Parallel evolution in the major haemoglobin genes of eight species of Andean waterfowl


*Institute of Arctic Biology, Department of Biology and Wildlife, and University of Alaska Museum, University of Alaska Fairbanks, Fairbanks, AK 99775, USA, †Illinois Natural History Survey, Champaign, IL 61820, USA, ‡Patagonia Outfitters, Perez 662, San Martín de los Andes, Neuquén 8370, Argentina, §Centro de Ornitología y Biodiversidad (CORBIDI), Sta. Rita 117, Urbana Huertos de San Antonio, Surco, Lima 33, Perú, ¶Museum of Natural Science, Louisiana State University, Baton Rouge, LA 70803, USA, **Department of Biology, Boston University, Boston, MA 02215, USA

Abstract

Theory predicts that parallel evolution should be common when the number of beneficial mutations is limited by selective constraints on protein structure. However, confirmation is scarce in natural populations. Here we studied the major haemoglobin genes of eight Andean duck lineages and compared them to 115 other waterfowl species, including the bar-headed goose (Anser indicus) and Abyssinian blue-winged goose (Cyanochen cyanopterus), two additional species living at high altitude. One to five amino acid replacements were significantly overrepresented or derived in each highland population, and parallel substitutions were more common than in simulated sequences evolved under a neutral model. Two substitutions evolved in parallel in the αA subunit of two (Ala-8) and five (Thr-77) taxa, and five identical βA subunit substitutions were observed in two (Ser-β4, Glu-β94, Met-β133) or three (Ser-β13, Ser-β116) taxa. Substitutions at adjacent sites within the same functional protein region were also observed. Five such replacements were in exterior, solvent-accessible positions on the A helix and AB corner of the αA subunit. Five others were in close proximity to inositolpentaphosphate binding sites, and two pairs of independent replacements occurred at two different αβ intersubunit contacts. More than half of the substitutions in highland lineages resulted in the acquisition of serine or threonine (18 gains vs. 2 losses), both of which possess a hydroxyl group that can hydrogen bond to a variety of polar substrates. The patterns of parallel evolution observed in these waterfowl suggest that adaptation to high-altitude hypoxia has resulted from selection on unique but overlapping sets of one to five amino acid substitutions in each lineage.

Keywords: Altiplano, Anatidae, haemoglobin, high-altitude hypoxia, oxygen affinity, parallel evolution, Patagonia, puna, South America, waterfowl

Received 2 April 2009; revision received 2 July 2009; accepted 10 August 2009

Introduction

Parallel evolution, characterized by similar adaptive responses to a particular set of ecological conditions, is of widespread interest (Zhang & Kumar 1997; Wood et al. 2005; Arendt & Reznick 2008). A few recent studies have examined the molecular basis of parallel adaptation (Kornegay et al. 1994; Golding & Dean 1998; Colosimo et al. 2005; Jost et al. 2008; Rokas & Carroll 2008), but such studies are still uncommon. Using simulations and extreme value theory, Orr (2005) concluded that parallel fixation of identical substitutions should be common when the number of possible beneficial mutations is limited, regardless of the distribution of fitness effects among alleles. Among closely related lineages, in

© 2009 Blackwell Publishing Ltd
which antecedent character states are identical, parallel evolution is thus expected to involve identical but historically independent nucleotide and amino acid substitutions. In addition, recurrent substitutions (see West-Eberhard 2003) with similar phenotypic effects will likely occur at different, but frequently adjacent, positions within the same gene or protein molecule if the number of mechanistic pathways by which adaptation can occur are limited.

Despite the importance of parallel evolution as a fundamental question in evolutionary biology, few studies have sampled a sufficient diversity of taxa to rigorously examine the theory. Parallel evolution has been observed in viral and bacterial populations subjected to directional selection (Bull et al. 1997; Wichman et al. 1999; Riehle et al. 2001; Bollback & Huelsenbeck 2009), but studies of natural populations have generally focused on a small number of populations and have lacked sufficient taxonomic breadth to confidently infer ancestral character states. Most phylogenetic studies have lacked the intra-specific sampling required to identify polymorphisms segregating in a population. Likewise, most studies have not incorporated null models to explicitly determine the expected number of parallel changes (but see Zhang & Kumar 1997; Rokas & Carroll 2008; Bollback & Huelsenbeck 2009). Finally, the genetic basis of many traits is complex or poorly understood, and the individual alleles or polymorphisms underlying adaptation in one population often show no association with the same trait in other closely related populations (Gilchrist & Partridge 1999; Hoekstra & Nachman 2003; Nachman et al. 2003). Thus, a comparative analysis of molecular evolution in a well-characterized protein coded by a small number of genes subject to well-defined selection pressures across contrasting ecological environments provides a fruitful approach to exploring the pattern and frequency of parallel adaptive evolution in non-traditional model systems.

High-altitude hypoxia, haemoglobins, and waterfowl

Hypoxia is one of the most important factors confronting organisms in high-altitude regions. At high elevations such as 4000 meters encountered in the Andes, Himalayas, or Ethiopian Plateau, the partial pressure of oxygen ($pO_2$) is approximately 60% of that at sea level. The low $pO_2$ of inspired air reduces the oxygen ($O_2$) saturation of arterial blood, which in turn can result in a reduced supply of $O_2$ to the tissues (Hopkins & Powell 2001; Hornbein & Schoene 2001; Hochachka & Somero 2002; Beall 2006). Several studies indicate that substitutions increasing the $O_2$-affinity of haemoglobins play an important role in mitigating the effects of chronic hypoxia in populations adapted to high-altitude environments (Perutz 1983; Storz et al. 2007, 2009; Weber 2007; Storz & Moriyama 2008).

