Global environmental factors impact the evolution of adult hemoglobins in squamata reptiles (lizards and snakes) and terrestrial turtles.

Juan D. Carvajal-Castro[1*]; Genrietta Yagudayeva[1]; Randy Ortiz[1]; Juan C. Santos[1] Department of Biological Sciences, St. John's University, Jamaica-Queens, NY, USA

[*]carvajaljcastro@gmail.com; juan.carvajal21@my.stjohns.edu

ABSTRACT

Convergent evolution of oxygen transport mechanisms arises from respiratory proteins adapting to similar environmental pressures. We examined this relationship between adult hemoglobin subunits (Hbs: HBA1, HBAD, HBB1, and HBB2) found in land reptiles (lizards, snakes, and turtles) with their global distribution variables: Altitude, latitude, ambient temperature, and biomass production. We found that biomass was positively associated with the synonymous substitution rate (*dS*) of HBAD, while it showed the opposite trend for HBB2 in snakes. Additionally, latitude was negatively related to the *dS* of HBB2 in snakes, but non-significant with other Hbs. Altitude was negatively associated with $\omega = dN/dS$ of HBA1 and HBAD, whereas temperature showed a similar negative trend with the ω of HBAD across reptiles and in HBB2 of snakes. At amino acid sites, we found most were conserved except for eleven (two near the heme-binding pocket) across Hbs. These fast-changing sites shifted from polar to nonpolar residues, showing a pattern seen in high-altitude mammals. Our results highlight that in reptiles (1) Hbs are diversifying at individual amino acid sites while generally some subunits exhibiting lower ω rates at higher altitudes and hotter temperatures, with the later and higher biomass ecosystems also linked to increases in *dS*; (2) HBBs are the most conserved of the Hbs; (3) latitudinal gradients only show a significant association with the *dS* of HBB2 in snakes; and (4) gene conversion events occurred across HBBs in reptiles, which confound their

homology assignation, except for snakes that evidenced a single major duplication in their HBBs.

KEYWORDS: Molecular Evolution, Environment, Adaptation, Hemoglobin, Oxygen-Affinity

SIGNIFICANCE STATEMENT

Hemoglobins are essential oxygen transporters, yet their evolution and the adaptations to various environmental factors remain poorly understood across diverse taxonomic groups. Our study addresses this key knowledge gap in squamata reptiles (i.e., lizards and snakes) and terrestrial turtles by examining how altitude, temperature, and primary production influence the evolution of hemoglobin subunits found in adult animals. Our findings indicate that high-altitude environments may constrain hemoglobin diversification due to low oxygen densities. At the same time, geographic regions with high productivity (e.g., tropical forests) and warmer environments (deserts) are also associated with increased genetic diversity (i.e., synonymous mutations) in these respiratory proteins, likely due to relaxed selection at such high oxygen density locations, usually at low-altitude locations. Additionally, we identified evidence of positive selection at several amino acid sites near the heme-binding regions of hemoglobins in reptiles, including substitutions trend from polar to nonpolar residues, highlighting adaptive changes in response to environmental pressures like low oxygen conditions, such as high-altitude mountains. Our results provide new insights into how environmental factors shape evolution across large taxonomic groups within a global context and highlighting the convergent patterns on respiration proteins in reptile species living in opposite corners of our planet.

INTRODUCTION

Environmental conditions impact biodiversity across various scales, ranging from populations to ecosystems (Jennions and Petrie 1997; Tilman 1999). Most comparative studies

aim to characterize adaptations as phenotypic responses to environmental changes by exploring how trait diversification occurs across lineages. Recent research has focused on elucidating the connection between adaptations and the underlying genes and genotypes by employing genomic-level molecular techniques such as transcriptomics and target-bait exome capture experiments. These studies aim to identify specific loci and their allelic variants underlying phenotypic adaptations. However, these molecular approaches can also help reveal how environmental pressures shaped physiological diversification. Thus, connecting environmental variables with the genotype is key for understanding how phenotypic and metabolic adaptations are encoded in the genome (Atkinson and Sibly 1997; Ligon et al. 2018).

