence of the major early embryonic beta-like globin in chickens. Journal of Biological Chemistry 256:5524-5531.

Chapman, M. A., S. J. Hiscock, and D. A. Filatov (2013). Genomic divergence during speciation driven by adaptation to altitude. Molecular Biology and Evolution:mst168.

Charlesworth, B. (2009). Effective population size and patterns of molecular evolution and variation. Nature Reviews Genetics 10:195-205.

Cheviron, Z., and R. Brumfield (2011). Genomic insights into adaptation to high-altitude environments. Heredity 108:354-361.

Cheviron, Z. A., G. C. Bachman, A. D. Connaty, G. B. McClelland, and J. F. Storz (2012). Regulatory changes contribute to the adaptive enhancement of thermogenic capacity in high-altitude deer mice. Proceedings of the National Academy of Sciences 109:8635-8640.

Cheviron, Z. A., C. Natarajan, J. Projecto-Garcia, D. K. Eddy, J. Jones, M. D. Carling, C. C. Witt, H. Moriyama, R. E. Weber, and A. Fago (2014). Integrating evolutionary and functional tests of adaptive hypotheses: a case study of altitudinal differentiation in hemoglobin function in an Andean sparrow, *Zonotrichia capensis*. Molecular Biology and Evolution 31:2948-2962.

Cheviron, Z. A., A. Whitehead, and R. T. Brumfield (2008). Transcriptomic variation and plasticity in rufous‐collared sparrows (*Zonotrichia capensis*) along an altitudinal gradient. Molecular Ecology 17:4556-4569.

Chohan, I. (2014). Blood coagulation changes at high altitude. Defence Science Journal 34:361-379.

Cirotto, C., I. Arangi, and F. Panara (1980). The haemoglobins of developing duck embryos. Development 60:389-404.

Cirotto, C., F. Panara, and I. Arangi (1987). The minor haemoglobins of primitive and definitive erythrocytes of the chicken embryo. Evidence for haemoglobin L. Development 101:805-813.

Clement, M., D. Posada, and K. A. Crandall (2000). TCS: a computer program to estimate gene genealogies. Molecular Ecology 9:1657-1659.

Compagnon, R., and C. Green (1994). PLINK User's Manual. Ecole Polytechnique Federale de Lausanna.

Conesa, A., S. Götz, J. M. García-Gómez, J. Terol, M. Talón, and M. Robles (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674-3676.

Conte, G. L., M. E. Arnegard, C. L. Peichel, and D. Schluter (2012). The probability of genetic parallelism and convergence in natural populations. Proceedings of the Royal Society of London B: Biological Sciences 279: 5039-5047

Cooper, G. M. (2000) The Cell: A Molecular Approach. 2nd edition. Sunderland (MA): Sinauer Associates. The Mechanism of Oxidative Phosphorylation. Available from: https://www.ncbi.nlm.nih.gov/books/NBK9885/

Cork, J. M., and M. D. Purugganan (2004). The evolution of molecular genetic pathways and networks. Bioessays 26:479-484.

Crews, S. T., and C.-M. Fan (1999). Remembrance of things PAS: regulation of development by bHLH–PAS proteins. Current Opinion in Genetics & Development 9:580-587.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, and S. T. Sherry (2011). The variant call format and VCFtools. Bioinformatics 27:2156-2158.

Das, J. (2006). The role of mitochondrial respiration in physiological and evolutionary adaptation. Bioessays 28:890-901.

Dawson, N. J., C. M. Ivy, L. Alza, R. Cheek, J. M. York, B. Chua, W. K. Milsom, K. G. McCracken, and G. R. Scott (2016). Mitochondrial physiology in the skeletal and cardiac muscles is altered in torrent ducks, *Merganetta armata*, from high altitudes in the Andes. Journal of Experimental Biology 219:3719-3728.

De Gobbi, M., V. Viprakasit, J. R. Hughes, C. Fisher, V. J. Buckle, H. Ayyub, R. J. Gibbons, D. Vernimmen, Y. Yoshinaga, and P. De Jong (2006). A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter. Science 312:1215-1217.

Dean, R., P. W. Harrison, A. E. Wright, F. Zimmer, and J. E. Mank (2015). Positive selection underlies Faster-Z evolution of gene expression in birds. Molecular Biology and Evolution 32:2646-2656.

DeFaveri, J., P. R. Jonsson, and J. Merilä (2013). Heterogeneous genomic differentiation in marine threespine sticklebacks: adaptation along an environmental gradient. Evolution 67:2530-2546.

DeLano, W. L. (2002). The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.

Dengler, V. L., M. D. Galbraith, and J. M. Espinosa (2014). Transcriptional regulation by hypoxia inducible factors. Critical Reviews in Biochemistry and Molecular Biology 49:1-15.

Dhami, K. K., L. Joseph, D. A. Roshier, and J. L. Peters (2016). Recent speciation and elevated Z‐chromosome differentiation between sexually monochromatic and dichromatic species of Australian teals. Journal of Avian Biology 47:92-102.

Dickerson, R. E., and I. Geis (1983). Hemoglobin: structure, function, evolution, and pathology. Benjamin-Cummings Publishing Company.

Dinkel, H., K. Van Roey, S. Michael, M. Kumar, B. Uyar, B. Altenberg, V. Milchevskaya, M. Schneider, H. Kühn, and A. Behrendt (2015). ELM 2016—data update and new functionality of the eukaryotic linear motif resource. Nucleic Acids Research:gkv1291.

Dowling, D. K., U. Friberg, and J. Lindell (2008). Evolutionary implications of non-neutral mitochondrial genetic variation. Trends in Ecology & Evolution 23:546-554.

Durrett, R. (2008). Probability models for DNA sequence evolution. Springer Science & Business Media.

Ehinger, M., P. Fontanillas, E. Petit, and N. Perrin (2002). Mitochondrial DNA variation along an altitudinal gradient in the greater white‐toothed shrew, *Crocidura russula*. Molecular Ecology 11:939-945.

Elgvin, T. O., J. S. Hermansen, A. Fijarczyk, T. Bonnet, T. Borge, S. A. Saether, K. L. Voje, and G. P. Saetre (2011). Hybrid speciation in sparrows II: a role for sex chromosomes? Molecular Ecology 20:3823-3837.

Ellegren, H. (2009). Genomic evidence for a large-Z effect. Proceedings of the Royal Society of London B: Biological Sciences 276:361-366.

Ellegren, H. (2010). Evolutionary stasis: the stable chromosomes of birds. Trends in Ecology & Evolution 25:283-291.

Ellegren, H., L. Smeds, R. Burri, P. I. Olason, N. Backström, T. Kawakami, A. Künstner, H. Mäkinen, K. Nadachowska-Brzyska, and A. Qvarnström (2012). The genomic landscape of species divergence in Ficedula flycatchers. Nature 491:756-760.

Ellison, C. K., and R. S. Burton (2006). Disruption of mitochondrial function in interpopulation hybrids of Tigriopus californicus. Evolution 60:1382-1391.

Ema, M., K. Hirota, J. Mimura, H. Abe, J. Yodoi, K. Sogawa, L. Poellinger, and Y. Fujii‐Kuriyama (1999). Molecular mechanisms of transcription activation by HLF and HIF1α in response to hypoxia: their stabilization and redox signal‐induced interaction with CBP/p300. The EMBO Journal 18:1905-1914.

Epstein, D. J. (2009). Cis-regulatory mutations in human disease. Briefings in Functional Genomics & Proteomics 8:310-316.

Erceg, J., T. E. Saunders, C. Girardot, D. P. Devos, L. Hufnagel, and E. E. Furlong (2014). Subtle changes in motif positioning cause tissue-specific effects on robustness of an enhancer's activity. PLoS Genetics 10:e1004060.

Excoffier, L., T. Hofer, and M. Foll (2009). Detecting loci under selection in a hierarchically structured population. Heredity 103:285-298.

Excoffier, L., G. Laval, and S. Schneider (2005). Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1:47.

Excoffier, L., and H. E. Lischer (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10:564-567.

Feder, J. L., S. P. Egan, and P. Nosil (2012). The genomics of speciation-with-gene-flow. Trends in Genetics 28:342-350.

Feder, J. L., and P. Nosil (2010). The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. Evolution 64:1729-1747.

Feldser, D., F. Agani, N. V. Iyer, B. Pak, G. Ferreira, and G. L. Semenza (1999). Reciprocal positive regulation of hypoxia-inducible factor 1 α and insulin-like growth factor 2. Cancer Research 59:3915-3918.