Haemoglobins are the primary blood $O_2$-transport protein in vertebrates and one of the best-studied macromolecular proteins. Most vertebrate haemoglobins are tetrameric proteins composed of four polypeptides, two $\alpha$ subunits and two $\beta$ subunits, each linked to a heme group that binds cooperatively and reversibly with molecular $O_2$. The $\alpha$ and $\beta$ subunits are coded on different chromosomes and, in birds such as the chicken and ducks, include three linked $\alpha$ chain genes ($\alpha_1$, $\alpha_2$, $\alpha_3$) and four linked $\beta$ chain genes ($\beta_1$, $\beta_2$, $\beta_3$, $\beta_4$), which are expressed in various combinations to produce different isoforms that vary in their affinity for $O_2$. The major haemoglobin (HbA) is composed of two $\alpha A$ subunits and two $\beta A$ subunits and is the most common isoform in the haematocrit of adult birds; the minor isoform (HbD) of adult birds is expressed at lower concentrations (Rowley & Ratcliffe 1988; Bulgarella et al. 2009).

Reversible $O_2$ loading and unloading results from small changes in the tertiary structure at the hemes and a large change in quaternary structure, which coincide with a rotation and translation of one $\beta$ dimer relative to the other (Perutz 1989). The deoxy or tense (T-state) structure has a low affinity for $O_2$ and high affinity for allosteric effectors such as protons, chloride, $CO_2$, and organic phosphate. The oxy or relaxed (R-state) structure generally has a much lower affinity for these allosteric effectors, but a high affinity for $O_2$. Haemoglobin $O_2$ affinity can be modified by amino acid substitutions that decrease the stability of the low-affinity deoxy structure, thereby shifting the allosteric equilibrium in favour of the high-affinity oxy structure (Perutz 1989), or by changing the affinity of haemoglobin for allosteric effectors, which stabilize the deoxy structure with salt bridges between the subunits. The principal allosteric effector in birds is inositolpentaphosphate (IPP), which binds to positively charged residues in the central cavity between the N- and C-termini of the $\alpha$ and $\beta$ subunits (Wang et al. 2000). $O_2$ affinity can thus be modified also by changing the net positive charge of IPP binding regions.

Two species of waterfowl have featured prominently in studies of haemoglobin adaptation. The bar-headed goose (Anser indicus) breeds at high elevations in Asia and migrates over the Himalayas at elevations >9000 m, where the $pO_2$ is ~30% of that at sea level (Scott & Milson 2006, 2007). Increased $O_2$ loading of haemoglobin in the lungs is achieved by a Pro $\rightarrow$ Ala-$\alpha 119$ substitution. The small R-group side chain of Ala-$\alpha 119$ eliminates an $\alpha 1\beta 1$ intersubunit van der Waals contact, destabilizing the deoxy (T-state) structure and increasing $O_2$ affinity (Jessen et al. 1991; Weber et al. 1993). Amino acid replacements resulting in small tertiary or...
quaternary structural changes also occur at α18, α63, and β125 (Liang et al. 2001a). Bar-headed goose thus achieves 50% saturation (P50) at lower pO2 than in lowland species such as greylag goose (Anser anser): 29.7 mmHg vs 39.5 mmHg for whole blood (Petschow et al. 1977). The Andean goose (Chloephaga melanoptera), which is not a true goose but instead belongs to a clade of ecologically convergent, goose-like ducks called sheldgeese (Livezey 1986), has independently evolved essentially the same mechanism to cope with chronic hypoxia in the Andes, but through a different amino acid substitution in a different gene. In this species, Leu → Ser-β55 also results in a smaller R-group side chain that loosens the same α1β1 intersubunit contact affected by the bar-headed goose substitution at α119. Andean goose blood has a P50 of 33.9 mmHg (Hall et al. 1936). Parallel evolution of increased haemoglobin-O2 affinity in these two distantly related highland waterfowl species thus involves the same basic mechanism, but the underlying sequence changes are different in each species.

Study design

We studied the molecular evolution of the major haemoglobin in eight Andean duck lineages that have independently colonized the same high-altitude wetlands and puna grasslands of the Altiplano and inter-Andean valleys of South America. Vouchered specimens were collected from a 6000 kilometer transect of the central and southern Andes from northern Peru to the Strait of Magellan (Fig. 1), sampling 574 individual ducks from eight paired lowland and highland populations or sister taxa experiencing contrasting pO2 at elevations ranging from sea level to 4600 m (Table S1). We sequenced the complete coding region and intervening intron sequences of their αA and βA haemoglobin subunits (HBA2 and HBB) and compared them to 115 other species of waterfowl (Table S1), including the bar-headed goose and the Abyssinian blue-winged goose (Cygnus cyanopterus); the latter is a highland species endemic to the Ethiopian Plateau.

Using two large datasets comprising 862 αA subunit and 683 βA subunit sequences representing most of the world’s waterfowl species, we identified amino acid polymorphisms that were overrepresented in highland populations (i.e. showing significantly elevated FST between paired lowland and highland populations of the same species) as well as amino acid residues exhibiting fixed differences between highland and lowland sister taxa. Maximum likelihood gene trees were constructed for the αA and βA haemoglobin subunits, and amino acid substitutions were mapped onto an independent mtDNA phylogeny. Each amino acid replacement was located on the oxy (R-state) crystal structure of the greylag goose haemoglobin (Liang et al. 2001a). The interatomic distances between each observed replacement and α1β1 inter-subunit contacts, IPP binding sites, and the heme groups were measured and characterized within the context of the haemoglobin literature, noting in particular the type of R-group substitution that occurred with each observed amino acid replacement.

Finally, we used simulated datasets incorporating the stochasticity of the coalescent and substitutional...
processes to calculate the probability of observing a given number of identical nonsynonymous codon replacements in ten replicate lineages. Simulations were conditioned on the same number of sampled sequences and empirical estimates of theta (Θ) and rate variation among sites obtained from the raw data. Additional analyses contrasting linkage disequilibrium, heterozygosity, and allelic migration rates between the αA and βA subunits and five autosomal reference loci are presented elsewhere (McCracken et al. 2009a,b).

