MOLECULAR ECOLOGY

Molecular Ecology (2009) doi: 10.1111/j.1365-294X.2009.04352.x

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

K. G. McCRACKEN,* 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**

*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 Martin 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-α8) and five (Thr-α77) 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 $\alpha^1\beta^1$ 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

Correspondence: Kevin G. McCracken, Fax: +1 (907) 474-6967; E-mail: fnkgm@uaf.edu

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 intraspecific 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 α subunits and two β subunits, each linked to a heme group that binds cooperatively and reversibly with molecular $O_2$. The α and β subunits are coded on different chromosomes and, in birds such as the chicken and ducks, include three linked α chain genes (απ, αD, αA) and four linked β chain genes (βρ, βH, βA, βε), 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 αA subunits and two βA 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 αβ 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 α and β 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 & Milsom 2006, 2007). Increased $O_2$ loading of haemoglobin in the lungs is achieved by a Pro → Ala-α119 substitution. The small R-group side chain of Ala-α119 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 ($P_{50}$) at lower $pO_2$ 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 $\alpha^1\beta^1$ intersubunit contact affected by the bar-headed goose substitution at α119. Andean goose blood has a $P_{50}$ of 33.9 mmHg (Hall *et al.* 1936). Parallel evolution of increased haemoglobin-$O_2$ 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 $pO_2$ 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 (*Cyanochen 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 $F_{ST}$ 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 $\alpha^1\beta^1$ 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.

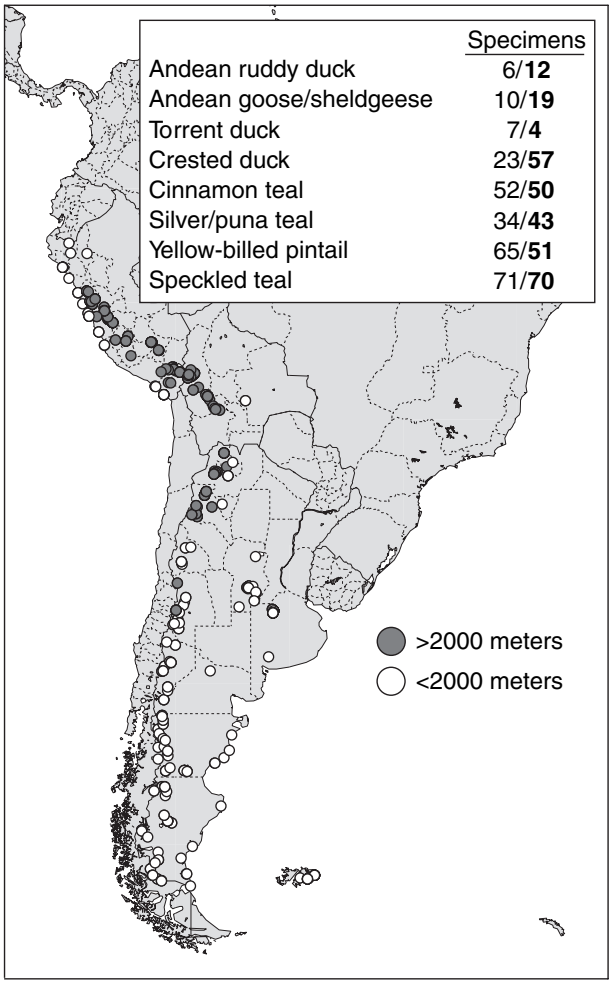


**Fig. 1** Localities for 574 ducks sampled from eight paired lowland and highland populations experiencing contrasting $pO_2$ in the Andes and adjacent lowlands of South America. Grey circles indicate localities >2000 meters elevation; white circles indicate localities <2000 meters. The numbers of lowland specimens per lineage are shown in plain text, and the numbers of highland specimens are shown in bold text. Scientific names: Andean ruddy duck (*Oxyura jamaicensis ferruginea*), Andean goose (*Chloephaga melanoptera*), torrent duck (*Merganetta armata*), crested duck (*Lophonetta specularioides*), cinnamon teal (*Anas cyanoptera*), silver teal (*Anas versicolor*), puna teal (*Anas puna*), yellow-billed pintail (*Anas georgica*), speckled teal (*Anas flavirostris*).

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 aligments 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).

