Molecular Phylogenetics and Evolution 56 (2010) 649–658

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Phylogenetic and structural analysis of the HbA (α^A/β^A) and HbD (α^D/β^A) hemoglobin genes in two high-altitude waterfowl from the Himalayas and the Andes: Bar-headed goose (*Anser indicus*) and Andean goose (*Chloephaga melanoptera*)

Kevin G. McCracken [a,b,*], Christopher P. Barger [a,b], Michael D. Sorenson [c]

[a] Institute of Arctic Biology, Department of Biology and Wildlife, University of Alaska Fairbanks, Fairbanks, AK 99775, USA
[b] University of Alaska Museum, University of Alaska Fairbanks, Fairbanks, AK 99775, USA
[c] Department of Biology, Boston University, Boston, MA 02215, USA

ARTICLE INFO

Article history:
Received 10 November 2009
Revised 17 April 2010
Accepted 22 April 2010
Available online 29 April 2010

Keywords:
Adaptation
Anatidae
Anserinae
Geese
Hypoxia
Sheldgeese

ABSTRACT

Two species of waterfowl living at high altitude provide a prominent example of parallel adaptation at the molecular level. The bar-headed goose (*Anser indicus*) breeds at high elevations in central Asia and migrates across the Himalayas, where the partial pressure of oxygen (O_2) is one-third of sea level. In South America, the distantly related Andean goose (*Chloephaga melanoptera*) is endemic to the high Andes. Both species exhibit increased blood–O_2 affinity, which has been attributed to the effects of single amino acid substitutions in the major hemoglobin. Here we present phylogenetic analyses of the swans and geese (Anserinae) and South American sheldgeese (Anatinae) using the three genes that encode the major (HbA) and minor (HbD) hemoglobin isoforms. We sought to determine whether two amino acid substitutions that have been the focus of extensive biochemical analysis (Ala-α^A119 and Ser-β^A55) are uniquely derived in bar-headed goose and Andean goose, respectively, and to examine evidence of molecular adaptation at other positions in hemoglobin genes by comparing these two high-altitude taxa to their closest relatives. Bayesian analysis of the α^A-, α^D-, and β^A-subunit genes produced well-resolved phylogenies, with high posterior probabilities and bootstrap values for most genera. The bar-headed goose is likely sister to all other *Anser* species. Andean goose, the sole highland representative of the South American sheldgeese is either sister to the other *Chloephaga* species or sister to *Neochen*. In the bar-headed goose, four derived substitutions were observed in HbA (α^A12, 18, 63, 119) and two in HbD (α^D2, 47). Four derived substitutions in Andean goose include three in HbA (α^A8, 77; β^A86) and two in HbD (α^D9; β^A86). Considering both highland species, four substitutions (Ala-α^A8, Ala-α^A12, Ser-α^A18, Leu-α^D9) were located at adjacent positions on the A helix (or AB corner) of the α-chains, three others (Thr-α^A77, Ser-β^A86, Ser-α^D2) were in close proximity to inositolpentaphosphate (IP_5) binding sites, and Ala-α^A119 occurred at an αβ intersubunit contact. Ser-β^A55, which is involved in the same αβ intersubunit contact and was previously shown to increase Hb–O_2 affinity, is not unique to Andean goose, but is a synapomorphy of the South American sheldgeese, a clade of predominantly lowland waterfowl. Our findings illustrate the importance of understanding phylogenetic relationships and polarity of character-state changes when making inferences about adaptive evolution.

1. Introduction

Independent evolutionary lineages challenged by similar environmental extremes provide an ideal context for the study of adaptive evolution. A critical component of this approach is comparative phylogenetic analysis of sequence evolution at functionally relevant genetic loci in closely related species inhabiting contrasting environments. When combined with structural and biochemical information, phylogenetic patterns of amino acid conservation and change can be used to identify changes of likely functional significance and to distinguish parallel evolution in independent lineages from shared ancestral traits (Colosimo et al., 2005; Jost et al., 2008; Feldman et al., 2009; McCracken et al., 2009a).

Two species of waterfowl living at high altitude have served as a prominent example of parallel adaptation to hypoxia and parallel

* Corresponding author. Address: Institute of Arctic Biology, 311 Irving 1, University of Alaska Fairbanks, Fairbanks, AK 99775, USA. Fax: +1 907 474 6419.
E-mail address: kmccrack@iab.alaska.edu (K.G. McCracken).

1055-7903/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2010.04.034

morphological evolution related to feeding ecology. The bar-headed goose (*Anser indicus*) breeds at high elevations in the steppes of central Asia and migrates twice annually across the Himalayas, where the partial pressure of oxygen (P_{O_2}) is one-third of that at sea level (Scott and Milsom, 2007). In South America, the Andean goose (*Chloephaga melanoptera*) is endemic to the high Andes, where it is similarly challenged by low O_2 availability (Hiebl et al., 1987). At high elevations, the low P_{O_2} of inspired air may reduce the O_2 saturation of arterial blood and limit O_2 supplies to the tissues (Powell et al., 2004). Bar-headed goose and Andean goose, like other native highland species, exhibit increased blood–O_2 affinity, which can be adaptive in high-altitude environments (Hall et al., 1936; Petschow et al., 1977; Monge and León-Velarde, 1991). In each species, increased O_2 affinity has been attributed to the effects of a single amino acid substitution in the major (HbA) hemoglobin (Jessen et al., 1991; Weber et al., 1993). In bar-headed goose, a Pro → Ala substitution at position 119 on the α^A subunit eliminates an αβ intersubunit van der Waals contact that shifts O_2 equilibrium in favor of the oxygenated (R-state) conformation of the protein and thereby increases hemoglobin–O_2 affinity (Fig. 1). Surprisingly, the same mechanism appears to increase O_2 affinity in Andean goose hemoglobin, but the effect depends on a different substitution on a different gene. In this species, a Leu → Ser substitution at position 55 on the β^A subunit eliminates the same αβ intersubunit contact as in bar-headed goose (Fig. 1). The mechanistic basis of parallel evolution in these two high-altitude waterfowl species thus appears to be essentially the same, but the underlying DNA sequence and amino acid changes are independent and occur on different globin-chain polypeptide subunits in each species.

Despite superficial morphological similarities related to upland grazing, the bar-headed goose and Andean goose are not close relatives. The bar-headed goose is a "true" goose (subfamily Anserinae) and belongs to a clade (*Anser* plus *Branta*) that is restricted to the Northern Hemisphere (Delacour and Mayr, 1945; Johnsgard, 1978; Livezey, 1986; Kear, 2005). The Andean goose, in contrast, is a member of a distantly related clade of sheldgeese, comprising six species (*Neochen* and five species of *Chloephaga*) that are endemic to South America. The sheldgeese, as part of the highly diverse subfamily Anatinae, are more closely related to various kinds of ducks than to the true geese.

Here we present phylogenetic analyses of the swans and geese and the South American sheldgeese using three genes that encode the major (HbA: α^A/β^A) and minor (HbD: α^D/β^A) hemoglobin isoforms, the two protein components in red blood cells that transport O_2 in adult birds (Weber, 2007; Bulgarella et al., 2009). Our objectives were (1) to test whether two amino acid substitutions (Ala-α^A119 and Ser-β^A55), which have been the focus of extensive biochemical analysis, are in fact uniquely derived in bar-headed goose and Andean goose, respectively, and (2) to examine phylogenetic and structural evidence of molecular adaptation at other positions in hemoglobin genes by comparing these two high-altitude taxa to their closest relatives. A recently completed phylogenetic analysis based on mtDNA sequence data includes almost all species of swans and geese (Gonzalez et al., 2009), but no comprehensive molecular phylogenetic analysis of the South American sheldgeese has been completed. Although the HbA isoform of bar-headed goose and greylag goose (*Anser anser*) has been examined using protein crystallography (Zhang et al., 1996; Wang et al., 2000; Liang et al., 2001a,b; Liu et al., 2001), no such studies have been undertaken for the Andean goose, and there has been no phylogenetic or comparative structural analysis of the HbD isoform. Comparing patterns of sequence evolution between the two paralogous genes (α^A vs. α^D) that differentiate the HbA and HbD isoforms is thus also of interest.