Materials and methods

Specimen collecting and DNA sequencing

Localities for each specimen are given in Fig. 1 and Table S1. Waterfowl collected at elevations >2000 meters were categorized as ‘highland’, and specimens collected at <2000 meters were categorized as ‘lowland’ following established criteria for studies of high-altitude adaptation in humans (Hornbein & Schoene 2001). Genomic DNA was isolated from muscle using DNeasy Tissue Kits (QIAGEN). Primers flanking the start and stop codons for the αA and βA haemoglobin subunit genes were designed using duck, chicken, and other DNA sequences in GenBank (Reitman et al. 1993; Flint et al. 2001). PCR was performed using AmpliTaq Gold PCR Master Mix (Applied Biosystems) and standardized thermal cycling protocols. The complete coding region of the αA subunit comprising three exons and two introns (668–711 bp) was sequenced as a single fragment, or in two overlapping fragments. The βA subunit (1567–1630 bp) was sequenced using nested or half-nested PCR. Product from an initial PCR spanning the complete βA coding region was used as the template for a second PCR using multiple overlapping combinations of internal and end primers. The αA and βA haemoglobin subunit primers are provided in Table S2. The gametic phase of each heterozygous site was determined for the eight Andean lineages by means of allele-specific PCR in combination with the software PHASE 2.1 (Stephens et al. 2001), as described by McCracken et al. (2009a,b). Sequences were aligned by eye, and the alignments are provided as supplementary NEXUS data files. Sequences are deposited in GenBank (accession numbers αA subunit FJ617587–FJ617702 and GQ271002–GQ271747; βA subunit FJ617703–FJ617816 and GQ271748–GQ272322).

F<sub>ST</sub> calculations

F<sub>ST</sub> based on the average number of pairwise differences (π) within and between each lowland and highland population were calculated for each polymorphic position in the αA and βA subunits using Arlequin 3.0 (Excoffier et al. 2005).

Probability of observing n identical codon replacements in ten replicate lineages

We used simulated datasets incorporating the stochasticity of coalescent and substitutional processes to calculate the probability of observing n identical codon replacements in ten replicate lineages with a given number of sampled sequences and empirical estimates of theta (Θ) and rate variation among sites. We first simulated genealogies using the software MS (Hudson 2002) and then evolved sequences on each genealogy using Seq-Gen 1.3.2 (Rambaut & Grassly 1997).

Ten replicate MS analyses were performed, each simulating 1000 independent genealogies, conditioned on the observed Θ values for the αA haemoglobin coding sequences of each highland lineage and the empirical sample sizes (n = 22 to 140 alleles; Table S3). Theta values for the five Andean dabbling duck species were obtained with LAMARC (Kuhner 2006) using two-population Bayesian coalescent models incorporating migration and recombination with 1 million recorded genealogies sampled every 50 steps and a burnin of 100 000 (10%) (McCracken et al. 2009a,b). Thetas for Andean ruddy duck (Oxyura j. ferruginea), Andean goose, and torrent duck (Merganetta armata) were obtained using the same search strategy but using a one-population coalescent model with recombination and no migration (Table S3). Empirical estimates of Θ were not available for bar-headed goose and blue-winged goose, because sufficient samples of these species do not exist. Two random values of Θ within the range of the eight observed values were selected, and we set the number of simulated samples for these two species as the average of the number of alleles sampled for the other eight species. Theta is scaled per gene in MS and was therefore calculated by multiplying the LAMARC Θ, which is scaled per site, by the sequence length (l = 423).

Nucleotide sequences of length 423 bp were simulated on each genealogy using Seq-Gen 1.3.2 and the best-fit substitution model identified with Modeltest 3.7 (Posada & Crandall 1998), which for αA haemoglobin coding sequences was the K80 model with equal base frequencies and a transition/transversion ratio equal to 9.1950. Relative substitution rates for 1st, 2nd, and 3rd codon positions were set to 0.770440, 0.343917, and 1.000000, respectively, using the mean pairwise K80 genetic distances between the snow goose (Anser caerulescens) sequence and sequences for the five Andean dabbling ducks. The same ancestral sequence inferred from the base of the αA subunit gene tree in Fig. 2

© 2009 Blackwell Publishing Ltd
using PAML 3.15 (Yang 1997) was used as the starting point for each simulated dataset. Branch lengths were scaled to $\Theta$ per site (Table S3) so that the overall numbers of segregating sites approximately matched the observed numbers of polymorphic sites in the empirical data. $P$-values were calculated from the posterior distribution of simulated sequences as the probability of obtaining $n$ identical codon replacements among ten independent lineages drawn one each from the simulated sets for each taxon.

Finally, we assessed whether the parallel amino acid substitutions we observed were significantly concentrated in taxa that lived at high elevation using the concentrated changes test (Maddison 1990) as implemented in MacClade 4.06 (Maddison & Maddison 2000). Elevation (highland or lowland) was treated as the independent character, and the observed numbers of gains and losses at individual amino acid positions were used as the dependent characters. Using the $\alpha A$ and $\beta A$ subunit genealogies comprising 123 and 93 taxa with the species-level topologies shown in Fig. 2, we identified positions that had the same amino acid substitution independently derived in at least two lineages (i.e., parallel substitutions). For positions in which more than one amino acid substitution type exhibited a parallel change, we repeated the test using all possible derived character states. We identified 15 such substitutions for the $\alpha A$ subunit and 14 for the $\beta A$ subunit. Significance was determined using the Bonferroni correction for multiple tests (Table S3).

**Structural analyses of the greylag goose haemoglobin**

The position of each amino acid replacement observed in highland waterfowl lineages was located on the oxy (R-state) crystal structure of greylag goose haemoglobin (Liang et al. 2001a). Inter-atomic distances were calculated using Cn3D 4.1 (National Institutes of Health, Bethesda, MD) and Swiss-PdbViewer 3.7 (Guex & Peitsch 1997) using a maximum van der Waals distance of 4.1 Å or less (Dall’Acqua et al. 1998). Changes in properties such as polarity, isoelectric point, molecular volume, and normalized van der Waals volume were calculated using published values (Grantham 1974), and protein structures were illustrated using Cn3D 4.1 or Python Molecular Viewer 1.5.2 (Scripps Research Institute, La Jolla, CA, USA). The $\beta A$ subunit positions $\beta1, 2, 82, 104, 135, 139, 143, 144,$ and $146$ comprise the 1st IPP-binding site (Zhang et al. 1996; Wang et al. 2000; Liang et al. 2001a,b; Liu et al. 2001), and the $\alpha A$ subunit positions $\alpha1, 95, 99, 134, 137, 138,$ and $141$ comprise the 2nd IPP-binding site (Tamburrini et al. 2000; Riccio et al. 2001).