### $F_{ST}$ *calculations*

$F_{ST}$ 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

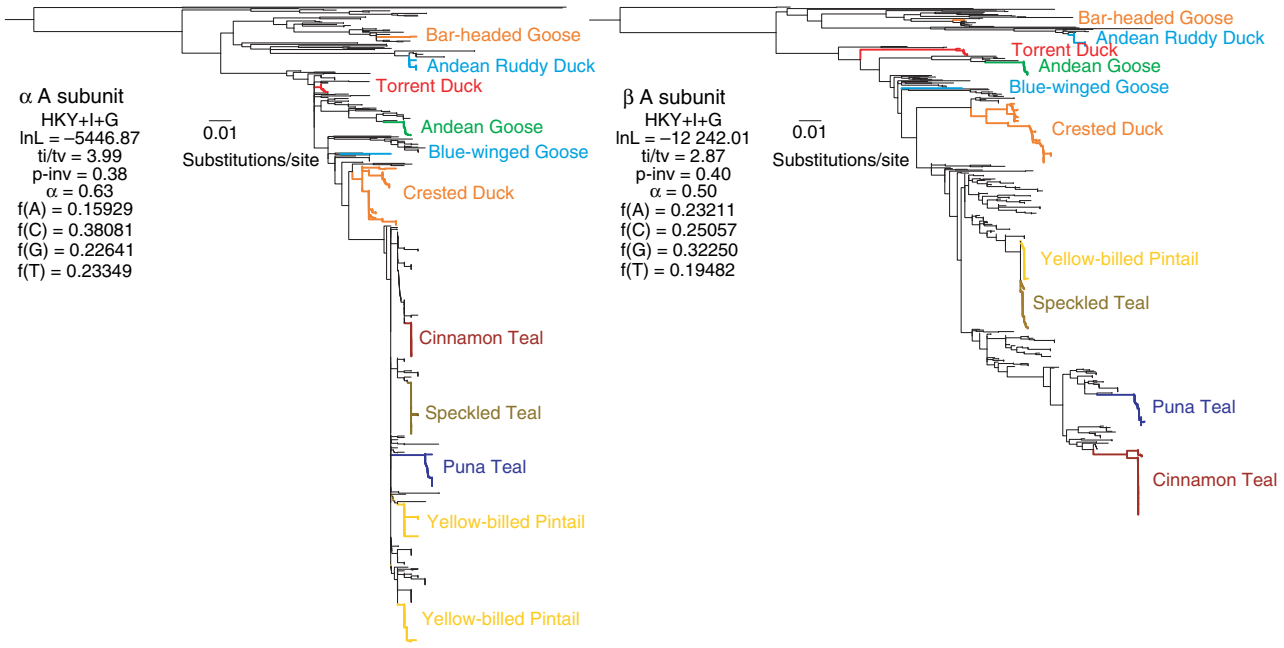


**Fig. 2** Maximum likelihood gene trees for the αA and βA haemoglobin subunits showing the positions of the ten highland waterfowl lineages.

using PAML 3.15 (Yang 1997) was used as the starting point for each simulated dataset. Branch lengths were scaled to Θ 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 αA and βA 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 αA subunit and 14 for the βA 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 βA subunit positions β1, 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 αA subunit positions α1, 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 αA and βA haemoglobin subunits (including both exons and introns) were estimated independently; 123 of 145 waterfowl species (85%) were analyzed for the αA subunit, and 93 species (64%) were analyzed for the βA 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 or parallel substitutions at three or more codon positions were never observed ($P < 0.001$). Each simulation was initiated from the same ancestral sequence, conditioned on theta (Θ)