2. Materials and methods

2.1. Specimen collection, PCR, and DNA sequencing

Sixty-eight tissue samples representing 20 of the 23 species of swans and geese ($n = 36$) and all six species of South American sheldgeese ($n = 32$) were collected from museums or captive waterfowl collections. Among these were three bar-headed geese, including two from Mongolia collected at 1160 and 2600 m, and 20 Andean geese collected at 3743–4611 m at widespread localities in Peru, Bolivia, and Argentina (Supplementary Table 1).

DNA was isolated using DNeasy Tissue Kits (QIAGEN, Valencia, CA). Primers flanking the start and stop codons for the α^D, α^A, and β^A hemoglobin subunit genes were designed using duck, chicken (*Gallus gallus*), and other sequences in GenBank (Supplementary Table 2). PCR was performed as follows using AmpliTaq Gold PCR Master Mix (Applied Biosystems, Foster City, CA): 7 min preheat at 94 °C, followed by 45 cycles of 20 s at 94 °C, 20 s at 60–64 °C, 1–3 min at 72 °C, and a final extension of 7 min at 72 °C. The complete protein-coding region of each gene, comprising three exons and two introns, was amplified either as a single fragment or in two or three overlapping fragments. Product from an initial PCR spanning the complete β^A subunit was used as the template for a second nested PCR using multiple combinations of internal and end primers. Gel-purified PCR products were then cycle-sequenced in both directions using BigDye Terminator Cycle Sequencing Kits on ABI 3100 automated DNA sequencers (Applied Biosystems, Foster City, CA). Sequences were reconciled using Sequencher 4.6 (Gene Codes, Ann Arbor, MI), and double-peaks, indicating the presence of two alleles, were coded with IUPAC degeneracy codes and treated as polymorphisms. Sequences from samples that were

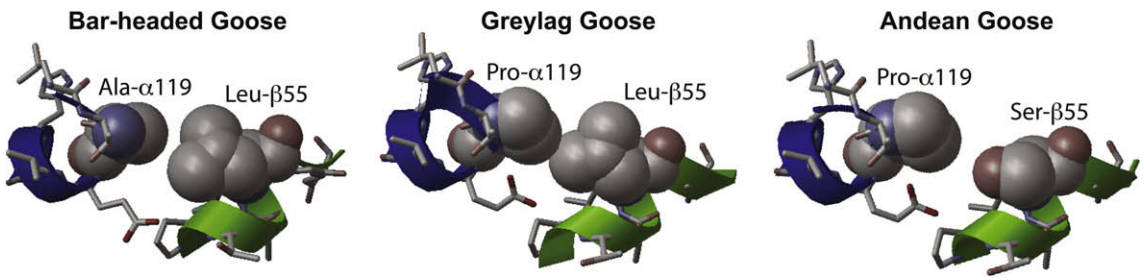

Fig. 1. Illustration of the α^A119/β^A55 intersubunit contact in bar-headed goose, greylag goose, and Andean goose. Pro-α^A119 and Leu-β^A55 lie adjacent to each other on different polypeptide subunits, and in the greylag goose make a van der Waals contact that stabilizes the deoxygenated (T-state) conformation of the HbA structure. van der Waals interactions between α^A119 and β^A55 are eliminated by substitution of Ala-α^A119 in bar-headed goose and Ser-β^A55 in Andean goose. Protein crystal structures for greylag goose and bar-headed goose were illustrated using Protein Data Bank 1FAW and 1A4F, respectively. The Andean goose structure was illustrated using 1FAW, with a Leu → Ser-β^A55 replacement modeled in Swiss-PdbViewer 3.7. The structures were illustrated with Python Molecular Viewer 1.5.2.

heterozygous for an indel were resolved by comparing forward and reverse strands, using the unambiguous 5′ sequence on one strand to interpret the offset peaks 3′ of the indel on the other strand (see Peters et al., 2007). Sequences were aligned by eye using Se–Al 2.0a11 (Rambaut, 2007) and are deposited in GenBank (α^D HM125082–HM125128; α^A GQ271002–GQ271003, GQ271019–GQ271050, GQ271481–GQ271513; β^A GQ271748–GQ271749, GQ271768–GQ271792, GQ272207–GQ272227). The α^D, α^A, and β^A subunit sequence alignments are provided as supplementary NEXUS files.

2.2. Phylogenetic analyses

Phylogenetic analyses were conducted separately for each gene, and using concatenated sequences from all three genes. Gene trees were constructed with MrBayes 3.1 (Ronquist and Huelsenbeck, 2003). Each MrBayes analysis included two independent Markov chain Monte Carlo runs of 3 heated chains and 1 cold chain; 10 million generations were completed with trees sampled every 1000 steps. The first 90% of sampled trees were discarded, and the final 1000 trees from each of the two independent runs were combined to compute posterior probabilities for each clade. Tracer 1.3 (Rambaut and Drummond, 2005) was used to verify convergence across independent runs. Substitution models were selected using the Bayesian Information Criterion (BIC; Schwarz, 1978) as calculated by Modeltest 3.7 (Posada and Crandall, 1998) using output from PAUP* (Swofford, 2002). Seven different data partitions (1st, 2nd, and 3rd codon positions, all three codon positions combined, intron 1, intron 2, intron 1 and 2 combined) and five mixed models were evaluated using hierarchical BIC, with gaps coded separately as binary characters. The best-fit models for the single data partitions were the HKY + I (α^A), HKY + G (α^D), and TrN + G (β^A), respectively, and the best-fit model overall for each locus was the five-partition model with separate rate parameters for 1st, 2nd, and 3rd positions, as well as intron 1 and 2, modeled separately (Supplementary Table 3). For the concatenated data set, we used a parameter-rich model, with five data partitions for each subunit. We also performed a non-parametric bootstrap of the concatenated data set using 1000 unweighted parsimony replicates in PAUP*, with 10 random addition replicates each limited to 100