**Phylogenetic analyses**

Maximum likelihood gene trees for the $\alpha A$ and $\beta A$ haemoglobin subunits (including both exons and introns) were estimated independently; 123 of 145 waterfowl species (85%) were analyzed for the $\alpha A$ subunit, and 93 species (64%) were analyzed for the $\beta A$ subunit. Gene trees for each locus were constructed using PhyML 3.0.
(Guindon & Gascuel 2003) with nearest neighbour interchange (NNI) and subtree pruning and regrafting (SPR) branch rearrangements. The best-fit substitution model for each locus was HKY+I+G, as determined using the Bayesian Information Criterion (BIC) (Schwarz 1978) as implemented in ModelTest 3.7 (Posada & Crandall 1998) and PAUP* (Swofford 2002). The PhyML gene tree topologies did not differ substantially from trees constructed using neighbour-joining with maximum likelihood distances in PAUP* or using mixed models in MrBayes 3.1 (Ronquist & Huelsenbeck 2003) with separate rate parameters for 1st, 2nd, and 3rd codon positions, and introns 1 and 2, modelled separately. Analysis of character evolution was conducted by mapping amino acid replacements onto a previously published mtDNA phylogeny of dabbling ducks (Johnson & Sorenson 1999) using unweighted parsimony and MacClade 4.06.

**Results**

*Parallel amino acid replacements in Andean ducks*

Molecular and morphological waterfowl phylogenies (Livezey 1986; Harshman 1996; Johnson & Sorenson 1999; McCracken et al. 1999; Sorenson et al. 1999; Donne-Gousse et al. 2002; Bulgarella et al. 2010) indicate that each of the ten highland waterfowl taxa considered here is an independent evolutionary lineage; no two highland forms are closely related as sister taxa (Figs 2 and 3).

Our analysis revealed numerous amino acid replacements in the αA and βA subunit genes that were present only in high-elevation waterfowl populations. Seventeen amino acid replacements at nine different positions in the two genes showed significant frequency differences between paired lowland and highland populations, with significantly elevated $F_{ST}$ values ranging from 0.06 to 1.00 (Fig. 4, Table 1, Fig. S1). Three additional amino acid replacements were rare alleles present only in highland populations ($F_{ST} \leq 0.01$). Five replacements were derived in species with highland populations, but lowland populations of the same species were also fixed for the novel amino acid ($F_{ST} = 0.00$). Two replacements were derived in the Abyssinian blue-winged goose but not its closest relatives. Similarly, five additional replacements were observed in bar-headed goose but not in other species of the same genus, including a Gly → Ala-α12 replacement not reported previously. In summary, all ten highland lineages possessed derived amino acid replacements, and most had multiple replacements across the two haemoglobin subunits (Fig. 4). The total number of amino acid replacements for any single highland species, however, was no greater than five. Two of the ten highland species had five replacements, three species each had three or four, respectively, and two species had one or two replacements (mean = 3.4; mode = 3 or 4; Fig. 4). Interestingly, bar-headed goose, Andean goose, and blue-winged goose, which likely have inhabited highland regions for longer time periods than the various dabbling duck lineages (*Anas* spp.), did not exhibit a greater number of amino acid replacements.

Parallel amino acid replacements were observed at 7 (33%) of 21 positions with derived substitutions (Fig. 4). Two amino acid replacements (Ala-α8 and Thr-α77) evolved in parallel on the αA subunit in two and five highland taxa, respectively, and five amino acids evolved in parallel on the βA subunit in two (Ser-β4, Glu-β94, Met-β133) or three (Ser-β13, Ser-β116) highland taxa. Each parallel acquisition of the same amino acid by highland taxa involved the same nucleotide substitution at the same codon position.

The parallel codon replacements we observed were not convergent per se because they evolved from the same antecedent character states (Wiens et al. 2003). Identical codon replacements may have resulted, in part, from translational selection associated with differences in tRNA abundance (Heger & Pontin 2007). Trans-specific ancestral polymorphisms, however, appear to have contributed little to parallel evolution of Andean waterfowl haemoglobins, as few amino acids that were derived in highland populations occurred elsewhere in lowland waterfowl populations sampled worldwide (Table 1; Fig. S1). No such residues occurred on the αA subunit, and those that were shared between different species on the βA subunit occurred in distantly related genera on different background sequences. Within the genus *Anas*, for example, none of the amino acids that were overrepresented in the five Andean lineages were found in 18 other lowland *Anas* species (Fig. 3; Fig. S1).

Simulations indicate that the probability of observing 19 parallel codon substitutions at seven different positions as in the observed data is exceedingly low under a null model of neutral evolution. Using ten replicate populations simulated 1000 times each with MS (Hudson 2002) and Seq-Gen 1.3.2 (Rambaut & Grassly 1997; see Materials and methods, Table S3), parallel nonsynonymous substitutions at a single codon position were observed in two lineages in 15% of replicates, whereas parallel substitutions at two different codon positions, each in two lineages, occurred in only 1.9% of replicates. Parallel substitutions at a single codon position in three or more lineages occurred in 0.001% of replicates. Each simulation was initiated from the same ancestral sequence, conditioned on theta ($\Theta$)
values for each highland lineage, empirical sample sizes, and the empirical level of rate variation among sites.