Among the most important agents of selection is the environment. Changes in abiotic factors (e.g., temperature) can drive natural selection in physiology (e.g., thermoregulation) and morphology (e.g., body size) of species (Angilletta 2009). Therefore, it important to note that environmental are not phenotypic traits, but these aspects of the environment have selective effect underlying eco-physiological adaptations to survive in a given habitat (Searle 1961; Hancock et al. 2008; Blanchet et al. 2023; Lasky et al. 2023). Thus, variations in the environment must related to expressed phenotypic traits that then can be characterized by exploring their underlying genotypic diversity, which includes gene expression, sequence polymorphisms (SNPs), and changes in non-coding regulatory regions. By focusing on higher taxonomic ranks, we can characterize convergent patterns of adaptation to similar environmental components at distant geographic locations by finding phenotypic convergences across taxa living in similar habitats within vast continental scales. In this study, we explored this relationship by focusing on reptiles (lizards, snakes, and turtles) and the evolution of their adult hemoglobin subunits (Hbs) in association with environmental factors within a planetary context (Fig. 1, Fig. S1). Abiotic habitat conditions can shape the primordial function of hemoglobin as an oxygen transporter, which phylogenetic comparative analyses with environmental variables can reveal. Thus, we can test whether specific sequence changes in Hbs are linked to habitat

conditions of the taxa under study using their geographic coordinates at the collection site of the genetic samples to extract the habitat conditions from global environmental grids.

To study environment-phenotype-genotype associations, we propose the following criteria to select key genes to characterize such connection: 1) These loci should encode peptides with critical importance for organism survival with respect to their physiology responses to specific environmental conditions; 2) These proteins must be ubiquitously expressed across tissues and life stages to be easily retrievable from diverse genomic experiments; 3) Such proteins must also have a direct interaction with a known environmental component (e.g., oxygen availability), and such biotic-abiotic correspondence should vary based on the geographical context (e.g., altitude), resulting in identifiable phenotypes (e.g., hypoxia-adapted populations); and 4) These loci must show enough genetic variability (polymorphisms) to allow convergent phenotypes to be expressed from different alleles that emerge independently as a consequence of similar environmental pressures. Other characteristics of these loci include relatively few paralogs (e.g., 4-6 copies) and short sequence length of the mature peptide (e.g., 300-400 amino acids, ~1,200 bp of corresponding open reading frame sequences). Here, we explore this connection by studying the Hbs of adult reptiles retrieved from a planetary-level collection of genomic data.

Using the criteria above, we focused our study on adult Hbs of reptiles because of their central role in respiration, specifically in oxygen transport and ambient gas exchange. Sequences of these complete Hbs are easily reconstructed or isolated from diverse types of tissues and sequence datasets, including transcriptomes, annotated genomes, exome capture experiments, and PCR-based studies. Likewise, Hbs are relatively short peptides at ~142 residues in length, and among the most studied respiratory proteins found in animals that are directly linked to ambient oxygen levels (Hardison 1998; Storz 2018). In adult vertebrates, most Hbs are found in the blood, forming four globular subunits that bind up to four oxygen molecules simultaneously. These proteins provide enormous oxygen-binding capacity to the blood, which

is fundamental for aerobic metabolism and cellular respiration (de Villota et al. 1981; Storz et al. 2013).

For land vertebrates, aerobic respiration is maintained by a steady supply of environmental oxygen to respiratory organs (e.g., lungs or skin), where Hbs exchange gas. From there, oxygen is bound to Hbs and transported to almost every tissue in the body. Once oxygen is released, the steady supply of this gas maintains the oxidation of glucose and the synthesis of energy molecules (e.g., ATP), which power metabolic functions. Likewise, Hbs participate in other physiological pathways that include binding CO_2 and nitric oxide (Hsia 1998), providing powerful antioxidant capacities inside diverse tissue types (e.g., neurons), and as a general iron repository of the body, which makes them active in the homeostasis of this gas (Giulivi and Davies 1990). Given all these key physiological functions, Hbs genes are under strong selective pressure, but can evolve rapidly in response to environmental changes affecting oxygen levels, such as hypoxia or hyperoxia (Storz 2007; Storz and Moriyama 2008; Hindle 2020).