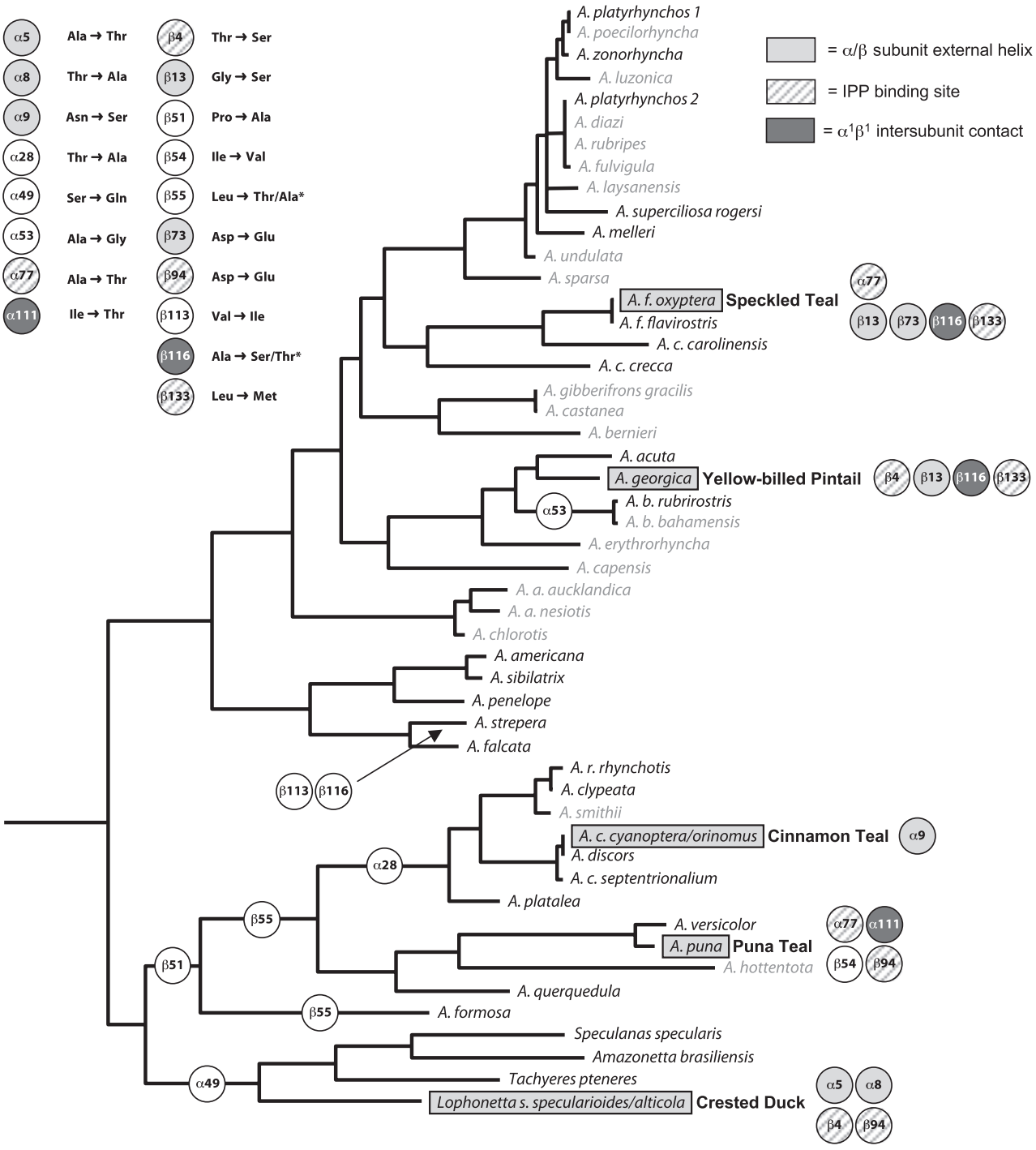


**Fig. 3** αA and βA haemoglobin subunit amino acid substitutions mapped onto a mtDNA phylogeny of dabbling ducks (Johnson & Sorenson 1999). The five species of Andean dabbling ducks (*Anas*, *Lophonetta*) are shown in bold text. Substitutions at external, helical positions on the αA or βA subunit, in close proximity to IPP binding sites, or at $\alpha^1\beta^1$ intersubunit contacts in highland lineages are shown in light grey, cross-hatched, and dark grey circles respectively. Substitutions in non-highland lineages are shown in white circles (Ala-β55 occurs in *A. formosa*; Thr-β116 occurs in *A. strepera*; and Val-β54 was found in a single *A. versicolor* in the lowlands). αA and βA subunits of species shown in faded grey were not sequenced.

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-

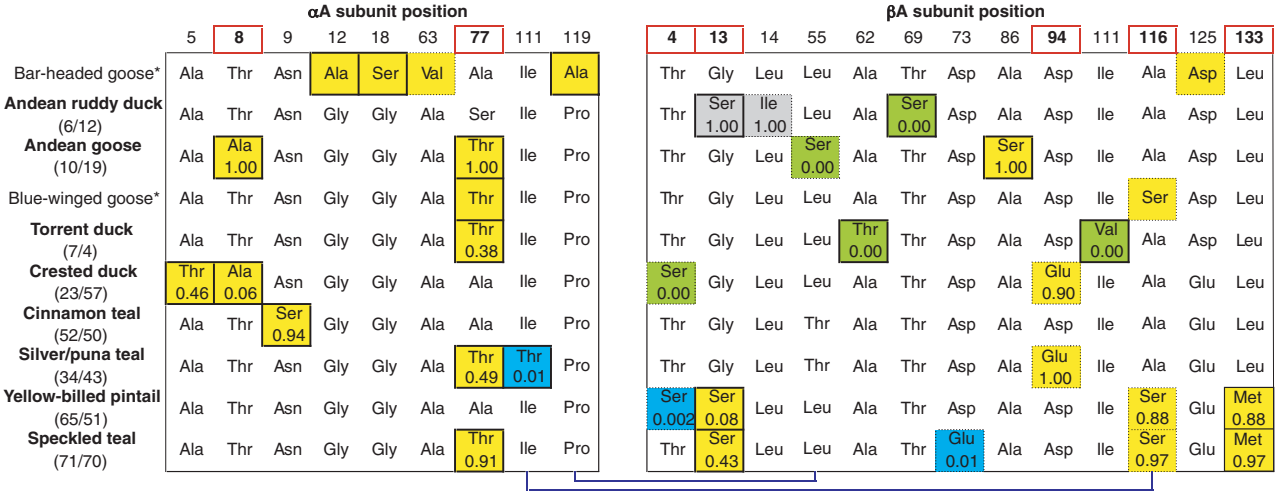


**Fig. 4** Amino acid replacements in the αA and βA haemoglobin subunits of eight Andean waterfowl species (bold text) and bar-headed goose and Abyssinian blue-winged goose (plain text with asterisks), which are endemic to the highlands of Asia and Ethiopia, respectively. $F_{ST}$ calculated for paired lowland and highland populations of the same species or closely related sibling species is shown for amino acid positions in the eight Andean lineages. Seventeen amino acid replacements at nine different positions were overrepresented in highland Andean populations with $F_{ST}$ values ranging from 0.06 to 1.00 (yellow boxes). Two and five additional replacements were observed in Abyssinian blue-winged goose and bar-headed goose but not in their closest relatives. Three amino acid replacements were found as rare alleles (Thr-α111, Ser-β4, Glu-β73; $F_{ST} \leq 0.01$) sampled only in highland populations (blue boxes), and five derived replacements were fixed for the same amino acid in both lowland and highland Andean populations of a given species ($F_{ST} = 0.00$; green boxes); three of the latter were not found in any other waterfowl species. Two derived replacements found in lowland but not highland Andean ruddy ducks are shown in grey boxes. Parallel replacements in independent highland lineages are indicated by red boxes. Replacements found only in waterfowl lineages with highland populations are shown inside bold boxes, whereas replacements also found in one or more unrelated waterfowl lineages without highland populations are shown in dotted boxes. Finally, two pairs of amino acid replacements affecting the same $\alpha^1\beta^1$ intersubunit contacts (α119/β55 and α111/β116) are connected by blue lines. Samples sizes (lowland/highland) for the eight Andean lineages are shown for each species.