Figueira, T. R., M. H. Barros, A. A. Camargo, R. F. Castilho, J. C. Ferreira, A. J. Kowaltowski, F. E. Sluse, N. C. Souza-Pinto, and A. E. Vercesi (2013). Mitochondria as a source of reactive oxygen and nitrogen species: from molecular mechanisms to human health. Antioxidants & Redox Signaling 18:2029-2074.

Fjeldså, J. (1985). Origin, evolution, and status of the avifauna of Andean wetlands. Ornithological Monographs:85-112.

Flint, P. L., K. Ozaki, J. M. Pearce, B. Guzzetti, H. Higuchi, J. P. Fleskes, T. Shimada, and D. V. Derksen (2009). Breeding-season sympatry facilitates genetic exchange among allopatric wintering populations of northern pintails in Japan and California. The Condor 111:591-598.

Foll, M. (2012). BayeScan v2. 1 user manual. Ecology 20:1450-1462.

Foll, M., and O. Gaggiotti (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180:977-993.

Foll, M., O. E. Gaggiotti, J. T. Daub, A. Vatsiou, and L. Excoffier (2014). Widespread signals of convergent adaptation to high altitude in Asia and America. The American Journal of Human Genetics 95:394-407.

Frankham, R. (2007). Efective population size/adult population size ratios in wildlife: a review. Genetical Research 89:491-503.

Freeman, H., K. Shimomura, R. Cox, and F. Ashcroft (2006). Nicotinamide nucleotide transhydrogenase: a link between insulin secretion, glucose metabolism and oxidative stress. Biochemical Society Transactions 34:806-810.

Fukuda, R., K. Hirota, F. Fan, Y. Do Jung, L. M. Ellis, and G. L. Semenza (2002). Insulin-like growth factor 1 induces hypoxia-inducible factor 1-mediated vascular endothelial growth factor expression, which is dependent on MAP kinase and phosphatidylinositol 3-kinase signaling in colon cancer cells. Journal of Biological Chemistry 277:38205-38211.

Gale, D. P., S. K. Harten, C. D. Reid, E. G. Tuddenham, and P. H. Maxwell (2008). Autosomal dominant erythrocytosis and pulmonary arterial hypertension associated with an activating HIF2 α mutation. Blood 112:919-921.

Ge, R.-L., T. S. Simonson, R. C. Cooksey, U. Tanna, G. Qin, C. D. Huff, D. J. Witherspoon, J. Xing, B. Zhengzhong, and J. T. Prchal (2012). Metabolic insight into mechanisms of high-altitude adaptation in Tibetans. Molecular Genetics and Metabolism 106:244-247.

Gershoni, M., A. R. Templeton, and D. Mishmar (2009). Mitochondrial bioenergetics as a major motive force of speciation. Bioessays 31:642-650.

Gilman, J., N. Mishima, X. Wen, F. Kutlar, and T. Huisman (1988). Upstream promoter mutation associated with a modest elevation of fetal hemoglobin expression in human adults. Blood 72:78-81.

Gnirke, A., A. Melnikov, J. Maguire, P. Rogov, E. M. LeProust, W. Brockman, T. Fennell, G. Giannoukos, S. Fisher, and C. Russ (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nature Biotechnology 27:182-189.

Gompel, N., and B. Prud'homme (2009). The causes of repeated genetic evolution. Developmental Biology 332:36-47.

Gordon, A., and G. Hannon (2010). Fastx-toolkit. Computer program distributed by the author, website http://hannonlab.cshl.edu/fastx_toolkit/index.html [accessed 2014–2015].

Gorr, T. A., J. D. Cahn, H. Yamagata, and H. F. Bunn (2004). Hypoxia-induced synthesis of hemoglobin in the crustacean *Daphnia magna* is hypoxia-inducible factor-dependent. Journal of Biological Chemistry 279:36038-36047.

Gorr, T. A., M. Gassmann, and P. Wappner (2006). Sensing and responding to hypoxia via HIF in model invertebrates. Journal of Insect Physiology 52:349-364.

Gou, X., N. Li, L. Lian, D. Yan, H. Zhang, Z. Wei, and C. Wu (2007). Hypoxic adaptations of hemoglobin in Tibetan chick embryo: high oxygen-affinity mutation and selective expression. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 147:147-155.

Gou, X., Z. Wang, N. Li, F. Qiu, Z. Xu, D. Yan, S. Yang, J. Jia, X. Kong, and Z. Wei (2014). Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia. Genome Research 24:1308-1315.

Goulet, B. E., F. Roda, and R. Hopkins (2017). Hybridization in plants: old ideas, new techniques. Plant Physiology 173:65-78.

Gouw, S. C., H. M. van den Berg, J. Oldenburg, J. Astermark, P. G. de Groot, M. Margaglione, A. R. Thompson, W. van Heerde, J. Boekhorst, and C. H. Miller (2012). F8 gene mutation type and inhibitor development in patients with severe hemophilia A: systematic review and meta-analysis. Blood 119(12):2622-34.

Gracey, A. Y., E. J. Fraser, W. Li, Y. Fang, R. R. Taylor, J. Rogers, A. Brass, and A. R. Cossins (2004). Coping with cold: an integrative, multitissue analysis of the transcriptome of a poikilothermic vertebrate. Proceedings of the National Academy of Sciences of the United States of America 101:16970-16975.

Grispo, M. T., C. Natarajan, J. Projecto-Garcia, H. Moriyama, R. E. Weber, and J. F. Storz (2012). Gene duplication and the evolution of hemoglobin isoform differentiation in birds. Journal of Biological Chemistry 287:37647-37658.

Groenen, M. A., H. H. Cheng, N. Bumstead, B. F. Benkel, W. E. Briles, T. Burke, D. W. Burt, L. B. Crittenden, J. Dodgson, and J. Hillel (2000). A consensus linkage map of the chicken genome. Genome Research 10:137-147.

Guntur, A. R., and C. J. Rosen (2013). IGF-1 regulation of key signaling pathways in bone. BoneKEy Reports 2:437.

Gutenkunst, R. N., R. D. Hernandez, S. H. Williamson, and C. D. Bustamante (2009). Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genetics 5:e1000695.

Gutenkunst, R. N., R. D. Hernandez, S. H. Williamson, and C. D. Bustamante (2010). Diffusion Approximations for Demographic Inference: *DaDi*. Nature Precedings.

Haase, V. H. (2013). Regulation of erythropoiesis by hypoxia-inducible factors. Blood Reviews 27:41-53.

Hahn, C., L. Bachmann, and B. Chevreux (2013). Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Research 41:e129-e129.

Haluska, P., A. Saleem, Z. Rasheed, F. Ahmed, E. W. Su, L. F. Liu, and E. H. Rubin (1999). Interaction between human topoisomerase I and a novel RING finger/arginine-serine protein. Nucleic Acids Research 27:2538-2544.

Hanaoka, M., Y. Droma, B. Basnyat, M. Ito, N. Kobayashi, Y. Katsuyama, K. Kubo, and M. Ota (2012). Genetic variants in EPAS1 contribute to adaptation to high-altitude hypoxia in Sherpas. PloS ONE 7:e50566.

Harris, S. L., and A. J. Levine (2005). The p53 pathway: positive and negative feedback loops. Oncogene 24:2899-2908.

Hassanin, A., A. Ropiquet, A. Couloux, and C. Cruaud (2009). Evolution of the mitochondrial genome in mammals living at high altitude: new insights from a study of the tribe Caprini (Bovidae, Antilopinae). Journal of Molecular Evolution 68:293-310.

He, Z., and J. E. Russell (2001). Expression, purification, and characterization of human hemoglobins Gower-1 (ζ 2 ε 2), Gower-2 (α 2 ε 2), and Portland-2 (ζ 2 β 2) assembled in complex transgenic‐knockout mice. Blood 97:1099-1105.

Hedenstrom, A., and T. Alerstam (1995). Optimal flight speed of birds. Philosophical Transactions of the Royal Society of London B: Biological Sciences 348:471-487.

Hedrick, P. W. (2013). Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Molecular Ecology 22:4606-4618.

Hoekstra, L. A., M. A. Siddiq, and K. L. Montooth (2013). Pleiotropic Effects of a Mitochondrial–Nuclear Incompatibility Depend upon the Accelerating Effect of Temperature in Drosophila. Genetics 195:1129-1139.

Hoffmann, F. G., J. C. Opazo, and J. F. Storz (2008). Rapid rates of lineage-specific gene duplication and deletion in the α-globin gene family. Molecular Biology and Evolution 25:591-602.

Hoffmann, F. G., J. F. Storz, T. A. Gorr, and J. C. Opazo (2010). Lineage-specific patterns of functional diversification in the α-and β-globin gene families of tetrapod vertebrates. Molecular Biology and Evolution 27:1126-1138.