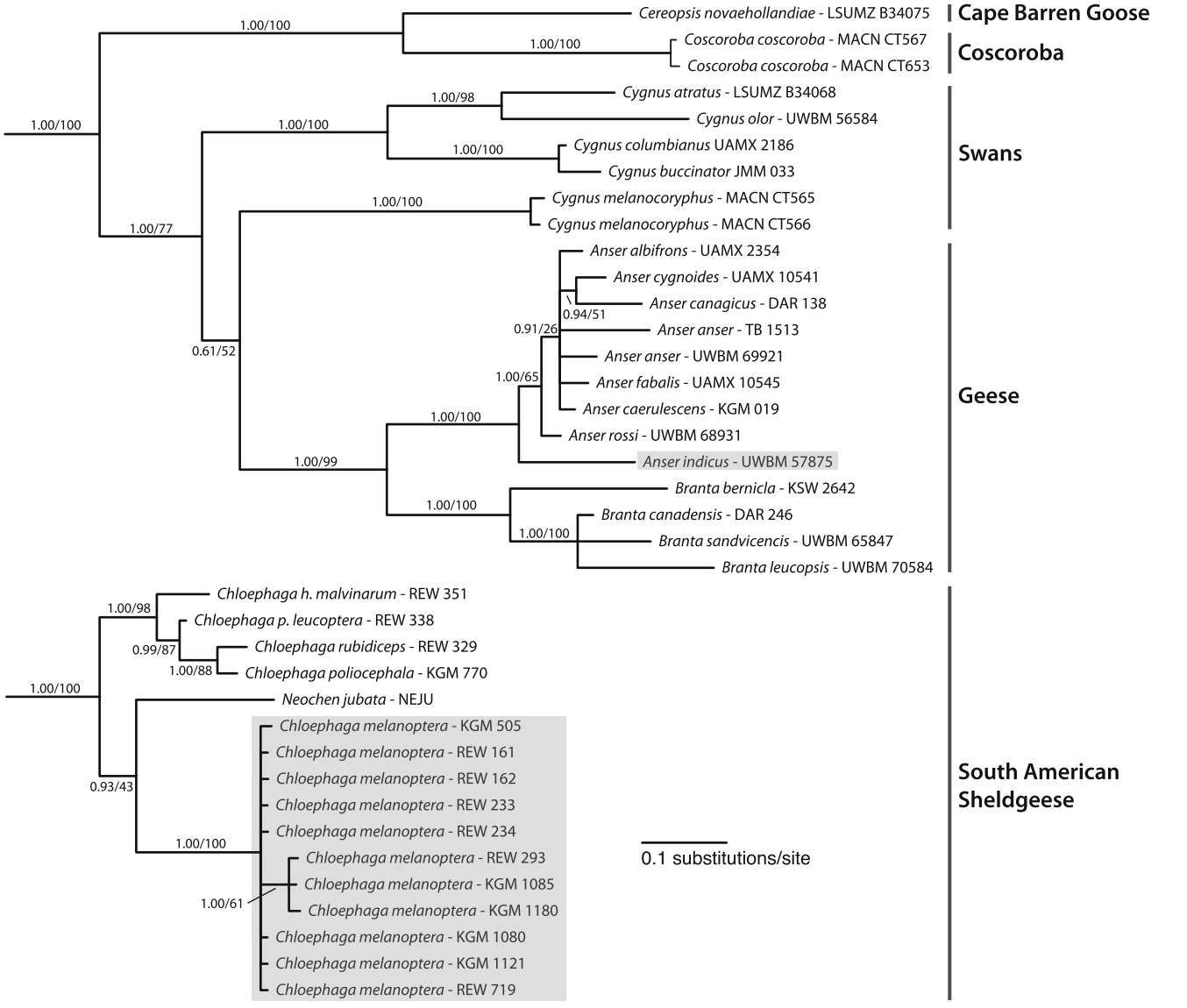

Fig. 2. Fifty-percent majority-rule consensus trees based on concatenated α^A, α^D, and β^A hemoglobin subunit sequences of the swans and geese and the South American sheldgeese, using a mixed model Bayesian phylogenetic analysis. Posterior probabilities (left) and parsimony bootstrap values (right) are shown for each clade. Highland lineages are shown in gray boxes.

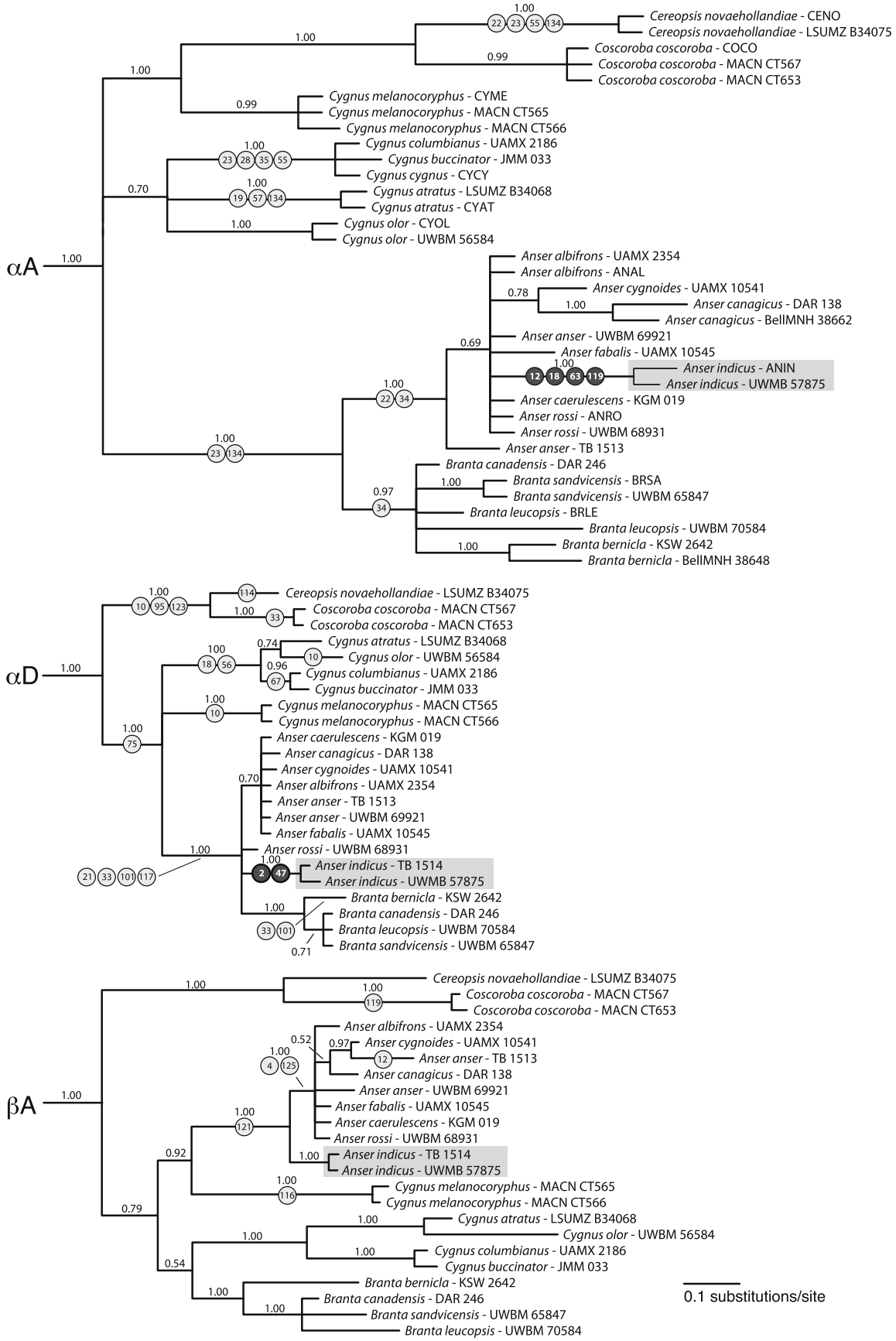

Fig. 3. Fifty-percent majority-rule consensus trees for the α^A, α^D, and β^A hemoglobin subunits of the swans and geese, based on mixed model Bayesian phylogenetic analyses. Posterior probabilities are shown for each clade. Highland lineages are shown in gray boxes. The positions of amino acid replacements within each polypeptide sequence are shown in light (lowland lineage) or dark (highland lineage) circles on each branch.

trees per replicate. Start and stop codons and identical sequences were omitted from each analysis. Trees were rooted on the branch separating the swans and geese (Anserinae) from the sheldgeese (Anatinae) because the reciprocal monophyly of these two groups is strongly supported by varied data and analyses (Livezey, 1986; Sibley and Ahlquist, 1990; Sorenson et al., 1999; Donne-Goussé et al., 2002; Bulgarella et al., 2010). Two monotypic genera of African "sheldgeese" (*Cyanochen*, *Alopochen*) were not included in our analysis. *Cyanochen* is not a true sheldgoose (Sorenson et al., 1999; Gonzalez et al., 2009; Bulgarella et al., 2010), and *Alopochen* is more closely related to shelducks (*Tadorna*) than to the South American sheldgeese (Donne-Goussé et al., 2002; Gonzalez et al., 2009; Bulgarella et al., 2010).

2.3. Structural analyses of amino acid replacements

The position of each α^A- and β^A-subunit amino acid replacement was located on the oxy (R-state) crystal structure of the bar-headed goose or greylag goose (*A. anser*) HbA hemoglobin (Zhang et al., 1996; Liang et al., 2001a; Protein Data Bank 1A4F and 1FAW). Since no HbD structure is available for any species of waterfowl, we used Swiss-PdbViewer 3.7 (Guex and Peitsch, 1997) and the SWISS-MODEL protein structure homology-modeling server (Arnold et al., 2006) to construct a model of the α^D subunit and the HbD tetramer using the chicken HbD oxy (R-state) hemoglobin (Knapp et al., 1999; 1HBR) as a structural template. Inter-atomic distances between neighboring amino acids were calculated using Cn3D 4.1 (National Institutes of Health, Bethesda, MD) and Swiss-PdbViewer 3.7, using a maximum van der Waals distance of 4.1 Å or less (Dall'Acqua et al., 1998). Changes in chemical properties such as polarity, isoelectric point, molecular volume, and normalized van der Waals volume were calculated using published values (e.g., Grantham, 1974). Amino acid replacements were reconstructed on each gene tree using unweighted parsimony in MacClade 4.6 (Maddison and Maddison, 2000). Protein crystal structures were illustrated using Python Molecular Viewer 1.5.2 (Scripps Research Institute, La Jolla, CA). The solvent accessible molecular surface was calculated using the default software settings with a probe radius of 1.5 Å. The electrostatic charge on the surface of the molecule was calculated using the Adaptive Poisson–Boltzmann Solver (APBS) v0.5.1 software (Baker et al., 2001).