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 $\alpha^1\beta^1$ 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 $\alpha^1\beta^1$ 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)

| Pos. | Amino acid | Structural position | Highland taxa | Amino acids in other Anatidae | $F_{ST}$ | High altitude taxa only | Notes/exceptions |
|---|---|---|---|---|---|---|---|
| α5 | Thr | A helix | Crested duck | Ala | 0.46 | Yes | — |
| **α8** | **Ala** | **A helix** | **Andean goose, crested duck** | **Thr (Ser, Asn)** | **0.06, 1.00** | **Yes** | **—** |
| α9 | Ser | A helix | Cinnamon teal | Asn | 0.94 | Yes | — |
| α12 | Ala | A helix | Bar-headed goose | Gly | n/a | Yes | — |
| α18 | Ser | AB corner | Bar-headed goose | Gly | n/a | Yes | — |
| α63 | Val | E helix | Bar-headed goose | Ala (Val) | n/a | No | *Stictonetta*, *Nettapus pulchellus* also have Val |
| **α77** | **Thr** | **2nd IPP binding site** | **Andean goose, blue-winged goose, torrent duck, puna teal, speckled teal** | **Ala (Gly, Ser)** | **1.00, n/a, 0.38, 0.49, 0.91** | **Yes** | **—** |
| α111 | Thr | $\alpha^1\beta^1$ intersubunit contact | Puna teal | Ile | 0.01* | Yes | rare highland allele in puna teal ($n$ = 2 heterozygotes) |
| α119 | Ala | $\alpha^1\beta^1$ intersubunit contact | Bar-headed goose | Pro | n/a | Yes | — |
| **β4** | **Ser** | **1st IPP binding site** | **Crested duck, yellow-billed pintail** | **Thr (Ser)** | **0.00, 0.002*** | **No** | **lowland and highland crested ducks fixed for Ser; rare highland allele in yellow-billed pintail ($n$ = 1 heterozygote); *Dendrocygna*, *Anser* also have Ser** |
| **β13** | **Ser** | **A helix** | **Andean ruddy duck, yellow-billed pintail, speckled teal** | **Gly (Ser)** | **1.00, 0.08, 0.43** | **(Yes)** | **reverse pattern in Andean ruddy duck - Ser in lowlands/Gly in highlands** |
| β14 | Ile | A helix | Andean ruddy duck | Leu (Ile) | 1.00 | No | reverse pattern in Andean ruddy duck - Ile in lowlands/Leu in highlands; *Heteronetta* also has Ile |
| β55 | Ser | $\alpha^1\beta^1$ intersubunit contact | Andean goose | Leu (Ser, Ala, Thr) | 0.00 | No | *Dendrocygna*, *Stictonetta*, *Callonetta*, *Neochen*, lowland *Chloephaga* spp. also have Ser |
| β62 | Thr | E helix | Torrent duck | Ala | 0.00 | (Yes) | lowland and highland torrent ducks fixed for Thr; Thr not found in other waterfowl |
| β69 | Ser | E helix | Andean ruddy duck | Thr | 0.00 | (Yes) | lowland and highland Andean ruddy ducks fixed for Ser; Ser not found in other waterfowl |
| β73 | Glu | E helix | Speckled teal | Asp (Glu) | 0.01* | No | rare highland allele in speckled teal ($n$ = 3 heterozygotes); *Dendrocygna* also has Glu |
| β86 | Ser | 1st IPP binding site | Andean goose | Ala | 1.00 | Yes | — |
| **β94** | **Glu** | **1st IPP binding site** | **Crested duck, puna teal** | **Asp (Glu)** | **0.90, 1.00** | **No** | ***Tadorna ferruginea*, *Tadorna tadornoides* also have Glu** |
| β111 | Val | $\alpha^1\beta^1$ intersubunit contact | Torrent duck | Ile | 0.00 | (Yes) | lowland and highland torrent ducks fixed for Val; Val not found in other waterfowl |
| **β116** | **Ser** | **$\alpha^1\beta^1$ intersubunit contact** | **Blue-winged goose, yellow-billed pintail, speckled teal** | **Ala (Ser, Thr)** | **n/a, 0.88, 0.97** | **No** | ***Cygnus melanocoryphus*, *Stictonetta*, *Tachyeres leucocephalus* also have Ser** |
| β125 | Asp | H helix | Bar-headed goose | Glu, Asp | n/a | No | Asp is present in most other basal waterfowl but not *other* Anser spp. |
| **β133** | **Met** | **1st IPP binding site** | **Yellow-billed pintail, speckled teal** | **Leu** | **0.88, 0.97** | **Yes** | **—** |

*$P$-value for pairwise $F_{ST} > 0.05$.