Hofmann, O., and T. Brittain (1996). Ligand binding kinetics and dissociation of the human embryonic haemoglobins. Biochemical Journal 315:65-70.

Hohenlohe, P. A., S. Bassham, P. D. Etter, N. Stiffler, E. A. Johnson, and W. A. Cresko (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS Genet 6:e1000862.

Hong, S.-S., H. Lee, and K.-W. Kim (2004). HIF-1 α : a valid therapeutic target for tumor therapy. Cancer Research and Treatment 36:343-353.

Hoogewijs, D., N. Terwilliger, K. Webster, J. Powell-Coffman, S. Tokishita, H. Yamagata, T. Hankeln, T. Burmester, K. Rytkönen, and M. Nikinmaa (2007). From critters to cancers: bridging comparative and clinical research on oxygen sensing, HIF signaling, and adaptations towards hypoxia. Integrative and Comparative Biology 47:552-577.

Hopkins, S. R., and F. L. Powell (2001). Common themes of adaptation to hypoxia. In Hypoxia. Springer 153-167.

Hoppeler, H., M. Vogt, E. R. Weibel, and M. Flück (2003). Response of skeletal muscle mitochondria to hypoxia. Experimental Physiology 88:109-119.

Hu, C., Q. Chang, W. Zhou, L. Yan, T. Pan, C. Xue, and B. Zhang (2015). Mitochondrial genome of the Anas crecca (Anatidae: Anas). Mitochondrial DNA 26:625-626.

Hu, C.-J., A. Sataur, L. Wang, H. Chen, and M. C. Simon (2007). The N-terminal transactivation domain confers target gene specificity of hypoxia-inducible factors HIF-1 α and HIF-2 α . Molecular Biology of the Cell 18:4528-4542.

Hu, C.-J., L.-Y. Wang, L. A. Chodosh, B. Keith, and M. C. Simon (2003). Differential roles of hypoxia-inducible factor 1 α (HIF-1 α) and HIF-2 α in hypoxic gene regulation. Molecular and Cellular Biology 23:9361-9374.

Huang, T.-T., M. Naeemuddin, S. Elchuri, M. Yamaguchi, H. M. Kozy, E. J. Carlson, and C. J. Epstein (2006). Genetic modifiers of the phenotype of mice deficient in mitochondrial superoxide dismutase. Human Molecular Genetics 15:1187-1194.

Huerta-Sánchez, E., M. DeGiorgio, L. Pagani, A. Tarekegn, R. Ekong, T. Antao, A. Cardona, H. E. Montgomery, G. L. Cavalleri, and P. A. Robbins (2013). Genetic signatures reveal high-altitude adaptation in a set of Ethiopian populations. Molecular Biology and Evolution 30:1877-1888.

Huerta-Sánchez, E., X. Jin, Z. Bianba, B. M. Peter, N. Vinckenbosch, Y. Liang, X. Yi, M. He, M. Somel, and P. Ni (2014). Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. Nature 512:194-197.

Isaacks, R. E., and D. R. Harkness (1980). Erythrocyte organic phosphates and hemoglobin function in birds, reptiles, and fishes. American Zoologist 20:115-129.

Jastroch, M., A. S. Divakaruni, S. Mookerjee, J. R. Treberg, and M. D. Brand (2010). Mitochondrial proton and electron leaks. Essays in Biochemistry 47:53-67.

Jeong, C., G. Alkorta-Aranburu, B. Basnyat, M. Neupane, D. B. Witonsky, J. K. Pritchard, C. M. Beall, and A. Di Rienzo (2014). Admixture facilitates genetic adaptations to high altitude in Tibet. Nature Communications 5:3281.

Kaelin Jr, W. G. (2005). Proline hydroxylation and gene expression. Annual Reviews in Biochemistry. 74:115-128.

Kaelin, W. G., and P. J. Ratcliffe (2008). Oxygen sensing by metazoans: the central role of the HIF hydroxylase pathway. Molecular Cell 30:393-402.

Kanehisa, M., and S. Goto (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Research 28:27-30.

Kang, L., H.-X. Zheng, F. Chen, S. Yan, K. Liu, Z. Qin, L. Liu, Z. Zhao, L. Li, and X. Wang (2013). MtDNA lineage expansions in Sherpa population suggest adaptive evolution in Tibetan highlands. Molecular Biology and Evolution:mst147.

Kasowski, M., F. Grubert, C. Heffelfinger, M. Hariharan, A. Asabere, S. M. Waszak, L. Habegger, J. Rozowsky, M. Shi, and A. E. Urban (2010). Variation in transcription factor binding among humans. Science 328:232-235.

Katoh, K., K. Misawa, K. i. Kuma, and T. Miyata (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Research 30:3059-3066.

Kearse, M., R. Moir, A. Wilson, S. Stones-Havas, M. Cheung, S. Sturrock, S. Buxton, A. Cooper, S. Markowitz, and C. Duran (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647-1649.

Kirkpatrick, M., and N. Barton (2006). Chromosome inversions, local adaptation and speciation. Genetics 173:419-434.

Knezetic, J. A., and G. Felsenfeld (1993). Mechanism of developmental regulation of alpha pi, the chicken embryonic alpha-globin gene. Molecular and Cellular Biology 13:4632-4639.

Koevoets, T., L. Van de Zande, and L. Beukeboom (2012). Temperature stress increases hybrid incompatibilities in the parasitic wasp genus Nasonia. Journal of Evolutionary Biology 25:304-316.

Kumar, S., G. Stecher, and K. Tamura (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular Biology and Evolution:msw054.

Lande, R. (2009). Adaptation to an extraordinary environment by evolution of phenotypic plasticity and genetic assimilation. Journal of Evolutionary Biology 22:1435-1446.

Lando, D., D. J. Peet, D. A. Whelan, J. J. Gorman, and M. L. Whitelaw (2002). Asparagine hydroxylation of the HIF transactivation domain: a hypoxic switch. Science 295:858-861.

Langmead, B., and S. L. Salzberg (2012). Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357-359.

Laslett, D., and B. Canbäck (2008). ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics 24:172-175.

Lavretsky, P., J. M. Dacosta, B. E. Hernández‐Baños, A. Engilis, M. D. Sorenson, and J. L. Peters (2015). Speciation genomics and a role for the Z chromosome in the early stages of divergence between Mexican ducks and mallards. Molecular Ecology 24:5364-5378.

Lavretsky, P., J. L. Peters, K. Winker, V. Bahn, I. Kulikova, Y. N. Zhuravlev, R. E. Wilson, C. Barger, K. Gurney, and K. G. McCracken (2016). Becoming pure: identifying generational classes of admixed individuals within lesser and greater scaup populations. Molecular Ecology.

Le Corre, V., and A. Kremer (2012). The genetic differentiation at quantitative trait loci under local adaptation. Molecular Ecology 21:1548-1566.

Ledent, V., and M. Vervoort (2001). The basic helix-loop-helix protein family: comparative genomics and phylogenetic analysis. Genome Research 11:754-770.

Lenhard, B., A. Sandelin, and P. Carninci (2012). Metazoan promoters: emerging characteristics and insights into transcriptional regulation. Nature Reviews Genetics 13:233-245.

Leon-Velarde, F., J. Whittembury, C. Carey, and C. Monge (1984). Permeability of eggshells of native chickens in the Peruvian Andes. In Respiration and metabolism of embryonic vertebrates. Springer 245-257.

Levine, A. J. (1997). p53, the cellular gatekeeper for growth and division. Cell 88:323-331.

Levis, N. A., and D. W. Pfennig (2016). Evaluating 'plasticity-first'evolution in nature: key criteria and empirical approaches. Trends in Ecology & Evolution 7:563-74.

Levo, M., and E. Segal (2014). In pursuit of design principles of regulatory sequences. Nature Reviews Genetics 15:453-468.

Li, H., and R. Durbin (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754-1760.

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

Li, M., S. Tian, L. Jin, G. Zhou, Y. Li, Y. Zhang, T. Wang, C. K. Yeung, L. Chen, and J. Ma (2013a). Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. Nature Genetics 45:1431-1438.

Li, W.-H., and M. Nei (1974). Stable linkage disequilibrium without epistasis in subdivided populations. Theoretical Population Biology 6:173-183.

Li, Y., Z. Ren, A. M. Shedlock, J. Wu, L. Sang, T. Tersing, M. Hasegawa, T. Yonezawa, and Y. Zhong (2013b). High altitude adaptation of the schizothoracine fishes (Cyprinidae) revealed by the mitochondrial genome analyses. Gene 517:169-178.