The concentrated changes test (Maddison 1990) further demonstrates that parallel substitutions of Ala-α8, Thr-α77, Ser-β4, Ser-β13, Glu-β94, Ser-β116, and Met-

© 2009 Blackwell Publishing Ltd
β133 were significantly associated with highland taxa; all but two of the these tests (Ser-β4 and Glu-β94) remained significant after correcting for multiple statistical tests on all parallel substitutions observed in each gene (Table S3). Thus there is strong evidence that the same amino acid replacements have occurred more times than expected by chance in highland taxa, and that these likely represent molecular adaptation to high-altitude hypoxia.

**Structural and functional patterns of parallelism**

Our datasets also include numerous examples in which unique amino acid replacements were observed at adjacent sites within the same functional regions of the haemoglobin protein (Table 1, Fig. 5). Five different replacements were located at exterior, solvent-accessible positions on the A helix and AB corner of the αA subunit (Fig. 5a), and five more occurred on the A and E helices of the βA subunit (Fig. 5b, c). Five different replacements occurred within van der Waals distance of IPP-binding sites at the N- and C-termini of the α and β subunits (Fig. 5d, e), and two pairs of independent amino acid replacements were observed at two different αβ intersubunit contacts (Fig. 5f, g). Parallel substitutions thus involved not only identical codon substitutions but also substitutions at adjacent amino acid positions on the same folded polypeptide and interacting positions (at αβ intersubunit contacts) in different polypeptide subunits coded on separate chromosomes.

**Discussion**

Thr-α77 provides the most striking example of parallel evolution in our dataset, evolving independently in five different highland taxa on two continents (Fig. 4). This residue probably plays a functional role in the second IPP binding site identified by Tamburrini _et al._ (2000) at the N- and C-termini of the two alpha subunits (Fig. 5e). Inositolpentaphosphate is a negatively charged ligand, and the IPP binding sites at α1, 95, 99, 134, 137, 138, and 141 are positively charged (Tamburrini _et al._ 2000; Riccio _et al._ 2001). Ala → Thr-α77 is a non-polar to polar change that results in the addition of a hydroxyl group and increased normalized van der Waals volume, thus adding positive charge to its surrounding environment. One possibility is that the
Table 1 Detailed information on the occurrence of individual amino acid replacements in lowland and highland populations of Andean and other waterfowl (parallel changes are shown in bold text)

<table>
<thead>
<tr>
<th>Pos.</th>
<th>Amino acid</th>
<th>Structural position</th>
<th>Highland taxa</th>
<th>Amino acids in other Anatidae</th>
<th>$F_{ST}$</th>
<th>High altitude taxa only</th>
<th>Notes/exceptions</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>Thr</td>
<td>A helix</td>
<td>Crested duck</td>
<td>Ala</td>
<td>0.46</td>
<td>Yes</td>
<td>—</td>
</tr>
<tr>
<td>8</td>
<td>Ala</td>
<td>A helix</td>
<td>Andean goose, crested duck</td>
<td>Thr (Ser, Asn)</td>
<td>0.06, 1.00</td>
<td>Yes</td>
<td>—</td>
</tr>
<tr>
<td>9</td>
<td>Ser</td>
<td>A helix</td>
<td>Cinnamon teal</td>
<td>Asn</td>
<td>0.94</td>
<td>Yes</td>
<td>—</td>
</tr>
<tr>
<td>12</td>
<td>Ala</td>
<td>A helix</td>
<td>Bar-headed goose</td>
<td>Gly</td>
<td>n/a</td>
<td>Yes</td>
<td>—</td>
</tr>
<tr>
<td>18</td>
<td>Ser</td>
<td>AB corner</td>
<td>Bar-headed goose</td>
<td>Gly</td>
<td>n/a</td>
<td>Yes</td>
<td>—</td>
</tr>
<tr>
<td>63</td>
<td>Val</td>
<td>E helix</td>
<td>Bar-headed goose</td>
<td>Ala (Val)</td>
<td>n/a</td>
<td>No</td>
<td>Stictonetta, Nettapus pulchellus also have Val</td>
</tr>
<tr>
<td>77</td>
<td>Thr</td>
<td>2nd IPP binding site</td>
<td>Andean goose, blue-winged goose, torrent duck, puna teal, speckled teal</td>
<td>Ala (Gly, Ser)</td>
<td>1.00, n/a, 0.38, 0.49, 0.91</td>
<td>Yes</td>
<td>—</td>
</tr>
<tr>
<td>111</td>
<td>Thr</td>
<td>2β1 intersubunit contact</td>
<td>Puna teal</td>
<td>Ile</td>
<td>0.01*</td>
<td>Yes</td>
<td>rare highland allele in puna teal (n = 2 heterozygotes)</td>
</tr>
<tr>
<td>119</td>
<td>Ala</td>
<td>2β1 intersubunit contact</td>
<td>Bar-headed goose</td>
<td>Pro</td>
<td>n/a</td>
<td>Yes</td>
<td>—</td>
</tr>
<tr>
<td>13</td>
<td>Ser</td>
<td>1st IPP binding site</td>
<td>Crested duck, yellow-billed pintail</td>
<td>Thr (Ser)</td>
<td>0.00, 0.002*</td>
<td>No</td>
<td>lowland and highland crested ducks fixed for Ser; rare highland allele in yellow-billed pintail (n = 1 heterozygote); Dendrocygna, Anser also have Ser reverse pattern in Andean ruddy duck - Ser in lowlands/Gly in highlands</td>
</tr>
<tr>
<td>13</td>
<td>Ser</td>
<td>A helix</td>
<td>Andean ruddy duck, yellow-billed pintail, speckled teal</td>
<td>Gly (Ser)</td>
<td>1.00, 0.08, 0.43 (Yes)</td>
<td>No</td>
<td>reverse pattern in Andean ruddy duck - Ile in lowlands/Gly in highlands; Heteronetta also has Ile Dendrocygna, Stictonetta, Gallonetta, Neochen, lowland Chlorophaga spp. also have Ser</td>
</tr>
<tr>
<td>14</td>
<td>Ile</td>
<td>A helix</td>
<td>Andean ruddy duck</td>
<td>Leu (Ile)</td>
<td>1.00</td>
<td>No</td>
<td>—</td>
</tr>
<tr>
<td>55</td>
<td>Ser</td>
<td>2β1 intersubunit contact</td>
<td>Andean goose</td>
<td>Leu (Ser, Ala, Thr)</td>
<td>0.00</td>
<td>No</td>
<td>lowland and highland and highland crested ducks fixed for Thr; Thr not found in other waterfowl</td>
</tr>
<tr>
<td>62</td>
<td>Thr</td>
<td>E helix</td>
<td>Torrent duck</td>
<td>Ala</td>
<td>0.00</td>
<td>(Yes)</td>
<td>—</td>
</tr>
<tr>
<td>69</td>
<td>Ser</td>
<td>E helix</td>
<td>Andean ruddy duck</td>
<td>Thr</td>
<td>0.00</td>
<td>(Yes)</td>
<td>lowland and highland Andean ruddy ducks fixed for Ser; Ser not found in other waterfowl</td>
</tr>
<tr>
<td>73</td>
<td>Glu</td>
<td>E helix</td>
<td>Speckled teal</td>
<td>Asp (Glu)</td>
<td>0.01*</td>
<td>No</td>
<td>—</td>
</tr>
<tr>
<td>86</td>
<td>Ser</td>
<td>1st IPP binding site</td>
<td>Andean goose</td>
<td>Ala</td>
<td>1.00</td>
<td>Yes</td>
<td>—</td>
</tr>
<tr>
<td>94</td>
<td>Glu</td>
<td>1st IPP binding site</td>
<td>Crested duck, puna teal</td>
<td>Asp (Glu)</td>
<td>0.90, 1.00</td>
<td>No</td>
<td>Tadorna ferruginea, Tadorna tadornoides also have Glu lowland and highland highland crested ducks fixed for Val; Val not found in other waterfowl</td>
</tr>
<tr>
<td>111</td>
<td>Val</td>
<td>2β1 intersubunit contact</td>
<td>Torrent duck</td>
<td>Ile</td>
<td>0.00</td>
<td>No</td>
<td>—</td>
</tr>
<tr>
<td>116</td>
<td>Ser</td>
<td>2β1 intersubunit contact</td>
<td>Blue-winged goose, yellow-billed pintail, speckled teal</td>
<td>Ala (Ser, Thr)</td>
<td>n/a, 0.88, 0.97</td>
<td>No</td>
<td>Cygnus melanocoryphus, Stictonetta, Tachyeres leucocephalus also have Ser</td>
</tr>
<tr>
<td>125</td>
<td>Asp</td>
<td>H helix</td>
<td>Bar-headed goose</td>
<td>Glu, Asp</td>
<td>n/a</td>
<td>No</td>
<td>Asp is present in most other basal waterfowl but not other Anser spp.</td>
</tr>
<tr>
<td>133</td>
<td>Met</td>
<td>1st IPP binding site</td>
<td>Yellow-billed pintail, speckled teal</td>
<td>Leu</td>
<td>0.88, 0.97</td>
<td>Yes</td>
<td>—</td>
</tr>
</tbody>
</table>