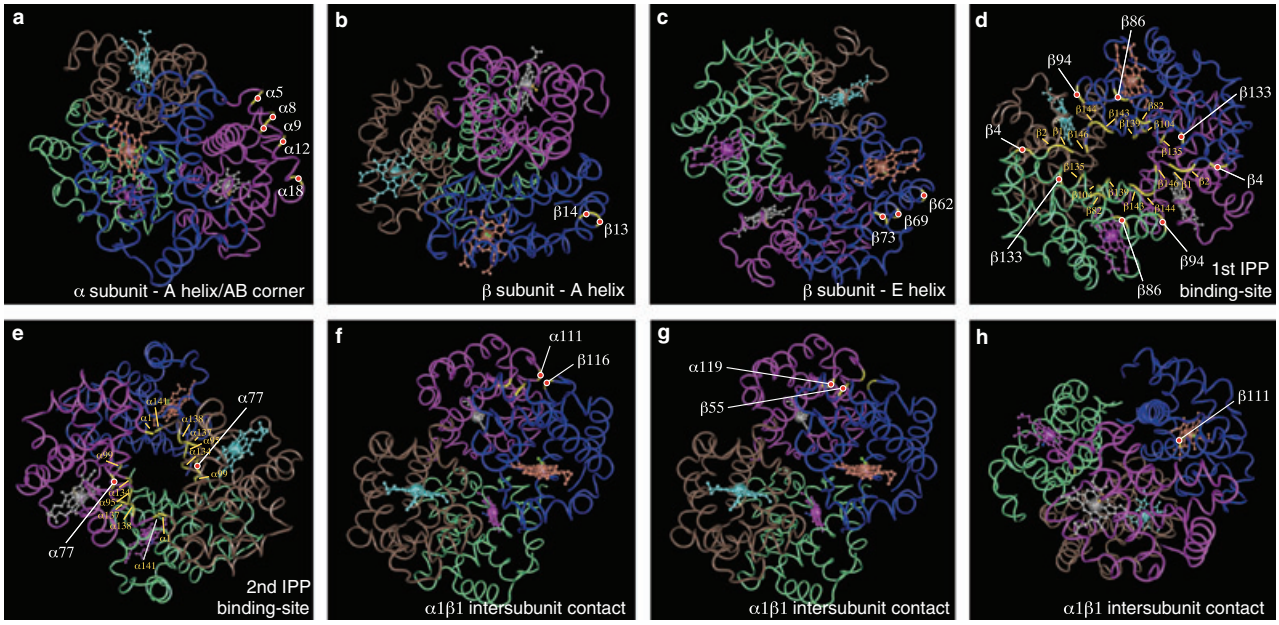


**Fig. 5** Structural positions of amino acid replacements in the oxy (R-state) major haemoglobin of greylag goose. (a) Helix A of the $\alpha^1$ subunit showing α5, α8, α9, α12, and α18. (b) Helix A of the $\beta^1$ subunit showing β13 and β14. (c) Helix E of the $\beta^1$ subunit showing β62, β69, 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) $\alpha^1\beta^1$ intersubunit contacts for α111 and β116. (g) $\alpha^1\beta^1$ intersubunit contacts for α119 and β55. (h) $\alpha^1\beta^1$ intersubunit contact for β111 on helix G of the $\beta^1$ subunit. Structures were illustrated with Cn3D 4.1.

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 homoglous 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

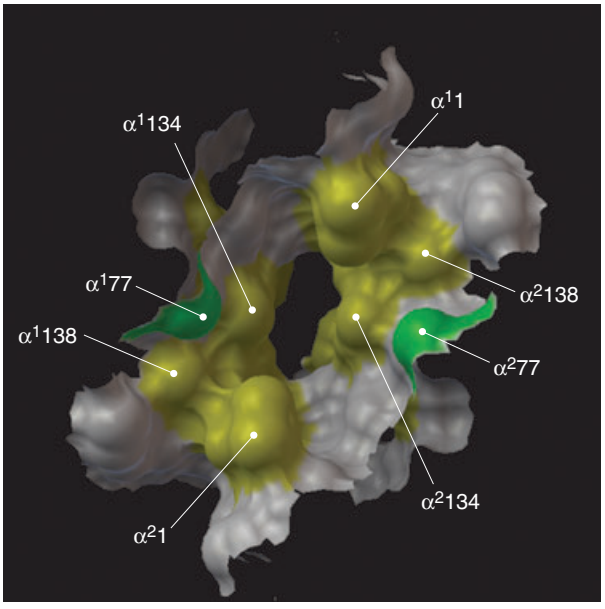


**Fig. 6** Surface of the greylag goose oxy (R-state) major haemoglobin illustrating α77 (green) and the 2nd IPP binding sites α1, α95, α99, α134, α137, α138, and α141 (yellow) at the entrance to the central cavity near the N- and C-termini of the two α subunits. Residues within 4.1Å of these binding 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-β94, whereas lowland populations in Argentina (but excluding Mendoza in *Lophonetta*) were fixed for the ancestral residue Asp-β94 (Fig. 4). This amino acid occurs within 3.57Å of the IPP binding site at postion β144 in greylag goose (Liang *et al.* 2001a). In human haemoglobin, Asp-β94 forms a salt bridge with the imidazole ring of the N-terminal His-β146 (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-β94 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-β13/Ile-β14 allele occurs in the lowland population. However, Ser-β69 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-β69 would be predicted to increase $O_2$-affinity, with Ser-β13/Ile-β14 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-β55, which has been the subject of extensive structural and physiological research (Jessen *et al.* 1991; Weber *et al.* 1993). Ser-β55 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 all other sheldgeese at three other positions (Ala-α8, Thr-α77, Ser-β86), 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-β55, 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/β55 or α111/β116 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

Arendt J, Reznick D (2008) Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *TREE*, **23**, 26–32.