Li, Y., D.-D. Wu, A. R. Boyko, G.-D. Wang, S.-F. Wu, D. M. Irwin, and Y.-P. Zhang (2014). Population variation revealed high-altitude adaptation of Tibetan mastiffs. Molecular Biology and Evolution 31:1200-1205.

Librado, P., and J. Rozas (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451-1452.

Linhart, C., Y. Halperin, and R. Shamir (2008). Transcription factor and microRNA motif discovery: the Amadeus platform and a compendium of metazoan target sets. Genome Research 18:1180-1189.

Lisy, K., and D. J. Peet (2008). Turn me on: regulating HIF transcriptional activity. Cell Death & Differentiation 15:642-649.

Londono, G. A., M. A. Chappell, M. d. R. Castañeda, J. E. Jankowski, and S. K. Robinson (2015). Basal metabolism in tropical birds: latitude, altitude, and the 'pace of life'. Functional Ecology 29:338-346.

Londoño, G. A., M. A. Chappell, J. E. Jankowski, and S. K. Robinson (2016). Do thermoregulatory costs limit altitude distributions of Andean forest birds? Functional Ecology.

Lorenzo, F. R., C. Huff, M. Myllymäki, B. Olenchock, S. Swierczek, T. Tashi, V. Gordeuk, T. Wuren, G. Ri-Li, and D. A. McClain (2014). A genetic mechanism for Tibetan high-altitude adaptation. Nature Genetics.

Losos, J. B. (2011). Convergence, adaptation, and constraint. Evolution 65:1827-1840.

Lott, M. T., J. N. Leipzig, O. Derbeneva, H. M. Xie, D. Chalkia, M. Sarmady, V. Procaccio, and D. C. Wallace (2013). mtDNA variation and analysis using MITOMAP and MITOMASTER. Current Protocols in Bioinformatics:1.23. 21-21.23. 26.

Luo, Y., W. Gao, Y. Gao, S. Tang, Q. Huang, X. Tan, J. Chen, and T. Huang (2008). Mitochondrial genome analysis of *Ochotona curzoniae* and implication of cytochrome c oxidase in hypoxic adaptation. Mitochondrion 8:352-357.

Majmundar, A. J., W. J. Wong, and M. C. Simon (2010). Hypoxia-inducible factors and the response to hypoxic stress. Molecular Cell 40:294-309.

Mank, J. E., E. Axelsson, and H. Ellegren (2007). Fast-X on the Z: rapid evolution of sex-linked genes in birds. Genome Research 17:618-624.

Manning, L. R., J. E. Russell, J. C. Padovan, B. T. Chait, A. Popowicz, R. S. Manning, and J. M. Manning (2007). Human embryonic, fetal, and adult hemoglobins have different subunit interface strengths. Correlation with lifespan in the red cell. Protein Science 16:1641-1658.

Martin, A., and V. Orgogozo (2013). The loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. Evolution 67:1235-1250.

Martin, S. H., K. K. Dasmahapatra, N. J. Nadeau, C. Salazar, J. R. Walters, F. Simpson, M. Blaxter, A. Manica, J. Mallet, and C. D. Jiggins (2013). Genome-wide evidence for speciation with gene flow in Heliconius butterflies. Genome Research 23:1817-1828.

Maurano, M. T., R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds, R. Sandstrom, H. Qu, and J. Brody (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science 337:1190-1195.

McCracken, K. G., C. P. Barger, M. Bulgarella, K. P. Johnson, M. K. Kuhner, A. V. Moore, J. L. Peters, J. Trucco, T. H. Valqui, K. Winker, and R. E. Wilson (2009a). Signatures of High-Altitude Adaptation in the Major Hemoglobin of Five Species of Andean Dabbling Ducks. American Naturalist 174:631-650.

McCracken, K. G., C. P. Barger, M. Bulgarella, K. P. Johnson, S. A. Sonsthagen, J. Trucco, T. H. Valqui, R. E. Wilson, K. Winker, and M. D. Sorenson (2009b). Parallel evolution in the major haemoglobin genes of eight species of Andean waterfowl. Molecular Ecology 18:3992-4005.

McCracken, K. G., C. P. Barger, and M. D. Sorenson (2010). Phylogenetic and structural analysis of the HbA (alpha(A)/beta(A)) and HbD (alpha(D)/beta(A)) hemoglobin genes in two high-altitude waterfowl from the Himalayas and the Andes: Bar-headed goose (*Anser indicus*) and Andean goose (*Chloephaga melanoptera*). Molecular Phylogenetics and Evolution 56:649-658.

McCracken, K. G., M. Bulgarella, K. P. Johnson, M. K. Kuhner, J. Trucco, T. H. Valqui, R. E. Wilson, and J. L. Peters (2009c). Gene Flow in the Face of Countervailing Selection: Adaptation to High-Altitude Hypoxia in the beta A Hemoglobin Subunit of Yellow-Billed Pintails in the Andes. Molecular Biology and Evolution 26:815-827.

McCracken, K. G., and R. E. Wilson (2011). Gene flow and hybridization between numerically imbalanced populations of two duck species in the Falkland Islands. PloS ONE 6:e23173.

McNab, B. K. (2003). Metabolism: Ecology shapes bird bioenergetics. Nature 426:620-621.

McNab, B. K. (2012). Extreme measures: the ecological energetics of birds and mammals. University of Chicago Press.

Medugorac, I., A. Graf, C. Grohs, S. Rothammer, Y. Zagdsuren, E. Gladyr, N. Zinovieva, J. Barbieri, D. Seichter, and I. Russ (2017). Whole-genome analysis of introgressive hybridization and characterization of the bovine legacy of Mongolian yaks. Nature Genetics 49(3):470-475.

Megens, H.-J., R. P. Crooijmans, J. W. Bastiaansen, H. H. Kerstens, A. Coster, R. Jalving, A. Vereijken, P. Silva, W. M. Muir, and H. H. Cheng (2009). Comparison of linkage disequilibrium and haplotype diversity on macro-and microchromosomes in chicken. BMC Genetics 10:86.

Meiklejohn, C. D., M. A. Holmbeck, M. A. Siddiq, D. N. Abt, D. M. Rand, and K. L. Montooth (2013). An incompatibility between a mitochondrial tRNA and its nuclear-encoded tRNA synthetase compromises development and fitness in Drosophila. PLoS Genetics 9:e1003238.

Melo-Ferreira, J., J. Vilela, M. M. Fonseca, R. R. da Fonseca, P. Boursot, and P. C. Alves (2014). The elusive nature of adaptive mitochondrial DNA evolution of an arctic lineage prone to frequent introgression. Genome Biology and Evolution 6:886-896.

Merika, M., and S. H. Orkin (1995). Functional synergy and physical interactions of the erythroid transcription factor GATA-1 with the Krüppel family proteins Sp1 and EKLF. Molecular and Cellular Biology 15:2437-2447.

Miller, M. R., J. P. Dunham, A. Amores, W. A. Cresko, and E. A. Johnson (2007). Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Research 17:240-248.

Mishmar, D., E. Ruiz-Pesini, P. Golik, V. Macaulay, A. G. Clark, S. Hosseini, M. Brandon, K. Easley, E. Chen, and M. D. Brown (2003). Natural selection shaped regional mtDNA variation in humans. Proceedings of the National Academy of Sciences 100:171-176.

Mitterboeck, T. F., and S. J. Adamowicz (2013). Flight loss linked to faster molecular evolution in insects. Proceedings of the Royal Society of London B: Biological Sciences 280:20131128.

Monge, C., and F. Leon-Velarde (1991). Physiological adaptation to high altitude: oxygen transport in mammals and birds. Physiological Reviews 71:1135-1172.

Moore, L., M. Shriver, L. Bemis, B. Hickler, M. Wilson, T. Brutsaert, E. Parra, and E. Vargas (2004). Maternal adaptation to high-altitude pregnancy: an experiment of nature—a review. Placenta 25:S60-S71.

Mugal, C. F., J. B. Wolf, and I. Kaj (2013). Why time matters: codon evolution and the temporal dynamics of dN/dS. Molecular Biology and Evolution:mst192.

Murrell, B., J. O. Wertheim, S. Moola, T. Weighill, K. Scheffler, and S. L. K. Pond (2012). Detecting individual sites subject to episodic diversifying selection. PLoS Genetics 8:e1002764.

Nadimpalli, S., A. V. Persikov, and M. Singh (2015). Pervasive variation of transcription factor orthologs contributes to regulatory network evolution. PLoS Genetics 11:e1005011.

Nakae, J., Y. Kido, and D. Accili (2001). Distinct and overlapping functions of insulin and IGF-I receptors. Endocrine Reviews 22:818-835.