*P-value for pairwise $F_{ST} > 0.05.$
hydroxyl group of the threonine hydrogen bonds with the residues at either α134 or α138, or alternatively the hydrophilic nature of this residue may draw an additional water molecule into this region of the molecule, thus perturbing binding of IPP (Fig. 6). Further substantiating this hypothesis is the observation that a serine occurs at this homologous position in the αD subunit of the HbD (minor) isoform in chicken and geese (Knapp et al. 1999; McCracken, KG, unpublished data); serine, like threonine, also possesses a hydroxyl group, and the HbD isoform exhibits higher oxygen affinity and cooperativity than the HbA major isoform (Cirotto & Geraci 1975; Baumann et al. 1984; Knapp et al. 1999).

Tamburrini et al. (2000) suggested that the second phosphate binding site contributes to controlled release of oxygen under conditions of hypoxic stress. Allosteric mechanisms of a second IPP binding and its role in adaptation to high-altitude hypoxia are in need of further exploration, and the effects of the Ala → Thr α77 substitution should now be tested experimentally.

Our study also revealed another striking pattern. More than half of the substitutions that were derived in highland lineages resulted in the acquisition or loss (18 gains vs. 2 losses) of a serine or threonine. Both serine and threonine possess a hydroxyl group that hydrogen-bonds to a wide variety of polar substrates, thus

![Fig. 5](image-url) Structural positions of amino acid replacements in the oxy (R-state) major haemoglobin of greylag goose. (a) Helix A of the α1 subunit showing α5, α8, α9, α12, and α18. (b) Helix A of the β1 subunit showing β13 and β14. (c) Helix E of the β1 subunit showing β62, β69, and β73. (d) β4, β86, β94, and β133 and the 1st IPP binding sites β1, β2, β82, β104, β135, β139, β143, β144, and β146 (yellow text). (e) α77 and the 2nd IPP binding sites α1, α95, α99, α134, α137, α138, and α141 (yellow text). (f) αβ1 intersubunit contacts for α111 and β116. (g) αβ1 intersubunit contacts for α119 and β55. (h) αβ1 intersubunit contact for β111 on helix G of the β1 subunit. Structures were illustrated with Cn3D 4.1.

![Fig. 6](image-url) Surface of the greylag goose oxy (R-state) major haemoglobin illustrating α77 (green) and the 2nd IPP binding sites α1, α95, α99, α134, α137, α138, and α141 (yellow) at the entrance to the central cavity near the N- and C-termini of the two α subunits. Residues within 4.1A of these binding sites are shown in grey. The structure was illustrated with Python Molecular Viewer 1.5.2.
implicating a role for modified allosteric interactions with effector molecules at the majority of these variable sites.