Baumann R, Goldbach E, Haller EA, Wright PG (1984) Organic phosphates increase the solubility of avian haemoglobin D and embryonic chicken haemoglobin. *Biochemical Journal*, **217**, 767–771.

Beall CM (2006) Andean, Tibetan, and Ethiopian patterns of adaptation to high-altitude hypoxia. *Integrative and Comparative Biology*, **46**, 18–24.

Bollback JP, Huelsenbeck JP (2009) Parallel genetic evolution within and between bacteriophage species of varying degrees of divergence. *Genetics*, **181**, 225–234.

Bulgarella M, Stewart NC, Fedorov VB, Moore AV, McCracken KG (2009) Hemoglobin transcript abundance in a cDNA library from bone marrow of crested ducks (*Lophonetta Specularioides*) in the Peruvian high Andes. *Auk*, **126**, 666–672.

Bulgarella M, Sorenson MD, Peters JL, Wilson RE, McCracken KG (2010) Phylogenetic relationships of *Amazonetta*, *Speculanas*, *Lophonetta*, and *Tachyeres*: four morphologically divergent genera of ducks endemic to South America. *Journal of Avian Biology*, (in press).

Bull JJ, Badgett MR, Wichman HA *et al.* (1997) Exceptional convergent evolution in a virus. *Genetics*, **147**, 1497–1507.

Cirotto C, Geraci G (1975) Embryonic chicken hemoglobins: studies on the oxygen equilibrium of two pure components. *Comparative Biochemistry and Physiology A*, **51**, 159–163.

Colosimo PF, Hosemann KE, Balabhadra S *et al.* (2005) Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*, **307**, 1928–1933.

Dall'Acqua W, Goldman ER, Lin W *et al.* (1998) A mutational analysis of binding interactions in an antigen-antibody protein-protein complex. *Biochemistry*, **37**, 7981–7991.

Donne-Gousse C, Laudet V, Hänni C (2002) A molecular phylogeny of anseriformes based on mitochondrial DNA analysis. *Molecular Phylogenetics and Evolution*, **23**, 339–356.

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0, an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online*, **1**, 47–50.

Flint J, Tufarelli C, Peden J *et al.* (2001) Comparative genome analysis delimits a chromosomal domain and identifies key regulatory elements in the α globin cluster. *Human Molecular Genetics*, **10**, 371–382.

Gilchrist AS, Partridge L (1999) A comparison of the genetic basis of wing size divergence in three parallel body size clines of *Drosophila melanogaster*. *Genetics*, **153**, 1775–1787.

Gillespie JH (1991) *The Causes of Molecular Evolution*. Oxford Univ. Press, Oxford.

Golding GB, Dean AM (1998) The structural basis of molecular adaptation. *Molecular Biology and Evolution*, **15**, 355–369.

Grantham R (1974) Amino acid difference formula to help explain protein evolution. *Science*, **185**, 862–864.

Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, **18**, 2714–2723.

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*, **52**, 696–704.

Hall FG, Dill DB, Guzman Barron ES (1936) Comparative physiology in high altitudes. *Journal of Cellular and Comparative Physiology*, **8**, 301–313.

Harshman J (1996) *Phyogeny, Evolutionary Rates, and Ducks*. Ph.D. Dissertation, Univ. Chicago, Chicago, Illinois.

Heger A, Pontin CP (2007) Variable strength of translational selection among twelve *Drosophila* species. *Genetics*, **177**, 1337–1348.

Hochachka PW, Somero GN (2002) *Biochemical Adaptation: Mechanism and Process in Physiological Evolution*. Oxford Univ. Press, Oxford.

Hoekstra HE, Nachman MW (2003) Different genes underly adaptive melanism in different populations of rock pocket mice. *Molecular Ecology*, **12**, 1185–1194.

Hopkins SR, Powell FL (2001) Common themes of adaptation to hypoxia. *Advances in Experimental Medicine and Biology*, **502**, 153–167.

Hornbein TF, Schoene RB (2001) *High Altitude: An Exploration of Human Adaptation*. Marcel Dekker, Inc., New York.

Hudson RR (2002) Generating samples under a Wright-Fisher neutral model. *Bioinformatics*, **18**, 337–338.

Jessen TH, Weber RE, Fermi G, Tame J, Braunitzer G (1991) Adaptation of bird hemoglobins to high altitudes: demonstrations of molecular mechanism by protein engineering. *Proceedings of the National Academy of Sciences of the United States of America*, **88**, 6519–6522.

Johnson KP, Sorenson MD (1999) Phylogeny and biogeography of dabbling ducks (genus: *Anas*): a comparison of molecular and morphological evidence. *Auk*, **116**, 792–805.