3. Results

3.1. Summary of the data

Sequences of the α^A, α^D, and β^A subunits varied from 662 to 680, 873 to 928, and 1563 to 1591 bp, respectively, encoding 141, 140, and 146 amino acids. Several small indels were observed in the two introns of each gene, but no codon insertions or deletions were observed. Amino acid sequences of the α^A and α^D subunits were 57.1–60.7% identical to each other within each species we sampled, whereas the α^D subunits were 80.7–83.6% identical to the chicken α^D subunit (Supplementary Table 4).

3.2. Phylogenetic relationships

Concatenated α^A, α^D, and β^A sequences produced a well-resolved phylogeny of the swans and geese at the generic level, with high posterior probabilities and bootstrap values for most clades (Fig. 2). *Cereopsis* and *Coscoroba* are each other's closest relatives and the sister group of other swans and geese. Due to uncertain placement of *Cygnus melanocoryphus*, the genus *Cygnus* was paraphyletic in most analyses, with the α^A, α^D, and β^A gene trees yielding conflicting results (Fig. 3). A sister relationship between the goose genera *Anser* and *Branta* was well supported in the combined analysis, and the bar-headed goose is likely sister to all other *Anser* species (but see Gonzalez et al., 2009).

Within the South American sheldgeese, the concatenated tree (Fig. 2) and all three Hb gene trees (Fig. 4) yielded strong support for a clade of lowland *Chloephaga* excluding Andean goose. The relationships among *Neochen*, Andean goose, and the other *Chloephaga* varied among genes (Fig. 4); the combined analysis suggests a sister relationship between *Neochen* and Andean goose, although this was not strongly supported (Fig. 2).

3.3. Amino acid replacements in highland lineages

Six derived amino acid replacements were found in the bar-headed goose, but not other species of the genus *Anser* (Fig. 5). These included four replacements on the α^A subunit (Gly → Ala-α^A12, Gly → Ser-α^A18, Ala → Val-α^A63, Pro → Ala-α^A119) and two on the α^D subunit (Thr → Ser-α^D2, Leu → Val-α^D47). One individual collected at 2600 m in Mongolia (UWBM 57875) was heterozygous for Gly/Ala-α^A12; the other individual was homozygous for Gly. No derived replacements were observed on the β^A subunit of bar-headed goose. Glu-β^A121 is a synapomorphy for *Anser*, and Thr-β^A4 and Asp-β^A125 are the presumably ancestral states found also in *Cygnus* and *Branta* (but not other *Anser*). The only other amino acid replacement occurring within the genus *Anser* was a Thr → Ala-β^A12 substitution in one of two greylag geese collected in Mongolia (UAM TB 1513); this specimen was homozygous for Ala-β^A12.

Four derived amino acid replacements were observed in Andean goose but not *Neochen* or other *Chloephaga* species

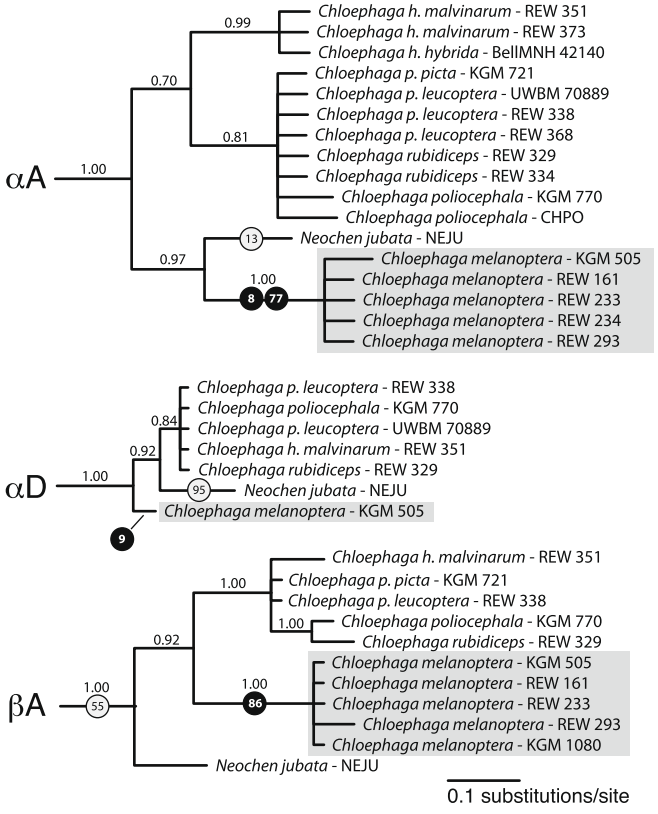

Fig. 4. Fifty-percent majority-rule consensus trees for the α^A, α^D, and β^A hemoglobin subunits of the South American sheldgeese, based on mixed model Bayesian phylogenetic analyses. Posterior probabilities are shown for each clade. Highland lineages are shown in gray boxes. The positions of amino acid replacements within each polypeptide sequence are shown in light (lowland lineage) or dark (highland lineage) circles on each branch.