Natarajan, C., N. Inoguchi, R. E. Weber, A. Fago, H. Moriyama, and J. F. Storz (2013). Epistasis among adaptive mutations in deer mouse hemoglobin. Science 340:1324-1327.

Natarajan, C., J. Projecto-Garcia, H. Moriyama, R. E. Weber, V. Muñoz-Fuentes, A. J. Green, C. Kopuchian, P. L. Tubaro, L. Alza, and M. Bulgarella (2015). Convergent Evolution of Hemoglobin Function in High-Altitude Andean Waterfowl Involves Limited Parallelism at the Molecular Sequence Level. PLoS Genetics 11:e1005681.

Nei, M., and T. Gojobori (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Molecular Biology and Evolution 3:418-426.

Nei, M., and S. Kumar (2000). Molecular evolution and phylogenetics. Oxford University Press, New York.

Nei, M., and W.-H. Li (1973). Linkage disequilibrium in subdivided populations. Genetics 75:213-219.

Nery, M. F., J. I. Arroyo, and J. C. Opazo (2013). Genomic organization and differential signature of positive selection in the alpha and beta globin gene clusters in two cetacean species. Genome Biology and Evolution 5:2359-2367.

Newman, J. H., T. N. Holt, J. D. Cogan, B. Womack, J. A. Phillips III, C. Li, Z. Kendall, K. R. Stenmark, M. G. Thomas, and R. D. Brown (2015). Increased prevalence of EPAS1 variant in cattle with high-altitude pulmonary hypertension. Nature Communications 6:6863.

Ng, P. C., and S. Henikoff (2006). Predicting the effects of amino acid substitutions on protein function. Annuual Review of Genomics and Human Genetics 7:61-80.

Niehuis, O., A. K. Judson, and J. Gadau (2008). Cytonuclear genic incompatibilities cause increased mortality in male F2 hybrids of Nasonia giraulti and N. vitripennis. Genetics 178:413-426.

Nielsen, R. (2005). Molecular signatures of natural selection. Annual Reviews of Genetics 39:197-218.

Norberg, U. M. (2012). Vertebrate flight: mechanics, physiology, morphology, ecology and evolution. Springer Science & Business Media.

Nosil, P., D. J. Funk, and D. Ortiz-Barrientos (2009). Divergent selection and heterogeneous genomic divergence. Molecular Ecology 18:375-402.

O'Brodovich, H. M., M. Andrew, G. Gray, and G. Coates (1984). Hypoxia alters blood coagulation during acute decompression in humans. Journal of Applied Physiology 56:666-670.

Odom, D. T., R. D. Dowell, E. S. Jacobsen, W. Gordon, T. W. Danford, K. D. MacIsaac, P. A. Rolfe, C. M. Conboy, D. K. Gifford, and E. Fraenkel (2007). Tissue-specific transcriptional regulation has diverged significantly between human and mouse. Nature Genetics 39:730-732.

Opazo, J. C., F. G. Hoffmann, C. Natarajan, C. C. Witt, M. Berenbrink, and J. F. Storz (2015). Gene turnover in the avian globin gene families and evolutionary changes in hemoglobin isoform expression. Molecular Biology and Evolution 32:871-887.

Pagani, L., Q. Ayub, D. G. MacArthur, Y. Xue, J. K. Baillie, Y. Chen, I. Kozarewa, D. J. Turner, S. Tofanelli, and K. Bulayeva (2012). High altitude adaptation in Daghestani populations from the Caucasus. Human Genetics 131:423-433.

Palmer, B. F., and D. J. Clegg (2014). Oxygen sensing and metabolic homeostasis. Molecular and Cellular Endocrinology 397:51-58.

Pan, T., L. Ren, H. Wang, J. Chen, and B. Zhang (2014). Mitochondrial genome of the Anas falcata (Anatidae: Anas). Mitochondrial DNA 25:111-112.

Paul, N. D., and D. Gwynn-Jones (2003). Ecological roles of solar UV radiation: towards an integrated approach. Trends in Ecology & Evolution 18:48-55.

Peng, Y., C. Cui, Y. He, H. Zhang, D. Yang, Q. Zhang, L. Yang, Y. He, K. Xiang, and X. Zhang (2017). Down-Regulation of EPAS1 Transcription and Genetic Adaptation of Tibetans to High-Altitude Hypoxia. Molecular Biology and Evolution:msw280.

Peng, Y., Z. Yang, H. Zhang, C. Cui, X. Qi, X. Luo, X. Tao, T. Wu, H. Chen, and H. Shi (2011). Genetic variations in Tibetan populations and high-altitude adaptation at the Himalayas. Molecular Biology and Evolution 28:1075-1081.

Percy, M. J., P. A. Beer, G. Campbell, A. W. Dekker, A. R. Green, D. Oscier, M. G. Rainey, R. van Wijk, M. Wood, and T. R. Lappin (2008a). Novel exon 12 mutations in the HIF2A gene associated with erythrocytosis. Blood 111:5400-5402.

Percy, M. J., P. W. Furlow, G. S. Lucas, X. Li, T. R. Lappin, M. F. McMullin, and F. S. Lee (2008b). A gain-of-function mutation in the HIF2A gene in familial erythrocytosis. New England Journal of Medicine 358:162-168.

Pereira, S. L., and A. J. Baker (2006). A mitogenomic timescale for birds detects variable phylogenetic rates of molecular evolution and refutes the standard molecular clock. Molecular Biology and Evolution 23:1731-1740.

Pérez‐Figueroa, A., M. García‐Pereira, M. Saura, E. Rolán‐Alvarez, and A. Caballero (2010). Comparing three different methods to detect selective loci using dominant markers. Journal of Evolutionary Biology 23:2267-2276.

Perutz, M. F. (1983). Species adaptation in a protein molecule. Molecular Biology and Evolution 1:1-28.

Perutz, M. F. (1989). Mechanisms of cooperativity and allosteric regulation in proteins. Quarterly Reviews of Biophysics 22:139-237.

Peters, J. L., K. A. Bolender, and J. M. Pearce (2012). Behavioural vs. molecular sources of conflict between nuclear and mitochondrial DNA: the role of male‐biased dispersal in a Holarctic sea duck. Molecular Ecology 21:3562-3575.

Peters, J. L., K. E. Omland, and K. Johnson (2007). Population structure and mitochondrial polyphyly in North American Gadwalls (*Anas strepera*). The Auk 124:444-462.

Phillips, P. C. (2008). Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. Nature Reviews Genetics 9:855-867.

Pigliucci, M., C. J. Murren, and C. D. Schlichting (2006). Phenotypic plasticity and evolution by genetic assimilation. Journal of Experimental Biology 209:2362-2367.

Pond, S. L. K., and S. D. Frost (2005a). Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21:2531-2533.

Pond, S. L. K., and S. D. Frost (2005b). Not so different after all: a comparison of methods for detecting amino acid sites under selection. Molecular Biology and Evolution 22:1208-1222.

Powell, F. L. (2003). Functional genomics and the comparative physiology of hypoxia. Annual Review of Physiology 65:203-230.

Projecto-Garcia, J., C. Natarajan, H. Moriyama, R. E. Weber, A. Fago, Z. A. Cheviron, R. Dudley, J. A. McGuire, C. C. Witt, and J. F. Storz (2013). Repeated elevational transitions in hemoglobin function during the evolution of Andean hummingbirds. Proceedings of the National Academy of Sciences 110:20669-20674.

Pryke, S. R. (2010). Sex chromosome linkage of mate preference and color signal maintains assortative mating between interbreeding finch morphs. Evolution 64:1301-1310.

Przeworski, M. (2002). The signature of positive selection at randomly chosen loci. Genetics 160:1179-1189.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. De Bakker, and M. J. Daly (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics 81:559-575.

Qiao, Q., Q. Wang, X. Han, Y. Guan, H. Sun, Y. Zhong, J. Huang, and T. Zhang (2016). Transcriptome sequencing of Crucihimalaya himalaica (Brassicaceae) reveals how Arabidopsis close relative adapt to the Qinghai-Tibet Plateau. Scientific Reports 6.

Qiu, Q., G. Zhang, T. Ma, W. Qian, J. Wang, Z. Ye, C. Cao, Q. Hu, J. Kim, and D. M. Larkin (2012). The yak genome and adaptation to life at high altitude. Nature Genetics 44:946-949.

Qu, Y., H. Zhao, N. Han, G. Zhou, G. Song, B. Gao, S. Tian, J. Zhang, R. Zhang, and X. Meng (2013). Ground tit genome reveals avian adaptation to living at high altitudes in the Tibetan plateau. Nature Communications 4:2071.