Jost MC, Hillis DM, Lu Y, Kyle JW, Fozzard HA, Zakon HH (2008) Toxin-resistant sodium channels: parallel adaptive evolution across a complete gene family. *Molecular Biology and Evolution*, **25**, 1016–1024.

Kilmartin JV, Fogg JH, Perutz MF (1980) Role of the C-terminal histidine in the alkaline Bohr effect of human hemoglobin. *Biochemistry*, **19**, 3189–3193.

Knapp JE, Oliveira MA, Xie Q, Ernst SR, Riggs AF, Hackert ML (1999) The structural and functional analysis of the hemoglobin D component from chicken. *Journal of Biological Chemistry*, **274**, 6411–6420.

Kornegay JR, Schilling JW, Wilson AC (1994) Molecular adaptation of a leaf-eating bird: stomach lysozyme of the hoatzin. *Molecular Biology and Evolution*, **11**, 921–928.

Kuhner MK (2006) LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. *Bioinformatics*, **22**, 768–770.

Liang YH, Liu XZ, Liu SH, Lu GY (2001a) The structure of greylag goose oxy haemoglobin: the roles of four mutations compared with bar-headed goose haemoglobin. *Acta Crystallographica Section D, Biological Crystallography*, **57**, 1850–1856.

Liang Y, Hua Z, Liang X, Xu Q, Lu G (2001b) The crystal structure of bar-headed goose hemoglobin in deoxy form: the allosteric mechanism of a hemoglobin species with high oxygen affinity. *Journal of Molecular Biology*, **313**, 123–137.

Liu XZ, Li SL, Jing H, Liang YH, Hua ZQ, Lu GY (2001) Avian haemoglobins and structural basis of high affinity for oxygen: structure of bar-headed goose aquomet haemoglobin. *Acta Crystallographica Section D, Biological Crystallography*, **57**, 775–783.

Livezey BC (1986) A phylogenetic analysis of recent anseriform genera using morphological characters. *Auk*, **103**, 737–754.

Maddison WP (1990) A method for testing the correlated evolution of two binary characters: are gains or losses concentrated on certain branches of a phylogenetic tree? *Evolution*, **44**, 539–557.

Maddison DR, Maddison WP (2000) *MacClade 4.0*. Sinaur Associates, Sunderland, Massachusetts.

McCracken KG, Sorenson MD (2005) Is homoplasy or lineage sorting the source of incongruent mtDNA and nuclear gene trees in the stiff-tailed ducks (*Nomonyx-Oxyura*). *Systematic Biology*, **54**, 35–55.

McCracken KG, Harshman J, McClellan DA, Afton AD (1999) Data set incongruence and correlated character evolution: an example of functional convergence in the hind-limbs of stifftail diving ducks. *Systematic Biology*, **48**, 683–714.

McCracken KG, Barger CP, Bulgarella M *et al.* (2009a) Signatures of high-altitude adaptation in the major hemoglobin of five species of Andean dabbling ducks. *American Naturalist*, (in press).

McCracken KG, Bulgarella M, Johnson KP *et al.* (2009b) Gene flow in the face of countervailing selection: adaptation to high-altitude hypoxia in the βA hemoglobin subunit of yellow-billed pintails in the Andes. *Molecular Biology and Evolution*, **26**, 815–827.

Nachman MW, Hoekstra HE, D'Agostino SL (2003) The genetic basis of adaptive melanism in pocket mice. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 5268–5273.

Orr HA (2005) The probability of parallel evolution. *Evolution*, **59**, 216–220.

Perutz MF (1970) The Bohr effect and combination with organic phosphates. *Nature*, **228**, 734–739.

Perutz MF (1983) Species adaptation in a protein molecule. *Molecular Biology and Evolution*, **1**, 1–28.

Perutz MF (1989) Mechanisms of cooperativity and allosteric regulation in proteins. *Quarterly Reviews of Biophysics*, **22**, 139–236.

Petschow D, Wurdinger I, Baumann R, Duhm J, Braunitzer G, Bauer C (1977) Cause of high blood $O_2$ affinity of animals living at high altitude. *Journal of Applied Physiology*, **42**, 139–143.

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.

Rambaut A, Grassly NC (1997) Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Computer Applications in the Biosciences*, **13**, 235–238.

Reitman M, Grasso JA, Blumenthal R, Lewit P (1993) Primary sequence, evolution, and repetitive elements of the *Gallus gallus* (chicken) β-globin cluster. *Genomics*, **18**, 616–626.

Riccio A, Tamburrini M, Giardina B, di Prisco G (2001) Molecular dynamics analysis of a second phosphate site in the hemoglobins of the seabird, south polar skua. Is there a site-site migratory mechanism along the central cavity?. *Biophysical Journal*, **81**, 1938–1946.

Riehle MM, Bennett AF, Long AD (2001) Genetic architecture of thermal adaptation in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 525–530.

Rokas A, Carroll SB (2008) Frequent and widespread parallel evolution of protein sequences. *Molecular Biology and Evolution*, **25**, 1943–1953.

Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, **19**, 1572–1574.