Highland populations of crested duck (Lophonetta specularioides) and puna teal (Anas puna) were fixed for Glu-b94, whereas lowland populations in Argentina (but excluding Mendoza in Lophonetta) were fixed for the ancestral residue Asp-b94 (Fig. 4). This amino acid occurs within 3.5Å of the IPP binding site at position β144 in greylag goose (Liang et al. 2001a). In human haemoglobin, Asp-b94 forms a salt bridge with the imidazole ring of the N-terminal His-b146 (Perutz 1970; Kilmartin et al. 1980; Shih et al. 1993), which stabilizes the low-affinity deoxy (T-state) structure and contributes to the Bohr effect. This salt bridge does not occur in bar-headed goose (Liang et al. 2001a), and the functional effect of the Glu-b94 replacement in Andean highland species remains to be elucidated.

We also identified a potential reversal in our study. Andean ruddy duck (O. j. ferruginea) exhibited a pattern opposite to that of other highland lineages (Fig. 4; Table 1). The derived Ser-b13/Ile-b14 allele occurs in the lowland population. However, Ser-b69 is fixed in both the highland and lowland populations. All three of these substitutions were absent in the North American ruddy duck (O. j. jamaicensis) and the other stiff-tailed ducks we sequenced (Fig. S1). One possible explanation consistent with a previously published biogeographical hypothesis is that the Andean highlands were first colonized by ruddy ducks dispersing from North America, and that colonization of the cordillera occurred ‘top-down’, north-to-south, from the central highlands to lower elevations in Patagonia (McCracken & Sorenson 2005). This would require the need to adapt first to hypoxia in the highlands and then again to normoxic conditions in the lowlands. Under this scenario, Ser-b69 would be predicted to increase O₂-affinity, with Ser-b13/Ile-b14 reversing the effect. This hypothesis could be tested with functional assays, but more generally speaking, this example highlights the need to consider the role of compensatory substitutions, as substitutions at one site may influence or reverse the effects of substitutions at other sites (Storz & Moriyama 2008).

Surprisingly, all of the sheldgeese (Chloephaga spp., Neochen jubata) were, like the Andean goose, homozygous for Ser-b55, which has been the subject of extensive structural and physiological research (Jessen et al. 1991; Weber et al. 1993). Ser-b55 is thus a synapomorphy for sheldgeese and is present in both lowland and highland taxa, raising a question about the physiological effects of this amino acid replacement in low-elevation species. Andean goose, however, differs from other sheldgeese at three other positions (Ala-28, Thr-277, Ser-b86), one of which is unique among waterfowl and two of which are shared by other highland species (Fig. 4, Table 1). The phenotypic effects of these substitutions, including the possibility that they might modify the effect of Ser-b55, should be examined experimentally.

Conclusion

Our study of the major haemoglobin in ten high-altitude waterfowl lineages revealed a striking pattern of parallel amino acid replacement, as well as recurrent substitutions at adjacent sites likely to produce similar phenotypic effects. No two highland lineages possessed exactly the same set of amino acid substitutions, but eight species possessed 1–4 parallel changes with as many as four other lineages (Fig. 4, Table 1). Additionally, eight species exhibited substitutions at external helical positions on the αA or βA subunits, seven had one or two substitutions in close proximity to IPP binding sites, and six species each had one substitution at either the α119/b55 or α111/b16 intersubunit contact sites. These simple observations suggest that the number of amino acid substitutions required to produce a haemoglobin protein adapted to low oxygen environments is small and involves only a few substitutions at key positions in the protein molecule (Perutz 1983; see also Gillespie 1991). While the substitutional pathways may not be identical in each species, similar underlying physiochemical mechanisms may be involved in each case.

Our findings are consistent with Orr (2005), who predicted using extreme value theory that replicate populations will fix the same beneficial mutation with probability $P = 2/(n + 1)$ when the number of possible beneficial mutations ($n$) is small (see also Wood et al. 2005; Weinrich et al. 2006). The simulations we employed indicate that the patterns of parallel evolution we observed were not caused by the stochastic forces of mutation and drift, but instead are consistent with the hypothesis that directional selection on the major haemoglobin has resulted in parallel evolution in independent lineages evolving in a common highland environment. Greater divergence in haemoglobin allele frequencies between highland and lowland populations of the same species than in five unlinked reference loci further supports this conclusion (see Table S4 for a summary of $\Phi_{ST}$ values from McCracken et al. 2009a,b). Migration of haemoglobin alleles between highland and lowland populations of the same species was also sharply restricted in these studies, suggesting that different genotypes may have different fitness rankings in different elevational zones.

In sum, the phylogenetic and protein structural analyses undertaken here provide a framework and impetus for comparative mechanistic analyses of haemoglobin function in native highland populations. The striking
patterns of parallel evolution we observed likely emerged because the major haemoglobin is a relatively simple protein coded by a small number of genes, in which the number of possible beneficial mutations is limited. While we have not yet demonstrated the functionality of the observed substitutions, by strong inference we can predict that parallel changes such as those observed here likely have an adaptive function even if those functions have not yet been determined.

Acknowledgements

We thank the people and the provincial and federal governments in South America who assisted us. Violeta Muñoz-Fuentes, Jeff Peters, Jay Storz, Naoki Takebayashi, and three anonymous reviewers provided helpful discussions or comments on the manuscript. Funding was provided by the Institute of Arctic Biology at the University of Alaska Fairbanks, Alaska EPSCoR (EPS-0092040, EPS-0346770), Delta Waterfowl Foundation, Frank M. Chapman Fund at the American Museum of Natural History, and NSF (DEB-0444748).

References


References

Museum of Natural History, and NSF (DEB-0444748).


Acknowledgements

We thank the people and the provincial and federal governments in South America who assisted us. Violeta Muñoz-Fuentes, Jeff Peters, Jay Storz, Naoki Takebayashi, and three anonymous reviewers provided helpful discussions or comments on the manuscript. Funding was provided by the Institute of Arctic Biology at the University of Alaska Fairbanks, Alaska EPSCoR (EPS-0092040, EPS-0346770), Delta Waterfowl Foundation, Frank M. Chapman Fund at the American Museum of Natural History, and NSF (DEB-0444748).

patterns of parallel evolution we observed likely emerged because the major haemoglobin is a relatively simple protein coded by a small number of genes, in which the number of possible beneficial mutations is limited. While we have not yet demonstrated the functionality of the observed substitutions, by strong inference we can predict that parallel changes such as those observed here likely have an adaptive function even if those functions have not yet been determined.