Quandt, K., K. Frech, H. Karas, E. Wingender, and T. Werner (1995). Matlnd and Matlnspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Research 23:4878-4884.

Quina, A., C. Bastos-Silveira, M. Miñarro, J. Ventura, R. Jiménez, O. Paulo, and M. da Luz Mathias (2015). p53 gene discriminates two ecologically divergent sister species of pine voles. Heredity 115:444-451.

Qvarnström, A., and R. I. Bailey (2009). Speciation through evolution of sex-linked genes. Heredity 102:4-15.

Racimo, F., S. Sankararaman, R. Nielsen, and E. Huerta-Sánchez (2015). Evidence for archaic adaptive introgression in humans. Nature Reviews Genetics 16:359-371.

Rajendra, R., D. Malegaonkar, P. Pungaliya, H. Marshall, Z. Rasheed, J. Brownell, L. F. Liu, S. Lutzker, A. Saleem, and E. H. Rubin (2004). Topors functions as an E3 ubiquitin ligase with specific E2 enzymes and ubiquitinates p53. Journal of Biological Chemistry 279:36440-36444.

Rand, D. M., R. A. Haney, and A. J. Fry (2004). Cytonuclear coevolution: the genomics of cooperation. Trends in Ecology & Evolution 19:645-653.

Revsbech, I. G., D. M. Tufts, J. Projecto-Garcia, H. Moriyama, R. E. Weber, J. F. Storz, and A. Fago (2013). Hemoglobin function and allosteric regulation in semi-fossorial rodents (family Sciuridae) with different altitudinal ranges. The Journal of Experimental Biology 216:4264-4271.

Reynafarje, C., J. Faura, D. Villavicencio, A. Curaca, B. Reynafarje, L. Oyola, L. Contreras, E. Vallenas, and A. Faura (1975). Oxygen transport of hemoglobin in high-altitude animals (Camelidae). Journal of Applied Physiology 38:806-810.

Rissanen, E., H. K. Tranberg, J. Sollid, G. E. Nilsson, and M. Nikinmaa (2006). Temperature regulates hypoxia-inducible factor-1 (HIF-1) in a poikilothermic vertebrate, crucian carp (*Carassius carassius*). Journal of Experimental Biology 209:994-1003.

Risso, A., D. Fabbro, G. Damante, and G. Antonutto (2012). Expression of fetal hemoglobin in adult humans exposed to high altitude hypoxia. Blood Cells, Molecules, and Diseases 48:147-153.

Ritchie, M. G. (2007). Sexual selection and speciation. Annual Review of Ecology, Evolution, and Systematics:79-102.

Rockman, M. V., and L. Kruglyak (2006). Genetics of global gene expression. Nature Reviews Genetics 7:862-872.

Ronen, R., N. Udpa, E. Halperin, and V. Bafna (2013). Learning natural selection from the site frequency spectrum. Genetics 195:181-193.

Rosenblum, E. B., C. E. Parent, and E. E. Brandt (2014). The molecular basis of phenotypic convergence. Annual Review of Ecology, Evolution, and Systematics 45:203-226.

Royds, J., S. Dower, E. Qwarnstrom, and C. Lewis (1998). Response of tumour cells to hypoxia: role of p53 and NFkB. Molecular Pathology 51:55.

Rytkönen, K. T., K. A. Vuori, C. R. Primmer, and M. Nikinmaa (2007). Comparison of hypoxia-inducible factor-1 alpha in hypoxia-sensitive and hypoxia-tolerant fish species. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics 2:177-186.

Rytkönen, K. T., T. A. Williams, G. M. Renshaw, C. R. Primmer, and M. Nikinmaa (2011). Molecular evolution of the metazoan PHD–HIF oxygen-sensing system. Molecular Biology and Evolution 28:1913-1926.

Sæther, S. A., G.-P. Sætre, T. Borge, C. Wiley, N. Svedin, G. Andersson, T. Veen, J. Haavie, M. R. Servedio, and S. Bureš (2007). Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. Science 318:95-97.

Sankaran, V. G., and S. H. Orkin (2013). The switch from fetal to adult hemoglobin. Cold Spring Harbor Perspectives in Medicine 3:a011643.

Savolainen, O., M. Lascoux, and J. Merilä (2013). Ecological genomics of local adaptation. Nature Reviews Genetics 14:807-820.

Savolainen, V., M.-C. Anstett, C. Lexer, I. Hutton, J. J. Clarkson, M. V. Norup, M. P. Powell, D. Springate, N. Salamin, and W. J. Baker (2006). Sympatric speciation in palms on an oceanic island. Nature 441:210-213.

Scheinfeldt, L. B., S. Soi, S. Thompson, A. Ranciaro, D. Woldemeskel, W. Beggs, C. Lambert, J. P. Jarvis, D. Abate, and G. Belay (2012). Genetic adaptation to high altitude in the Ethiopian highlands. Genome Biology 13:R1.

Scheinfeldt, L. B., and S. A. Tishkoff (2010). Living the high life: high-altitude adaptation. Genome Biology 11:133.

Schofield, C. J., and P. J. Ratcliffe (2004). Oxygen sensing by HIF hydroxylases. Nature reviews Molecular Cell Biology 5:343-354.

Scott, G. R., S. Egginton, J. G. Richards, and W. K. Milsom (2009). Evolution of muscle phenotype for extreme high altitude flight in the bar-headed goose. Proceedings of the Royal Society B: Biological Sciences 276:3645-3653.

Scott, G. R., T. S. Elogio, M. A. Lui, J. F. Storz, and Z. A. Cheviron (2015). Adaptive modifications of muscle phenotype in high-altitude deer mice are associated with evolved changes in gene regulation. Molecular Biology and Evolution:msv076.

Scott, G. R., P. M. Schulte, S. Egginton, A. L. Scott, J. G. Richards, and W. K. Milsom (2011). Molecular evolution of cytochrome c oxidase underlies high-altitude adaptation in the bar-headed goose. Molecular Biology and Evolution 28:351-363.

Seebacher, F., M. D. Brand, P. L. Else, H. Guderley, A. J. Hulbert, and C. D. Moyes (2010). Plasticity of oxidative metabolism in variable climates: molecular mechanisms. Physiological and Biochemical Zoology 83:721-732.

Semenza, G. L. (2007a). Hypoxia-inducible factor 1 (HIF-1) pathway. Science Signaling 2007:cm8.

Semenza, G. L. (2007b). Life with oxygen. Science 318:62-64.

Semenza, G. L. (2011). Oxygen sensing, homeostasis, and disease. New England Journal of Medicine 365:537-547.

Sermeus, A., and C. Michiels (2011). Reciprocal influence of the p53 and the hypoxic pathways. Cell Death & Disease 2:e164.

Shen, Y.-Y., L. Liang, Z.-H. Zhu, W.-P. Zhou, D. M. Irwin, and Y.-P. Zhang (2010). Adaptive evolution of energy metabolism genes and the origin of flight in bats. Proceedings of the National Academy of Sciences 107:8666-8671.

Shen, Y.-Y., P. Shi, Y.-B. Sun, and Y.-P. Zhang (2009). Relaxation of selective constraints on avian mitochondrial DNA following the degeneration of flight ability. Genome Research 19:1760-1765.

Simonson, T. S., D. A. McClain, L. B. Jorde, and J. T. Prchal (2012). Genetic determinants of Tibetan high-altitude adaptation. Human Genetics 131:527-533.

Simonson, T. S., Y. Yang, C. D. Huff, H. Yun, G. Qin, D. J. Witherspoon, Z. Bai, F. R. Lorenzo, J. Xing, and L. B. Jorde (2010). Genetic evidence for high-altitude adaptation in Tibet. Science 329:72-75.

Slatkin, M. (1975). Gene flow and selection in a two-locus system. Genetics 81:787-802.

Slatkin, M. (2008). Linkage disequilibrium—understanding the evolutionary past and mapping the medical future. Nature Reviews Genetics 9:477-485.

Solaini, G., A. Baracca, G. Lenaz, and G. Sgarbi (2010). Hypoxia and mitochondrial oxidative metabolism. Biochimica et Biophysica Acta (BBA)-Bioenergetics 1797:1171-1177.

Song, D., L.-s. Li, P. R. Arsenault, Q. Tan, A. W. Bigham, K. J. Heaton-Johnson, S. R. Master, and F. S. Lee (2014). Defective Tibetan PHD2 binding to p23 links high altitude adaption to altered oxygen sensing. Journal of Biological Chemistry 289:14656-14665.