Different environments are associated with the adaptive evolution of Hbs because of the differences in oxygen availability due to altitudinal variation, temperature, and air density (Karpov and Novikov 1980; Andersen et al. 2009; Baalsrud et al. 2017). Altitude is especially important because of the hypoxia-induced extreme physiology at high elevations, where oxygen densities are lower than those at sea level and require efficient aerobic metabolism for daily life and exercise performance (Storz et al. 2010; Projecto-Garcia et al. 2013; Natarajan et al. 2018). However, for taxa that have thrived and diversified in such environments (e.g., mountaintops), we might observe unique adaptations in their Hbs for efficient respiratory homeostasis (Winslow 2007). For instance, species living in mountaintop locations tend to show multiple amino acid substitutions in their Hbs that experimentally have been shown to increase the oxygen affinity of their blood, as evidenced in deer mice (Storz et al. 2010), Andean waterfowl (Natarajan et al. 2015), and hummingbirds (Projecto-Garcia et al. 2013). Consequently, we could expect to find

multiple instances of amino acid substitutions in Hbs in other high-altitude taxa and many events of convergence as a telltale of parallel adaptations to such hypoxic environments (Storz et al. 2009; Scott et al. 2011; Storz et al. 2019). Most importantly, if a large collection of Hbs were available for a diverse clade with known distributions and across vast altitudinal ranges and geographic distances, we would most likely see evidence of unique substitutions associated with low-oxygen environments as potentially adaptive convergences and highlight what type of changes have occurred (e.g., polar to non-polar amino acid replacements) and where (e.g., near or at heme-binding regions). These substitutions can then be tested experimentally for their physiological significance, such as in *in vitro* or molecular modeling experiments, to determine whether they increase the oxygen affinity of Hbs as predicted.

We note that some Hbs paralogs are expressed only during larval/juvenile life stages. Still, we excluded these in our analyses for the following reasons: (1) Available molecular data on these Hbs variants are reduced unless the sequencing experiments targeted or included larvae or juveniles, which make these Hbs to be life-stage specific and rare among most genetic experiments; (2) Transcriptomic experiments of adult tissues rarely recover partial larval/juvenile Hbs as by-products, which limits phylogenetic comparisons across most species to include partial sequences of these loci; and (3) Homology determination of juvenile/larval Hbs is extremely difficult give the poor knowledge on such variants among reptiles (Storz 2018). Given all these caveats, we focused only on adult variants that could be entirely recovered from most sequence experiments or fully reconstructed from genomic/transcriptomic repositories.

In this study, we explored the evolution of adult Hbs in squamata reptiles (i.e., lizards and snakes) and terrestrial turtles globally. These genes are ubiquitously expressed and easily obtained from diverse molecular data, which allowed us to determine how environmental variables might have shaped their evolution. To our knowledge, no studies have explored the molecular evolution of Hbs across Reptilia within a planetary context. To maximize homology, our study focused on the two most diverse orders of reptiles: Squamata (lizards and snakes)

and Testudines (turtles and terrapins). We excluded other groups of reptiles (e.g., birds, crocodilians, and aquatic/fossorial species) because of their highly specialized aerobic physiology, difficulty in assigning homology to their Hbs, and noncomparable environmental data (e.g., marine species). To facilitate further comparisons, we restricted our study to four adult Hbs genes: HBBA, HBAD, HBB1, and HBB2. Full gene transcripts of these loci are easily reconstructed or isolated from most transcriptomic and genomic archival data (i.e., genome annotations and PCR-based studies). Using the metadata associated with the source samples, we geolocated and retrieved the following environmental variables: altitude, mean annual temperature, annual primary productivity (i.e. Aridity Index, Potential Evapotranspiration, Net Primary Productivity, Enhanced Vegetation Index), and latitude. We chose these variables because they characterize key aspects of the environmental envelope of each species and are essential predictors of reptilian biodiversity and oxygen homeostasis (Gaston 2000; Vitt and Caldwell 2013; Pough et al. 2016). For example, high temperature and high primary productivity could be associated with faster metabolic rates in reptiles, which in turn could increase mutation rates in Hbs genes (Allen et al. 2006; Gillman and Wright 2014). With Hbs sequences and environmental data, we tested for associations between the environment and whole-protein physicochemical properties and with per-site amino acid changes. For the latter analysis, we compared whether individual sites were under positive selection and whether they might have allosteric effects on nearby sites where heme groups bind, which are usually highly conserved (Storz 2018). All analyses were performed under a phylogenetic comparative framework to address evolutionary signal and bias. Lastly, our sequence dataset also allowed us to test whether gene duplication and conversion events occurred across the adult Hbs of land reptiles, which we also report here.