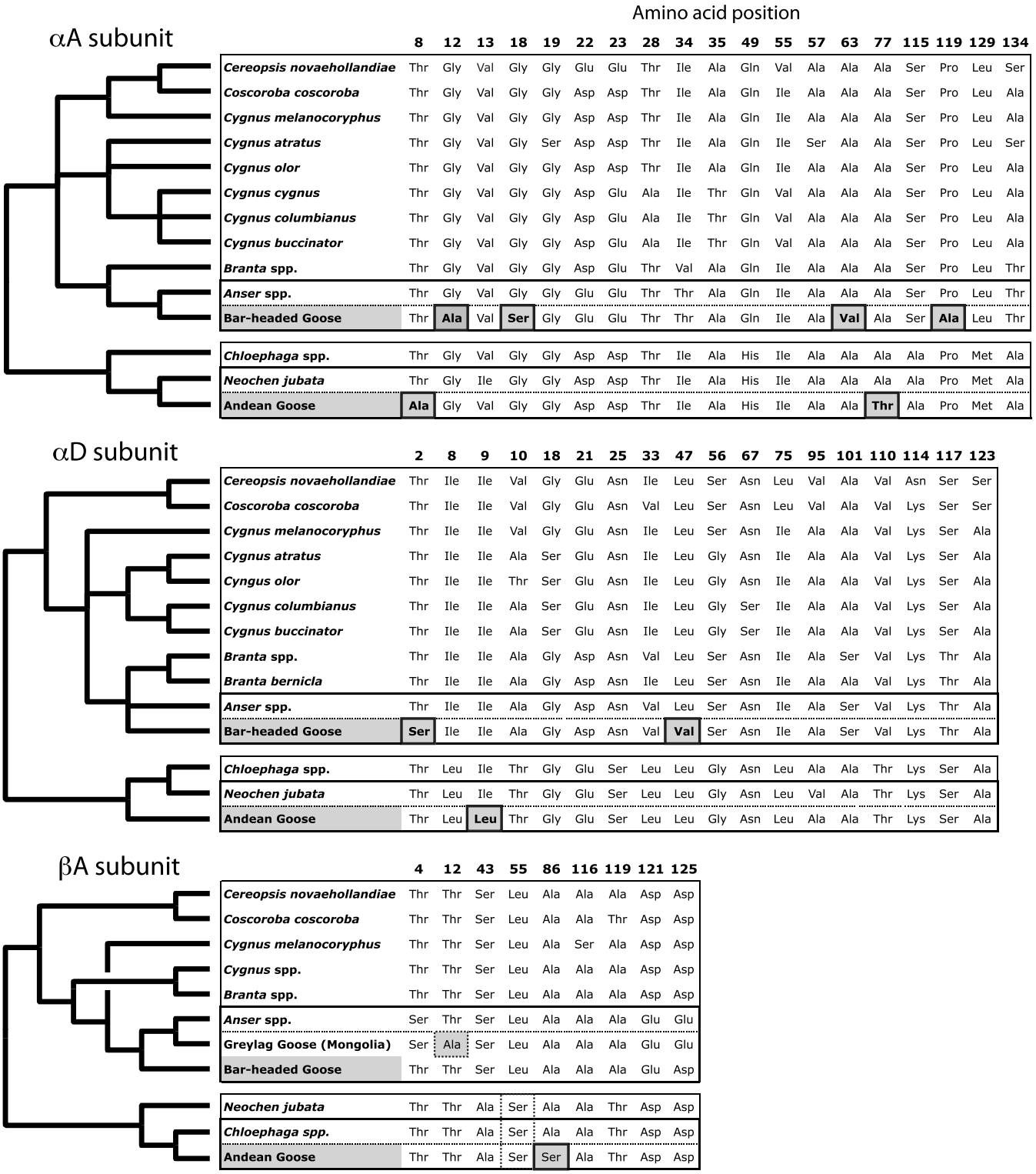

Fig. 5. Variable amino acid positions in the α^A, α^D, and β^A hemoglobin subunits of the swans and geese and the South American sheldgeese. Highland lineages are shown in gray boxes. The 50% majority-rule consensus tree for each gene is shown at left.

(Fig. 5). These included two replacements on the α^A subunit (Thr → Ala-α^A8, Ala → Thr-α^A77), one replacement on the α^D subunit (Ile → Leu-α^D9), and one replacement on the β^A subunit (Ala → Ser-β^A86). All 20 Andean geese that were sampled were homozygous for all of these residues. Ser-β^A55, which has been the focus of previous analyses of Andean goose hemoglobin, is a synapomorphy for *Neochen* and *Chloephaga* and is thus present in five additional sheldgeese species that are presently found only in lowland environments. Two nonsynonymous substitutions were found in *Neochen* (Val → Ile-α^A13, Ala → Val-α^D95), but none were found in any of the five lowland *Chloephaga* species.

4. Discussion

4.1. Taxonomic conclusions

Our findings based on the α^A, α^D, and β^A hemoglobin subunits are generally in agreement with previous analyses of waterfowl based on morphology (Livezey, 1996a,b, 1997). Monophyly and the close relationship of *Anser* and *Branta* were strongly supported, whereas our data were equivocal regarding the monophyly of *Cygnus* due to an uncertain placement of *C. melanocoryphus*. Consistent with a recent analysis based on mtDNA (Gonzalez et al., 2009), the hemoglobin data indicate that *Cygnus atratus* is sister to *Cygnus olor*, rather than to *C. melanocoryphus* (see Livezey, 1996a). Likewise, both hemoglobin and mtDNA indicate that the South American *C. melanocoryphus* is a divergent lineage that is likely the sister taxon of the remaining *Cygnus* species (Gonzalez et al., 2009). Constraining *Cygnus* monophyly in our analysis of the concatenated data produces this result and requires just one additional parsimony step, an insignificant increase in tree length ($P > 0.52$; Kishino–Hasegawa (1989) test).

As in other molecular analyses, we also found strong support for a sister relationship between *Cereopsis* and *Coscoroba* (Laskowski and Fitch, 1989; Donne-Goussé et al., 2002; Gonzalez et al., 2009), a result confirmed by shared CR1 retrotransposons (St. John et al., 2005). Similarly, the hemoglobin data suggest that *Cereopsis* and *Coscoroba* form the sister group of all other swans and geese. In contrast, Livezey (1996a) placed *Cereopsis* alone as the sister group to other Anserinae and placed *Coscoroba* sister to *Cygnus* based in part on six synapomorphic characters in the synsacrum, sternum, coracoid, and pelvis. Similarities between *Coscoroba* and *Cygnus* have thus likely arisen through morphological convergence.

The hemoglobin data suggest that bar-headed goose is the sister group of all other *Anser* species included in this study. In contrast, Gonzalez et al. (2009) found bar-headed goose to be embedded within *Anser* based on sequences from two mitochondrial genes. In our study, intra-generic relationships within *Anser* and *Branta* were poorly resolved, owing to short branch lengths, with different gene trees supporting different topologies. In the case of the three hemoglobin genes sequenced here, it is likely that the small number of variable sites and incomplete lineage sorting both contribute to lack of resolution within genera. Additional data from the mitochondrial genome or many more nuclear loci likely will be required to resolve these relationships.

Like Livezey (1997), we also found strong support for a clade of lowland *Chloephaga* excluding Andean goose. The concatenated data suggest that *Neochen* is sister to Andean goose, and that these two species from northern South America form a clade, sister to the four lowland *Chloephaga* species endemic to southern South America. Additional data will also be required to resolve the relationships of these taxa.

4.2. Amino acid replacements in highland lineages

Structural changes in the HbA isoform arising from substitutions at α^A18, α^A63, α^A119, and β^A125 have been examined extensively using the protein crystal structures of bar-headed goose and greylag goose (Zhang et al., 1996; Wang et al., 2000; Liang et al., 2001a,b; Liu et al., 2001). Additionally, the effects of Ala-α^A119 also have been measured experimentally using site-directed mutagenesis to incorporate this substitution into human hemoglobin (Jessen et al., 1991; Weber et al., 1993). Our findings provide insight into the evolutionary pathways by which these amino acid replacements occurred within the Anserinae. Ser-α^A18, Val-α^A63, and Ala-α^A119 are each derived residues, present only in bar-headed goose and absent in greylag goose and other *Anser* species. In contrast, Asp-β^A125 is not derived, but is the ancestral residue present in most Anserinae and other basal waterfowl lineages (McCracken et al., 2009a). Based on phylogenetic analysis of the concatenated hemoglobin sequences, Glu-β^A125 is a synapomorphy for a clade comprising all *Anser* species except bar-headed goose. We also observed a Gly → Ala-α^A12 replacement in a single bar-headed goose specimen from Mongolia. This residue occurs at an exterior, solvent accessible position on the A helix of the α^A subunit (Fig. 6), a region of the HbA isoform that has experienced parallel changes in three other high-altitude waterfowl species, including the Andean goose (as discussed below; see also McCracken et al., 2009a,b). Thus, it would be worthwhile to measure the frequency of this polymorphism in the bar-headed goose population and also determine if it affects hemoglobin–O_2 affinity. Additionally, we do not yet know whether this polymorphism represents two allelic copies of a single α^A-subunit gene, or whether bar-headed goose possesses duplicate α^A-subunit genes distinguished by a single amino acid change. An example of this kind of similarity between gene copies is maintained by recurrent gene conversion in the

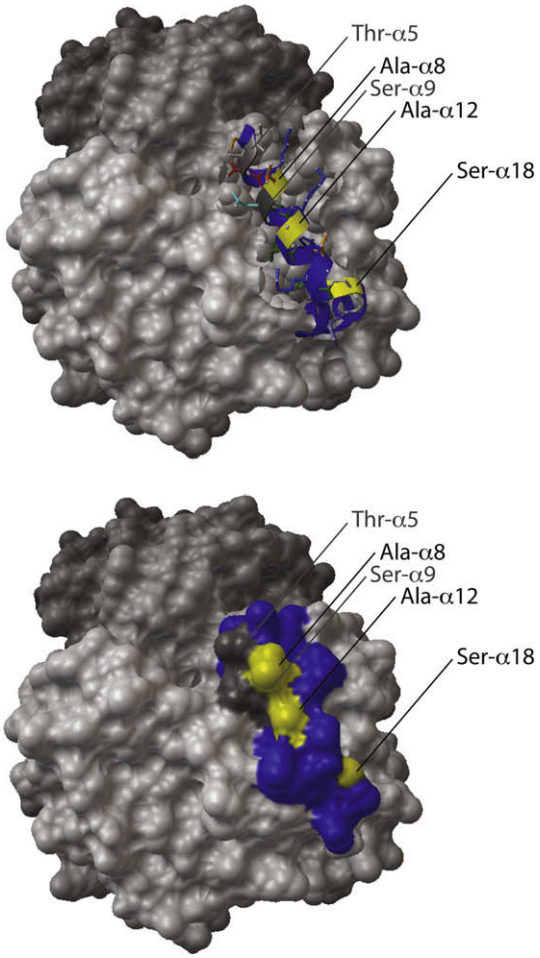

Fig. 6. HbA structure illustrating three amino acid substitutions (Thr → Ala-α^A8, Gly → Ala-α^A12, and Gly → Ser-α^A18 shown in yellow) at external, solvent accessible positions on the A helix and AB corner of the α^A subunit (shown in blue). Substitutions at two adjacent sites that occurred in other highland waterfowl species (McCracken et al., 2009a) are shown in faded gray. The top panel shows the internal three-dimensional 'ribbon' structure of the A helix and AB corner, whereas the bottom panel shows the external, solvent-accessible surface of the protein with a probe radius of 1.5 Å. The protein crystal structure was illustrated using Protein Data Bank 1FAW and Python Molecular Viewer 1.5.2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