Rowley AF, Ratcliffe NA (1988) *Vertebrate Blood Cells*. Cambridge University Press, Cambridge.

Schwarz G (1978) Estimating the dimension of a model. *Annals of Statistics*, **6**, 461–464.

Scott GR, Milsom WK (2006) Flying high: a theoretical analysis of the factors limiting exercise performance in birds at altitude. *Repiratory Physiology and Neurobiology*, **154**, 284–301.

Scott GR, Milsom WK (2007) Control of breathing and adaptation to high altitude in the bar-headed goose. *American Journal of Physiology: Regulatory, Integrative and Comparative Physiology*, **293**, R379–R391.

Shih DT, Luisi BF, Miyazaki G, Perutz MF (1993) A mutagenic study of the allosteric linakge of HisHC3(146)β in haemoglobin. *Journal of Molecular Biology*, **230**, 1291–1296.

Sorenson MD, Cooper A, Paxinos EE *et al.* (1999) Relationships of the extinct moa-nalos, flightless Hawaiian waterfowl, based on ancient DNA. *Proceedings of the Royal Society of London. Series B*, **266**, 2187–2193.

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *American Journal of Human Genetics*, **68**, 978–989.

Storz JF, Moriyama H (2008) Mechanisms of hemoglobin adaptation to high altitude hypoxia. *High Altitude Medicine Biology*, **9**, 148–157.

Storz JF, Sabatino SJ, Hoffmann FG *et al.* (2007) The molecular basis of high-altitude adaptation in deer mice. *PLoS Genetics*, **3**, 448–459.

Storz JF, Runck AF, Sabatino SJ *et al.* (2009) Evolutionary and functional insights into the mechanism underlying high-altitude adaptation of deer mouse hemoglobin. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 14 450–14 455.

Swofford DL (2002) *PAUP*: Phylogenetic Analysis Using Parsimony (and Other Methods)*, version 4. Sinauer Associates, Sunderland, Massachusetts.

Tamburrini M, Riccio A, Romano M, Giardina B, di Prisco G (2000) Structural and functional analysis of the two haemoglobins of the Antarctic seabird *Catharacta maccormicki*: characterization of an additional phosphate binding site by molecular modelling. *European Journal of Biochemistry*, **267**, 6089–6098.

Wang HC, Liang YH, Zhu JP, Lu GY (2000) Crystallization and preliminary crystallographic studies of bar-headed goose fluoromethaemoglobin with inostitol hexaphosphate. *Acta Crystallographica Section D, Biological Crystallography*, **56**, 1183–1184.

Weber RE (2007) High altitude adaptations in vertebrate hemoglobins. *Respiratory Physiology and Neurobiology*, **158**, 132–142.

Weber RE, Jessen TH, Malte H, Tame J (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.

Weinrich DM, Delaney NF, DePristo MA, Hartl DL (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science*, **312**, 111–114.

West-Eberhard MJ (2003) *Developmental Plasticity and Evolution*. Oxford Univ. Press, Oxford.

Wichman HA, Badgett MR, Scott LA, Boulianne CM, Bull JJ (1999) Parallel evolution during viral adaptation. *Science*, **285**, 422–424.

Wiens JJ, Chippindale PT, Hillis DM (2003) When are phylogenetic analyses misled by convergence? A case study in Texas cave salamanders. *Systematic Biology*, **52**, 501–514.

Wood TE, Burke JM, Rieseberg LH (2005) Parallel genotypic adaptation: when evolution repeats itself. *Genetica*, **123**, 157–170.

Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood Comput. *Applied Bioscience*, **13**, 555–556.

Zhang JZ, Kumar S (1997) Detection of convergent and parallel evolution at the amino acid sequence level. *Molecular Biology and Evolution*, **14**, 527–536.

Zhang J, Hua Z, Tame JRH, Guangying L, Zhang R, Gu X (1996) The crystal structure of a high oxygen affinity species of haemoglobin (bar-headed goose haemoglobin in the oxy form). *Journal of Molecular Biology*, **255**, 484–493.

---

Kevin McCracken studies population genetics and molecular evolution of waterfowl. He has longstanding interests in South American ducks and the evolutionary processes by which they have adapted to high-altitude regions in the Andes. Christopher Barger, Mariana Bulgarella, and Robert Wilson are graduate students at the University of Alaska Fairbanks, with research interests in waterfowl, biogeography, and ecological adaptation. Kevin Johnson's research focuses on phylogenetics and comparative analyses of birds and their ectoparasitic lice, particularly in waterfowl and doves. Sarah Sonsthagen is a research scientist at the U.S. Geological Survey's Alaska Science Center, studying waterfowl and other Arctic birds. Jorge Trucco is the owner of Patagonia Outfitters, a fly-fishing and wing shooting outfitter in San Martin de los Andes, Argentina. Thomas Valqui is the director of the Centro de Ornitologia y Biodiversidad in Lima, Peru, His research interests focus on the ecology and evolution of Neotropical birds. Kevin Winker is Curator of Birds at the University of Alaska. His research focuses on the patterns and processes of ecological differentiation in birds. Michael Sorenson has longstanding interests in the systematics and behavioural ecology of waterfowl and has in recent years been working on a comprehensive molecular phylogeny for the group.

---

## 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 $\Phi_{ST}$ 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.