Acknowledgements

We thank the people and the provincial and federal governments in South America who assisted us. Violeta Muñoz-Fuentes, Jeff Peters, Jay Storz, Naoki Takebayashi, and three anonymous reviewers provided helpful discussions or comments on the manuscript. Funding was provided by the Institute of Arctic Biology at the University of Alaska Fairbanks, Alaska EPSCoR (EPS-0092040, EPS-0346770), Delta Waterfowl Foundation, Frank M. Chapman Fund at the American Museum of Natural History, and NSF (DEB-0444748).

patterns of parallel evolution we observed likely emerged because the major haemoglobin is a relatively simple protein coded by a small number of genes, in which the number of possible beneficial mutations is limited. While we have not yet demonstrated the functionality of the observed substitutions, by strong inference we can predict that parallel changes such as those observed here likely have an adaptive function even if those functions have not yet been determined.

Acknowledgements

We thank the people and the provincial and federal governments in South America who assisted us. Violeta Muñoz-Fuentes, Jeff Peters, Jay Storz, Naoki Takebayashi, and three anonymous reviewers provided helpful discussions or comments on the manuscript. Funding was provided by the Institute of Arctic Biology at the University of Alaska Fairbanks, Alaska EPSCoR (EPS-0092040, EPS-0346770), Delta Waterfowl Foundation, Frank M. Chapman Fund at the American Museum of Natural History, and NSF (DEB-0444748).

patterns of parallel evolution we observed likely emerged because the major haemoglobin is a relatively simple protein coded by a small number of genes, in which the number of possible beneficial mutations is limited. While we have not yet demonstrated the functionality of the observed substitutions, by strong inference we can predict that parallel changes such as those observed here likely have an adaptive function even if those functions have not yet been determined.

Acknowledgements

We thank the people and the provincial and federal governments in South America who assisted us. Violeta Muñoz-Fuentes, Jeff Peters, Jay Storz, Naoki Takebayashi, and three anonymous reviewers provided helpful discussions or comments on the manuscript. Funding was provided by the Institute of Arctic Biology at the University of Alaska Fairbanks, Alaska EPSCoR (EPS-0092040, EPS-0346770), Delta Waterfowl Foundation, Frank M. Chapman Fund at the American Museum of Natural History, and NSF (DEB-0444748).

patterns of parallel evolution we observed likely emerged because the major haemoglobin is a relatively simple protein coded by a small number of genes, in which the number of possible beneficial mutations is limited. While we have not yet demonstrated the functionality of the observed substitutions, by strong inference we can predict that parallel changes such as those observed here likely have an adaptive function even if those functions have not yet been determined.

Acknowledgements

We thank the people and the provincial and federal governments in South America who assisted us. Violeta Muñoz-Fuentes, Jeff Peters, Jay Storz, Naoki Takebayashi, and three anonymous reviewers provided helpful discussions or comments on the manuscript. Funding was provided by the Institute of Arctic Biology at the University of Alaska Fairbanks, Alaska EPSCoR (EPS-0092040, EPS-0346770), Delta Waterfowl Foundation, Frank M. Chapman Fund at the American Museum of Natural History, and NSF (DEB-0444748).

patterns of parallel evolution we observed likely emerged because the major haemoglobin is a relatively simple protein coded by a small number of genes, in which the number of possible beneficial mutations is limited. While we have not yet demonstrated the functionality of the observed substitutions, by strong inference we can predict that parallel changes such as those observed here likely have an adaptive function even if those functions have not yet been determined.

Acknowledgements

We thank the people and the provincial and federal governments in South America who assisted us. Violeta Muñoz-Fuentes, Jeff Peters, Jay Storz, Naoki Takebayashi, and three anonymous reviewers provided helpful discussions or comments on the manuscript. Funding was provided by the Institute of Arctic Biology at the University of Alaska Fairbanks, Alaska EPSCoR (EPS-0092040, EPS-0346770), Delta Waterfowl Foundation, Frank M. Chapman Fund at the American Museum of Natural History, and NSF (DEB-0444748).

patterns of parallel evolution we observed likely emerged because the major haemoglobin is a relatively simple protein coded by a small number of genes, in which the number of possible beneficial mutations is limited. While we have not yet demonstrated the functionality of the observed substitutions, by strong inference we can predict that parallel changes such as those observed here likely have an adaptive function even if those functions have not yet been determined.

Acknowledgements

We thank the people and the provincial and federal governments in South America who assisted us. Violeta Muñoz-Fuentes, Jeff Peters, Jay Storz, Naoki Takebayashi, and three anonymous reviewers provided helpful discussions or comments on the manuscript. Funding was provided by the Institute of Arctic Biology at the University of Alaska Fairbanks, Alaska EPSCoR (EPS-0092040, EPS-0346770), Delta Waterfowl Foundation, Frank M. Chapman Fund at the American Museum of Natural History, and NSF (DEB-0444748).


Supporting information

Additional supporting information may be found in the online version of this article:

Table S1 Localities of specimens included in this study
Table S2 Primers used to amplify and sequence the αA and βA haemoglobin subunits
Table S3 Results of the MS and Seq-Gen 1.3.2 simulations and concentrated changes tests
Table S4 Summary of ΦST values between lowland and highland populations for the αA and βA subunits compared to five unlinked reference loci

Fig. S1 Summary of the variable amino acid positions in the αA haemoglobin subunit of 123 waterfowl species (862 individuals) and the βA haemoglobin subunit of 93 waterfowl species (683 individuals).

Supplementary data file 1 Sequence alignment (NEXUS file) for the αA haemoglobin subunit.

Supplementary data file 2 Sequence alignment (NEXUS file) for the βA haemoglobin subunit.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.