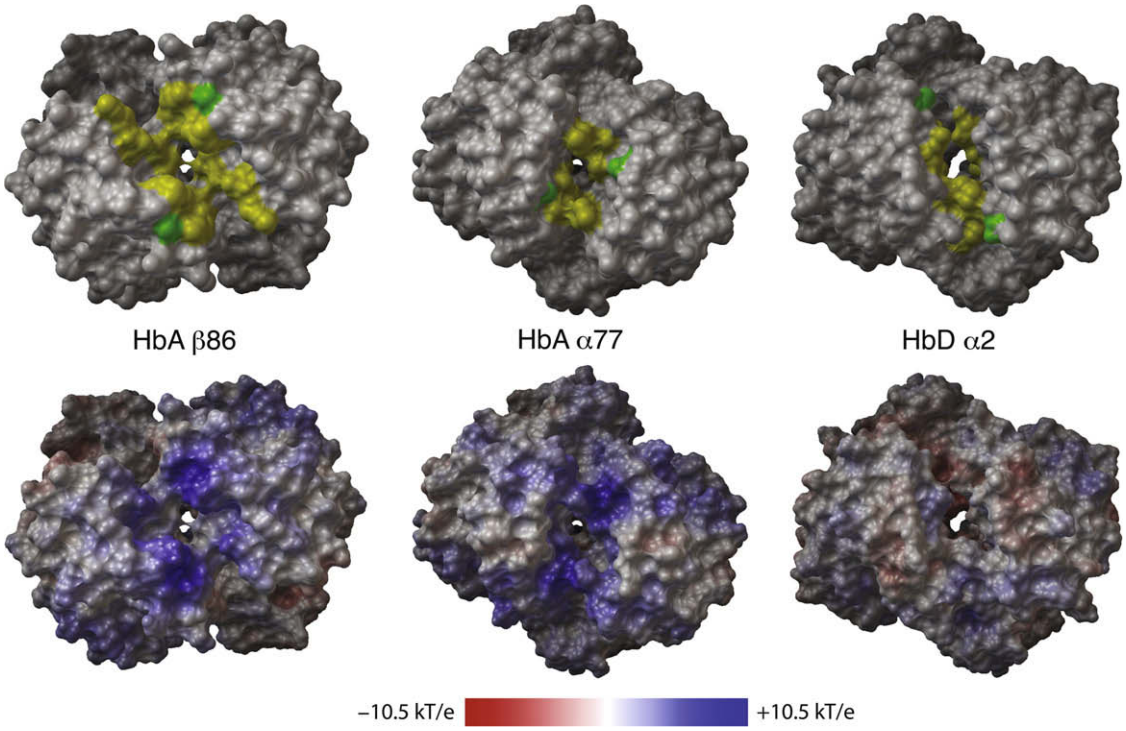

Fig. 7. HbA and HbD structures illustrating three amino acid substitutions (Ala → Ser-β^A86, Ala → Thr-α^A77, and Thr → Ser-α^D2 shown in green) that occurred in close proximity to inositolpentaphosphate (IP_5) binding sites (shown in yellow). The HbA IP_5 binding sites include β^A1, 2, 84, 104, 135, 139, 143, 144, and 146 (Wang et al., 2000; top left), and α^A1, 95, 99, 134, 137, 138, and 141 (Tamburrini et al., 2000; Riccio et al., 2001; top middle); the same homologous α^D sites are shown for the HbD isoform of chicken (top-right). The electrostatic charge on the surface of the molecule is illustrated for each structure at bottom. Blue surfaces correspond to positively charged regions, and red surfaces correspond to negatively charged regions. In the HbA isoform, the IP_5 binding sites shown in yellow (top row) occur within regions of dense positive charge (bottom row). Protein crystal structures for HbA and HbD were illustrated using Protein Data Bank 1FAW and 1HBR, respectively, and Python Molecular Viewer 1.5.2. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

duplicated α and β globin genes of deer mice (*Peromyscus maniculatus*; Storz et al., 2007, 2009).

We also observed two derived substitutions on the α^D subunit of bar-headed goose (Thr → Ser-α^D2 and Leu → Val-α^D47) that were previously identified by protein sequencing (Hiebl et al., 1986). Thr → Ser-α^D2 occurs at the N-terminus of the α^D subunit in a region of the protein tetramer that may function in inositolpentaphosphate (IP_5) binding (Fig. 7), based on corresponding IP_5 binding sites identified in the South Polar skua (*Catharacta maccormicki*; Tamburrini et al., 2000; Riccio et al., 2001). IP_5 is the principal allosteric effector in avian red blood cells. It influences the equilibrium between deoxygenated (T-state) and oxygenated (R-state) hemoglobin by binding to positively charged sites between in the central cavity between the N- and C-termini of the α- and β-chain subunits, thereby stabilizing the deoxygenated conformation of the protein with salt-bridges. Changes that suppress phosphate binding are thus expected to reduce hemoglobin–O_2 affinity (Storz et al., 2009).

The total number of derived amino acid substitutions in the HbA and HbD isoforms of bar-headed goose was thus four and two, respectively. No replacements were observed on the β^A subunit, which is a component of both HbA and HbD.

Andean goose exhibited a total of four derived amino acid substitutions. Two substitutions occurred on the α^A subunit (Thr → Ala-α^A8 and Ala → Thr-α^A77), one occurred on the α^D subunit (Ile → Leu-α^D9), and one occurred on the β^A subunit (Ala → Ser-β^A86). Thr-α^A77 and Ser-β^A86 each occur in close proximity to IP_5 binding sites and in regions of positive charge (Fig. 7; Wang et al., 2000). Substitution of threonine and serine each result in non-polar to polar changes that correspond to the addition of a hydroxyl group and increased molecular volume. It is thus possible that the hydroxyl group substitutions influence IP_5 binding indirectly; IP_5 is a negatively charged ligand. The addition of threonine or serine might also enhance hemoglobin solubility, and this could be beneficial in desiccating, high-altitude environments. The likely functional significance of Thr-α^A77 is also supported by the fact that the same substitution is independently derived in four other waterfowl species inhabiting high-altitude regions (McCracken et al., 2009a,b).

Intriguingly, the two other derived replacements we observed in Andean goose, Ala-α^A8 and Leu-α^D9, occur at nearby positions on the A helix of the α^A and α^D subunits (α^D9 corresponds to α^A10 in HbD and HbA, respectively, due to one amino acid difference in the length of the α^D- and α^A-chain subunits). Highland populations of two other Andean waterfowl species also have derived amino acid substitutions in this structural domain of the HbA tetramer; crested duck (*Lophonetta specularioides*; Thr-α^A5 and Ala-α^A8) and cinnamon teal (*Anas cyanoptera*; Ser-α^A9; McCracken et al., 2009a). One noteworthy difference, however, is that while α^A5, 8, 9, 12, and 18 each occupy external, solvent accessible positions in the HbA structure (Fig. 6), α^D9 occurs at an internal position within the chicken HbD structure.