Song, S., N. Yao, M. Yang, X. Liu, K. Dong, Q. Zhao, Y. Pu, X. He, W. Guan, and N. Yang (2016). Exome sequencing reveals genetic differentiation due to high-altitude adaptation in the Tibetan cashmere goat (*Capra hircus*). BMC genomics 17:122.

Soria-Carrasco, V., Z. Gompert, A. A. Comeault, T. E. Farkas, T. L. Parchman, J. S. Johnston, C. A. Buerkle, J. L. Feder, J. Bast, and T. Schwander (2014). Stick insect genomes reveal natural selection's role in parallel speciation. Science 344:738-742.

Spitz, F., and E. E. Furlong (2012). Transcription factors: from enhancer binding to developmental control. Nature Reviews Genetics 13:613-626.

Stamatoyannopoulos, G. (2005). Control of globin gene expression during development and erythroid differentiation. Experimental Hematology 33:259-271.

Stapley, J., T. R. Birkhead, T. Burke, and J. Slate (2010). Pronounced inter-and intrachromosomal variation in linkage disequilibrium across the zebra finch genome. Genome Research 20:496-502.

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

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

Stern, D. L., and V. Orgogozo (2009). Is genetic evolution predictable? Science 323:746-751.

Stölting, K. N., R. Nipper, D. Lindtke, C. Caseys, S. Waeber, S. Castiglione, and C. Lexer (2013). Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. Molecular Ecology 22:842-855.

Storchová, R., J. Reif, and M. W. Nachman (2010). Female heterogamety and speciation: reduced introgression of the Z chromosome between two species of nightingales. Evolution 64:456-471.

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

Storz, J. F., and Z. A. Cheviron (2016). Functional Genomic Insights into Regulatory Mechanisms of High-Altitude Adaptation. In Hypoxia. Springer 113-128.

Storz, J. F., and H. Moriyama (2008). Mechanisms of hemoglobin adaptation to high altitude hypoxia. High Altitude Medicine & Biology 9:148-157.

Storz, J. F., J. C. Opazo, and F. G. Hoffmann (2013). Gene duplication, genome duplication, and the functional diversification of vertebrate globins. Molecular Phylogenetics and Evolution 66:469-478.

Storz, J. F., A. M. Runck, S. J. Sabatino, J. K. Kelly, N. Ferrand, H. Moriyama, R. E. Weber, and A. Fago (2009). Evolutionary and functional insights into the mechanism underlying high-altitude adaptation of deer mouse hemoglobin. Proceedings of the National Academy of Sciences 106:14450-14455.

Storz, J. F., G. R. Scott, and Z. A. Cheviron (2010). Phenotypic plasticity and genetic adaptation to high-altitude hypoxia in vertebrates. The Journal of Experimental Biology 213:4125-4136.

Stranger, B. E., A. C. Nica, M. S. Forrest, A. Dimas, C. P. Bird, C. Beazley, C. E. Ingle, M. Dunning, P. Flicek, and D. Koller (2007). Population genomics of human gene expression. Nature Genetics 39:1217.

Subramani, P. A., B. Hameed, and R. D. Michael (2015). Effect of UV-B radiation on the antibody response of fish–Implication on high altitude fish culture. Journal of Photochemistry and Photobiology B: Biology 143:1-4.

Swanson, R. (1984). A unifying concept for the amino acid code. Bulletin of Mathematical Biology 46:187-203.

Tabach, Y., R. Brosh, Y. Buganim, A. Reiner, O. Zuk, A. Yitzhaky, M. Koudritsky, V. Rotter, and E. Domany (2007). Wide-scale analysis of human functional transcription factor binding reveals a strong bias towards the transcription start site. PloS ONE 2:e807.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585-595.

Taylor, W. R. (1986). The classification of amino acid conservation. Journal of Theoretical Biology 119:205-218.

Terova, G., S. Rimoldi, S. Corà, G. Bernardini, R. Gornati, and M. Saroglia (2008). Acute and chronic hypoxia affects HIF-1 α mRNA levels in sea bass *Dicentrarchus labrax*). Aquaculture 279:150-159.

Thompson, M., and C. Jiggins (2014). Supergenes and their role in evolution. Heredity 113:1-8.

Thornton, K. R., and J. D. Jensen (2007). Controlling the false-positive rate in multilocus genome scans for selection. Genetics 175:737-750.

Tirosh, I., A. Weinberger, D. Bezalel, M. Kaganovich, and N. Barkai (2008). On the relation between promoter divergence and gene expression evolution. Molecular Systems Biology 4:159.

Toews, D. P., M. Mandic, J. G. Richards, and D. E. Irwin (2014). Migration, mitochondria, and the yellow‐rumped warbler. Evolution 68:241-255.

Tomasco, I. H., and E. P. Lessa (2011). The evolution of mitochondrial genomes in subterranean caviomorph rodents: Adaptation against a background of purifying selection. Molecular Phylogenetics and Evolution 61:64-70.

Tuch, B. B., D. J. Galgoczy, A. D. Hernday, H. Li, and A. D. Johnson (2008). The evolution of combinatorial gene regulation in fungi. PLoS Biology 6:e38.

Tufts, D. M., C. Natarajan, I. G. Revsbech, J. Projecto-Garcia, F. G. Hoffmann, R. E. Weber, A. Fago, H. Moriyama, and J. F. Storz (2014). Epistasis constrains mutational pathways of hemoglobin adaptation in high-altitude pikas. Molecular Biology and Evolution:msu311.

Turrens, J. F. (2003). Mitochondrial formation of reactive oxygen species. The Journal of Physiology 552:335-344.

Via, S. (2001). Sympatric speciation in animals: the ugly duckling grows up. Trends in Ecology & Evolution 16:381-390.

Via, S. (2012). Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. Philosophical Transactions of the Royal Society B: Biological Sciences 367:451-460.

Via, S., and J. West (2008). The genetic mosaic suggests a new role for hitchhiking in ecological speciation. Molecular Ecology 17:4334-4345.

Visschedijk, A., A. Ar, H. Rahn, and J. Piiper (1980). The independent effects of atmospheric pressure and oxygen partial pressure on gas exchange of the chicken embryo. Respiration physiology 39:33-44.

Wagner, G. P., and V. J. Lynch (2008). The gene regulatory logic of transcription factor evolution. Trends in Ecology & Evolution 23:377-385.

Wagner, G. P., and V. J. Lynch (2010). Evolutionary novelties. Current Biology 20:R48-R52.

Wake, D. B. (1991). Homoplasy: the result of natural selection, or evidence of design limitations? The American Naturalist 138:543-567.

Wang, G.-D., R.-X. Fan, W. Zhai, F. Liu, L. Wang, L. Zhong, H. Wu, H.-C. Yang, S.-F. Wu, and C.-L. Zhu (2014). Genetic convergence in the adaptation of dogs and humans to the high-altitude environment of the tibetan plateau. Genome Biology and Evolution 6:2122-2128.

Wang, G. L., B.-H. Jiang, E. A. Rue, and G. L. Semenza (1995). Hypoxia-inducible factor 1 is a basic-helix-loop-helix-PAS heterodimer regulated by cellular O2 tension. Proceedings of the National Academy of Sciences 92:5510-5514.

Wang, M.-S., Y. Li, M.-S. Peng, L. Zhong, Z.-J. Wang, Q.-Y. Li, X.-L. Tu, Y. Dong, C.-L. Zhu, and L. Wang (2015). Genomic analyses reveal potential independent adaptation to high altitude in Tibetan chickens. Molecular Biology and Evolution:msv071.

Wang, Y., C. Wan, S. R. Gilbert, and T. L. Clemens (2007). Oxygen sensing and osteogenesis. Annals of the New York Academy of Sciences 1117:1-11.

Wang, Z., T. Yonezawa, B. Liu, T. Ma, X. Shen, J. Su, S. Guo, M. Hasegawa, and J. Liu (2011). Domestication relaxed selective constraints on the yak mitochondrial genome. Molecular Biology and Evolution 28:1553-1556.

Webb, J. D., M. L. Coleman, and C. W. Pugh (2009). Hypoxia, hypoxia-inducible factors (HIF), HIF hydroxylases and oxygen sensing. Cellular and Molecular Life Sciences 66:3539-3554.

Weber, R. E. (2007). High-altitude adaptations in vertebrate hemoglobins. Respiratory Physiology and Neurobiology 158:132-142.

Weber, R. E., T. H. Jessen, H. Malte, and J. Tame (1993). Mutant hemoglobins (alpha 119-Ala and beta 55-Ser): functions related to high-altitude respiration in geese. Journal of Applied Physiology 75:2646-2655.

Weger, S., E. Hammer, and R. Heilbronn (2005). Topors acts as a SUMO-1 E3 ligase for p53 in vitro and in vivo. FEBS Letters 579:5007-5012.