Finally, the finding that Ser-β^A55 is not derived in Andean goose, but is a synapomorphy for the South American sheldgeese (*Neochen* + *Chloephaga*) raises questions about its physiological function in species native to lowland environments. Perutz (1983, 1989) hypothesized that Leu → Ser-β^A55 destabilizes the deoxygenated (T-state) conformation of the protein by eliminating a van der Waals contact at an important αβ intersubunit contact (Fig. 1), thereby shifting the allosteric equilibrium in favor of the oxygenated (R-state) structure. When Ser-β^A55 was experimentally introduced into recombinant human hemoglobin, this pre-

dicted effect was observed (Jessen et al., 1991; Weber et al., 1993). This raises the intriguing hypothesis that Ser-β^A55 may have resulted in elevated hemoglobin–O_2 affinity in species that presently are restricted to lowland environments, such that the South American sheldgeese exhibit higher average blood–O_2 affinity than other lowland endemic waterfowl species. This hypothesis remains to be tested experimentally, but if so, it might be used to infer the direction of colonization of the Andes by the sheldgeese. If the ancestral allele for this clade confers elevated hemoglobin–O_2 affinity, the South American sheldgeese might have originated in the highlands, but more recently colonized lowland regions.

5. Conclusion

Our study of the swans and geese and the South American sheldgeese represents one of the first comprehensive molecular phylogenetic analyses of these taxa and provides a historical basis for understanding patterns of amino acid substitution in the HbA and HbD genes of bar-headed goose and Andean goose, two species of waterfowl that have featured prominently in studies of high-altitude adaptation. Bar-headed goose and Andean goose have adapted to similar hypoxic environments in the Himalayas and Andes, and thus experience the same hypobaric selection pressures. While these two species share no identical amino acid substitutions in either the HbA or HbD subunit genes, similar types of substitutions have occurred within the same structural regions of the protein, including the same structural domains of the α^A and α^D subunits. The mechanistic basis of hypoxia resistance in these two waterfowl species has yet to be fully determined, but the patterns identified here provide an example of parallel evolution of sites within the same hemoglobin protein domains. In addition to the previously described amino acid substitutions that influence an important $\alpha\beta$ intersubunit contact, additional substitutions that suppress the phosphate sensitivity of hemoglobin–O_2 affinity may also likely be important. Further experimental studies are required to determine how the observed substitutions influence hemoglobin–O_2 binding properties.

It is also important to identify the molecular mechanisms responsible for differences between the HbA and HbD isoforms in their intrinsic O_2 affinity, allosteric regulation, and oxygen-linked polymerization (Cirotto and Geraci, 1975; Baumann et al., 1984; Riggs, 1998; Knapp et al., 1999; Rana et al., 2008). HbD has been shown to have higher O_2 affinity than HbA, and the ratio of these two isoforms in avian red blood cells may play an important role in modulating blood–O_2 affinity (Hiebl et al., 1988; Weber et al., 1988). For example, it is possible that highland species maintain elevated blood–O_2 affinity by upregulating expression of the α^D gene, thereby increasing the relative abundance of the HbD isoform (Hoffmann and Storz, 2007; Bulgarella et al., 2009).

In conclusion, our findings illustrate the importance of understanding phylogenetic relationships and polarity of character-state changes when making inferences about adaptive evolution. A comparative phylogenetic approach reveals the evolutionary origins of substitutions of putative functional significance. For example, we found that a substitution showing evidence of physiological adaptation was not uniquely derived in a high-altitude lineage. Similarly, recurrent patterns of substitution in relation to protein structure across independent lineages can lead to testable hypotheses about function that can be used to guide future experimental and comparative studies.

Acknowledgments

We thank Mariana Bulgarella, Cecilia Kopuchian, Thomas Valqui, Robert Wilson, Instituto Nacional de Recursos Naturales (INRENA) and Centro de Ornitología y Biodiversidad (CORBIDI) in Peru, Colección Boliviana de Fauna, Subsecretaría del Ambiente de Catamarca, Dirección de Fauna y Flora Silvestre del Chubut, Dirección de Fauna de Santa Cruz in Argentina, and the Falkland Islands Government for their assistance collecting specimens. Sharon Birks at the Burke Museum of Natural History and Culture, Pablo Tubaro at the Museo Argentino de Ciencias Naturales "Bernardino Rivadavia", Mike Lubbock at Sylvan Heights Waterfowl, and Kevin Winker at the University of Alaska Museum provided tissue loans. Jay Storz, Jeremy Tame, and an anonymous reviewer provided helpful comments that improved the manuscript. Funding was provided by the Institute of Arctic Biology, Alaska EPSCoR (NSF EPS-0092040 and EPS-0346770), Grants from the Delta Waterfowl Foundation, Frank M. Chapman Fund, and NSF (DEB-0089760 and DEB-0444748).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2010.04.034.

References

Arnold, K., Bordoli, L., Kopp, J., Schwede, T., 2006. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics 22, 195–201.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., McCammon, J.A., 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. Proc. Natl. Acad. Sci. USA 98, 10037–10041.

Baumann, R., Goldbach, E., Haller, E.A., Wright, P.G., 1984. Organic phosphates increase the solubility of avian haemoglobin D and embryonic chicken haemoglobin. Biochem. J. 217, 767–771.

Bulgarella, M., Stewart, N.C., Fedorov, V.B., Moore, A.V., McCracken, K.G., 2009. Hemoglobin transcript abundance in a cDNA library from bone marrow of crested ducks in the high Andes. Auk 126, 666–672.

Bulgarella, M., Sorenson, M.D., Peters, J.L., Wilson, R.E., McCracken, K.G., 2010. Phylogenetic relationships of *Amazonetta*, *Speculanas*, *Lophonetta*, and *Tachyeres*: four morphologically divergent genera of ducks endemic to South America. J. Avian Biol. 41, 186–199.

Cirotto, C., Geraci, G., 1975. Embryonic chicken hemoglobins: studies on the oxygen equilibrium of two pure components. Comp. Biochem. Physiol. A 51, 159–163.

Colosimo, P.F., Hosemann, K.E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., Schmutz, J., Meyers, R.M., Schluter, D., Kingsley, D.M., 2005. Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. Science 307, 1928–1933.

Dall'Acqua, W., Goldman, E.R., Lin, W., Teng, C., Tsuchiya, D., Li, H., Ysern, X., Braden, B.C., Li, Y., Smith-Gill, S.J., Mariuzza, R.A., 1998. A mutational analysis of binding interactions in an antigen–antibody protein–protein complex. Biochemistry 37, 7981–7991.

Delacour, J., Mayr, E., 1945. The family Anatidae. Wilson Bull. 57, 3–55.

Donne-Goussé, C., Laudet, V., Hänni, C., 2002. A molecular phylogeny of anseriformes based on mitochondrial DNA analysis. Mol. Phylogen. Evol. 23, 339–356.

Feldman, C.R., Brodie Jr., E.D., Brodie III, E.D., Pfrender, M.E., 2009. The evolutionary origins of beneficial alleles during the repeated adaptation of garter snakes to deadly prey. Proc. Natl. Acad. Sci. USA 106, 13415–13420.

Gonzalez, J., Düttmann, H., Wink, M., 2009. Phylogenetic relationships based on two mitochondrial genes and hybridization patterns in Anatidae. J. Zool. 279, 310–318.

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

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

Hall, F.G., Dill, D.B., Guzman Barron, E.S., 1936. Comparative physiology in high altitudes. J. Cell. Comp. Physiol. 8, 301–313.

Hiebl, I., Schneeganss, D., Braunitzer, G., 1986. The primary structures of the α^D-chains of the bar-headed goose (*Anser indicus*), the greylag goose (*Anser anser*) and the Canada goose (*Branta canadensis*). Biol. Chem. Hoppe-Seyler 367, 591–599.

Hiebl, I., Braunitzer, G., Schneeganss, D., 1987. The primary structures of the major and minor hemoglobin-components of adult Andean goose (*Chloephaga melanoptera*, Andatidae): the mutation Leu → Ser in position 55 of the β-chain). Biol. Chem. Hoppe-Seyler 368, 1559–1569.

Hiebl, I., Weber, R.E., Schneeganss, D., Kosters, J., Braunitzer, G., 1988. High-altitude respiration of birds. Structural adaptations in the major and minor hemoglobin components of adult Ruppell's griffon (*Gyps ruppellii*, Aegypiinae): a new molecular pattern for hypoxic tolerance. Biol. Chem. Hoppe-Seyler 369, 217–232.

Hoffmann, F.G., Storz, J.F., 2007. The αD globin gene originated via duplication of an embryonic α-like globin gene in the ancestor of tetrapod vertebrates. Mol. Biol. Evol. 24, 1982–1990.
Jessen, T.H., Weber, R.E., Fermi, G., Tame, J., Braunitzer, G., 1991. Adaptation of bird hemoglobins to high altitudes: demonstrations of molecular mechanism by protein engineering. Proc. Natl. Acad. Sci. USA 88, 6519–6522.
Johnsgard, P.A., 1978. Ducks, Geese, and Swans of the World. University of Nebraska Press, Lincoln, Nebraska.
Jost, M.C., Hillis, D.M., Lu, Y., Kyle, J.W., Fozzard, H.A., Zakon, H.H., 2008. Toxin-resistant sodium channels: parallel adaptive evolution across a complete gene family. Mol. Biol. Evol. 25, 1016–1024.
Kear, J., 2005. Bird Families of the World: Ducks, Geese, and Swans. Oxford University Press, New York.
Kishino, H., Hasegawa, M., 1989. Evaluation of maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. 29, 170–179.
Knapp, J.E., Oliveira, M.A., Xie, Q., Ernst, S.R., Riggs, A.F., Hackert, M.L., 1999. The structural and functional analysis of the hemoglobin D component from chicken. J. Biol. Chem. 274, 6411–6420.
Laskowski Jr., M., Fitch, W.M., 1989. Evolution of avian ovomucoids and of birds. In: Fernholm, B., Bremer, K., Jörnvall, H. (Eds.), The Hierarchy of Life. Elsevier Press, Amsterdam, pp. 371–387.
Liang, Y.H., Liu, X.Z., Liu, S.H., Lu, G.Y., 2001a. The structure of greylag goose oxy haemoglobin: the roles of four mutations compared with bar-headed goose haemoglobin. Acta Crystallogr. D Biol. Crystallogr. 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. J. Mol. Biol. 313, 123–137.
Liu, X.Z., Li, S.L., Jing, H., Liang, Y.H., Hua, Z.Q., Lu, G.Y., 2001. Avian haemoglobins and structural basis of high affinity for oxygen: structure of bar-headed goose aquomet haemoglobin. Acta Crystallogr. D Biol. Crystallogr. 57, 775–783.
Livezey, B.C., 1986. A phylogenetic analysis of recent anseriform genera using morphological characters. Auk 103, 737–754.
Livezey, B.C., 1996a. A phylogenetic analysis of geese and swans (Anseriformes: Anserinae), including selected fossil species. Syst. Biol. 45, 415–450.
Livezey, B.C., 1996b. A phylogenetic analysis of the tadornine–anatine divergence (Aves: Anseriformes: Anatidae). Ann. Carnegie Mus. 65, 27–88.
Livezey, B.C., 1997. A phylogenetic analysis of modern sheldgeese and shelducks (Anatidae, Tadornini). Ibis 139, 51–66.
Maddison, D.R., Maddison, W.P., 2000. MacClade 4.0. Sinaur Associates, Sunderland, Massachusetts.
McCracken, K.G., Barger, C.P., Bulgarella, M., Johnson, K.P., Sonsthagen, S.A., Trucco, J., Valqui, T.H., Wilson, R.E., Winker, K., Sorenson, M.D., 2009a. Parallel evolution in the major haemoglobin genes of eight species of Andean waterfowl. Mol. Ecol. 18, 3992–4005.
McCracken, K.G., Barger, C.P., Bulgarella, M., Johnson, K.P., Kuhner, M.K., Moore, A.V., Peters, J.L., Trucco, J., Valqui, T.H., Winker, K., Wilson, R.E., 2009b. Signatures of high-altitude adaptation in the major hemoglobin of five species of Andean dabbling ducks. Am. Nat. 174, 631–650.
Monge, C., León-Velarde, F., 1991. Phsiological adaptation to high altitude: oxygen transport in mammals and birds. Physiol. Rev. 71, 1136–1172.
Perutz, M.F., 1983. Species adaptation in a protein molecule. Mol. Biol. Evol. 1, 1–28.
Perutz, M.F., 1989. Mechanisms of cooperativity and allosteric regulation in proteins. Q Rev. Biophys. 22, 139–236.
Peters, J.L., Zhuravlev, Y., Fefelov, I., Logie, A., Omland, K.E., 2007. Nuclear loci and coalescent methods support ancient hybridization as cause of mitochondrial paraphyly between gadwall and falcated duck (*Anas* spp.). Evolution 61, 1992–2006.
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. J. Appl. Physiol. 42, 139–143.
Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.
Powell, F.L., Shams, H., Hempleman, S.C., Mitchell, G.S., 2004. Breathing in thin air: acclimatization to altitude in ducks. Respir. Physiol. Neurobiol. 144, 225–235.
Rambaut, A., 2007. Se–Al 2.0a11. University of Oxford, Oxford.
Rambaut, A., Drummond, A., 2005. Tracer 1.3: MCMC Trace Analysis Tool. University of Oxford, Oxford.
Rana, M.S., Knapp, J.E., Holland, R.A.B., Riggs, A.F., 2008. Component D of chicken hemoglobin and the hemoglobin of the embryonic Tammar wallaby (*Macropus eugenii*) self-association upon deoxygenation: effect on oxygen binding. Proteins 70, 553–561.
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? Biophys. J. 81, 1938–1946.
Riggs, A.F., 1998. Self-association, cooperativity and super-cooperativity of oxygen binding by hemoglobins. J. Exp. Biol. 201, 1073–1084.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6, 461–464.
Scott, G.R., Milsom, W.K., 2007. Control of breathing and adaptation to high altitude in the bar-headed goose. Am. J. Physiol. Regul. Integr. Comp. Physiol. 293, R379–R391.
Sibley, C.G., Ahlquist, J.E., 1990. Phylogeny and Classification of Birds: A Study in Molecular Evolution. Yale University Press, New Haven, Connecticut.
Sorenson, M.D., Cooper, A., Paxinos, E.E., Quinn, T.W., James, H.F., Olson, S.L., Fleischer, R.C., 1999. Relationships of the extinct moa-nalos, flightless Hawaiian waterfowl, based on ancient DNA. Proc. R. Soc. Lond. B 266, 2187–2193.
St. John, J., Cotter, J.P., Quinn, T.W., 2005. A recent chicken repeat 1 retrotransposon confirms the Coscoroba-Cape Barren goose clade. Mol. Phylogen. Evol. 37, 83–90.
Storz, J.F., Sabatino, S.J., Hoffmann, F.G., Gering, E.J., Moriyama, H., Ferrand, N., Monteiro, B., Nachman, M.W., 2007. The molecular basis of high-altitude adaptation in deer mice. PLoS Genet. 3, 448–459.
Storz, J.F., Runck, A.F., Sabatino, S.J., Kelly, J.K., Ferrand, N., Moriyama, H., Weber, R.E., Fago, A., 2009. Evolutionary and functional insights into the mechanism underlying high-altitude adaptation of deer mouse hemoglobin. Proc. Natl. Acad. Sci. USA 106, 14450–14455.
Swofford, D.L., 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. Eur. J. Biochem. 267, 6089–6098.
Wang, H.C., Liang, Y.H., Zhu, J.P., Lu, G.Y., 2000. Crystallization and preliminary crystallographic studies of bar-headed goose fluoromethaemoglobin with inostitol hexaphosphate. Acta Crystallogr. D Biol. Crystallogr. 56, 1183–1184.
Weber, R.E., 2007. High altitude adaptations in vertebrate hemoglobins. Respir. Physiol. Neurobiol. 158, 132–142.
Weber, R.E., Hiebl, I., Braunitzer, G., 1988. High altitude and hemoglobin function in the vultures *Gyps ruepelli* and *Aeqypius monarchus*. Biol. Chem. Hoppe-Seyler 369, 233–240.
Weber, R.E., Jessen, T.H., Malte, H., Tame, J., 1993. Mutant hemoglobins (α^{119}-Ala and β^{55}-Ser): functions related to high-altitude respiration in geese. J. Appl. Physiol. 75, 2646–2655.
Zhang, J., Hua, Z., Tame, J.R.H., 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). J. Mol. Biol. 255, 484–493.