Weir, B. S., and C. C. Cockerham (1984). Estimating F-statistics for the analysis of population structure. Evolution:1358-1370.

Weir, E. K., J. López-Barneo, K. J. Buckler, and S. L. Archer (2005). Acute oxygen-sensing mechanisms. New England Journal of Medicine 353:2042-2055.

Welch, A. J., O. C. Bedoya-Reina, L. Carretero-Paulet, W. Miller, K. D. Rode, and C. Lindqvist (2014). Polar bears exhibit genome-wide signatures of bioenergetic adaptation to life in the arctic environment. Genome Biology and Evolution 6:433-450.

Wenger, R. H., D. P. Stiehl, and G. Camenisch (2005). Integration of oxygen signaling at the consensus HRE. Science Signaling 2005:re12.

Whitehead, A., and D. L. Crawford (2006). Neutral and adaptive variation in gene expression. Proceedings of the National Academy of Sciences 103:5425-5430.

Whitlock, M. C., and D. E. McCauley (1999). Indirect measures of gene flow and migration: FST≠ 1/(4Nm+ 1). Heredity 82:117-125.

Wilson, R. E., M. Eaton, K. G. McCracken, and N. y Sudamérica (2008). color divergence among cinnamon Teal (*Anas cyanoptera*) subspecies from North America and South America. Ornitologia Neotropical 19:307-314.

Wilson, R. E., M. D. Eaton, S. A. Sonsthagen, J. L. Peters, K. P. Johnson, B. Simarra, and K. G. McCracken (2011). Speciation, subspecies divergence, and paraphyly in the Cinnamon Teal and Blue-winged Teal. The Condor 113:747-761.

Wilson, R. E., J. L. Peters, and K. G. McCracken (2013). Genetic and phenotypic divergence between low- and high- altitude populations of two recently diverged Cinnamon teal subspecies. Evolution 67:170-184.

Wilson, R. E., T. H. Valqui, and K. G. McCracken (2010). Ecogeographic variation in Cinnamon Teal (*Anas cyanoptera*) along elevational and latitudinal gradients. Ornithological Monographs 67:141-161.

Wit, P., and S. R. Palumbi (2013). Transcriptome‐wide polymorphisms of red abalone (*Haliotis rufescens*) reveal patterns of gene flow and local adaptation. Molecular Ecology 22:2884-2897.

Wittenhagen, L. M., and S. O. Kelley (2003). Impact of disease-related mitochondrial mutations on tRNA structure and function. Trends in Biochemical Sciences 28:605-611.

Wittkopp, P. J., and G. Kalay (2011). Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nature Reviews Genetics 13:59-69.

Wolff, J. N., E. D. Ladoukakis, J. A. Enríquez, and D. K. Dowling (2014). Mitonuclear interactions: evolutionary consequences over multiple biological scales. Philosophical Transactions of the Royal Society B: Biological Sciences 369:20130443.

Wong, G. K.-S., B. Liu, J. Wang, Y. Zhang, X. Yang, Z. Zhang, Q. Meng, J. Zhou, D. Li, and J. Zhang (2004). A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. Nature 432:717-722.

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

Wu, T., and B. Kayser (2006). High altitude adaptation in Tibetans. High Altitude Medicine & Biology 7:193-208.

Xiang, K., Y. Peng, Z. Yang, X. Zhang, C. Cui, H. Zhang, M. Li, Y. Zhang, T. Wu, and H. Chen (2013). Identification of a Tibetan-specific mutation in the hypoxic gene EGLN1 and its contribution to high-altitude adaptation. Molecular Biology and Evolution 30:1889-1898.

Xing, J., T. Wuren, T. S. Simonson, W. S. Watkins, D. J. Witherspoon, W. Wu, G. Qin, C. D. Huff, L. B. Jorde, and R.-L. Ge (2013). Genomic analysis of natural selection and phenotypic variation in high-altitude mongolians. PLoS Genetics 9:e1003634.

Xu, S., S. Li, Y. Yang, J. Tan, H. Lou, W. Jin, L. Yang, X. Pan, J. Wang, and Y. Shen (2011). A genome-wide search for signals of high-altitude adaptation in Tibetans. Molecular Biology and Evolution 28:1003-1011.

Xu, S., J. Luosang, S. Hua, J. He, A. Ciren, W. Wang, X. Tong, Y. Liang, J. Wang, and X. Zheng (2007). High altitude adaptation and phylogenetic analysis of tibetan horse based on the mitochondrial genome. Journal of Genetics and Genomics 34:720-729.

Yan, L., C. Zhang, T. Pan, W. Zhou, C. Hu, Q. Chang, and B. Zhang (2015). Mitochondrial genome of the *Anas acuta* (Anatidae: Anas). Mitochondrial DNA 26:297-298.

Yang, J., W.R. Li, F.H. Lv, S.G. He, S.L. Tian, W.F. Peng, Y.W. Sun, Y.X. Zhao, X.L. Tu, and M. Zhang (2016). Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. Molecular Biology and Evolution 33:2576-2592.

Yang, Y., L. Wang, J. Han, X. Tang, M. Ma, K. Wang, X. Zhang, Q. Ren, Q. Chen, and Q. Qiu (2015). Comparative transcriptomic analysis revealed adaptation mechanism of *Phrynocephalus erythrurus*, the highest altitude lizard living in the Qinghai-Tibet plateau. BMC Evolutionary Biology 15:101.

Yeaman, S., and M. C. Whitlock (2011). The genetic architecture of adaptation under migration–selection balance. Evolution 65:1897-1911.

Yi, X., Y. Liang, E. Huerta-Sanchez, X. Jin, Z. X. P. Cuo, J. E. Pool, X. Xu, H. Jiang, N. Vinckenbosch, and T. S. Korneliussen (2010). Sequencing of 50 human exomes reveals adaptation to high altitude. Science 329:75-78.

Yu, L., X. Wang, N. Ting, and Y. Zhang (2011). Mitogenomic analysis of Chinese snub-nosed monkeys: evidence of positive selection in NADH dehydrogenase genes in high-altitude adaptation. Mitochondrion 11:497-503.

Zhang, G., B. Li, C. Li, M. T. P. Gilbert, E. D. Jarvis, and J. Wang (2014). Comparative genomic data of the avian phylogenomics project. GigaScience 3:26.

Zhang, Q., W. Gou, X. Wang, Y. Zhang, J. Ma, H. Zhang, Y. Zhang, and H. Zhang (2016). Genome resequencing identifies unique adaptations of Tibetan chickens to hypoxia and high-dose ultraviolet radiation in high-altitude environments. Genome Biology and Evolution:evw032.

Zhang, Z.Y., B. Chen, D.J. Zhao, and L. Kang (2013). Functional modulation of mitochondrial cytochrome c oxidase underlies adaptation to high-altitude hypoxia in a Tibetan migratory locust. Proceedings of the Royal Society B: Biological Sciences 280:20122758.

Zhao, D., Z. Zhang, A. Cease, J. Harrison, and L. Kang (2013a). Efficient utilization of aerobic metabolism helps Tibetan locusts conquer hypoxia. BMC Genomics 14:1.

Zhao, X., N. Wu, Q. Zhu, U. Gaur, T. Gu, and D. Li (2016). High-altitude adaptation of Tibetan chicken from MT-COI and ATP-6 perspective. Mitochondrial DNA Part A 27:3280-3288.

Zhao, Y., J.L. Ren, M.Y. Wang, S.T. Zhang, Y. Liu, M. Li, Y.B. Cao, H.Y. Zu, X.C. Chen, and C.-I. Wu (2013b). Codon 104 variation of p53 gene provides adaptive apoptotic responses to extreme environments in mammals of the Tibet plateau. Proceedings of the National Academy of Sciences 110:20639-20644.

Zhou, T., X. Shen, D. M. Irwin, Y. Shen, and Y. Zhang (2014). Mitogenomic analyses propose positive selection in mitochondrial genes for high-altitude adaptation in galliform birds. Mitochondrion 18:70-75.

Zhou, W., C. Zhang, T. Pan, L. Yan, C. Hu, C. Xue, Q. Chang, and B. Zhang (2015). The complete mitochondrial genome of *Anas poecilorhyncha* (Anatidae: Anas). Mitochondrial DNA 26:265-266.

Zink, R. M., and G. F. Barrowclough (2008). Mitochondrial DNA under siege in avian phylogeography. Molecular Ecology 17:2107-2121.

Zwickl, D. (2006). GARLI: genetic algorithm for rapid likelihood inference. See http://www.bio.utexas.edu/faculty/antisense/garli/Garli.html.