Results

Phylogenetic signal: We found that related species of reptiles tend to live in habitats with similar environmental variables as expected (Table S1). In other words, these animals

might have adapted and show similar eco-physiological traits to survive in habitat conditions reflected in absolute latitude, altitude, mean annual temperature, aridity index (AWI), and primary production measured by potential evapotranspiration (PET), net primary production (NPP), and Enhanced Vegetation Index (EVI) estimates. We note that these variables are environmental and not phenotypic traits, but these habitat conditions must relate to a set of underlying eco-physiological adaptations that each species (including reptiles) and populations must have to survive in a set environment (Searle 1961; Hancock et al. 2008; Blanchet et al. 2023; Lasky et al. 2023). We summarize the phylogenetic signal of each of these environmental variables as a proxy for the underlying eco-physiological traits using Pagel's lambda (λ), which has values from 0 (no signal) to 1 (strong phylogenetic signal). Our interpretation of no signal (i.e., $\lambda \sim 0$) is that eco-physiological adaptations for a particular environmental variable evolved independently between taxa along their phylogenetic tree, and these traits are not evolutionary correlated with the diversification patterns; while $\lambda \sim 1$ suggests that these eco-physiological adaptations follow the phylogenetic relationships among taxa such that similar values reflect shared phenotypic patterns among closely related species living in similar habitat conditions (Revell et al. 2008). Our results suggest that related species share similar eco-physiological traits for most environmental variables if they live in geographic areas with similar latitude, altitude, mean annual temperature, PET, NPP, and EVI. Our interpretation is supported by the strong phylogenetic signal evidenced by all λ estimates > 0 for all these environmental variables (i.e., λ equal to 0.8860, 0.4962, 0.8376, 0.5536, 0.6922, and 0.5500, respectively) and all these λs with significant p-values < 0.001 for N = 273-278 taxa. In contrast, the phylogenetic signal result associated with AWI (aridity) was not significant, and it did not follow the phylogeny of reptiles with a $\lambda = 0.0001$ and p-value = 1; N = 273 taxa. This result suggests that this aridity estimate did not support any underlying association with desiccation resistance adaptations in the reptiles of our dataset.

Hemoglobin subunits (Hbs) in land reptiles exhibit gene duplications and conversions

We reconstructed the phylogeny of each Hbs with the sequences of 278 species, which included 17 species of turtles (Testudinea), 134 species of snakes (Serpentes), and 127 species of lizards (Table S2). The phylogeny of all four Hbs shows that HBAD and HBA1 formed two monophyletic groups with corresponding 99% and 92% nonparametric bootstrap support. We found strong evidence that land reptiles have only one copy of HBA1 and HBAD without gene duplications or conversions. In contrast, the two HBBs showed several gene duplication events in reptiles, except for these snake subunits, where they were found to be sister clades (Fig. S2).

We used a modified host-parasite evolution approach using cophylogeny and parafit tests (Legendre et al. 2002) to compare the gene tree and species. These methods are based on the idea that individual genes (i.e., each Hbs) mostly should follow the species tree topology, and discrepancies between gene and species trees might reveal duplications or incorrect homology assignment of individual genes. For HBAD and HBA1, we retained a total of 264 and 256 sequences, respectively, and individual sequences for each Hbs correspond to a single reptile species. For HBAD, we found the maximum likelihood (ML) phylogeny had a significant cophylogenetic signal with the Squamate+Testudinea–ToL reference species tree (m^2_{XY} = 3442875, p-value < 0.001, n =1000). This result was also supported by a significant Parafit test (ParaFitGlobal = 153464026, p-value < 0.001 based on 999 permutations). For HBA1, we also found that this gene ML phylogeny had a significant cophylogenetic signal with Squamate+Testudinea–ToL tree (m2XY = 3364423, p-value < 0.001, n=1000) as well as a significant Parafit test (ParaFitGlobal = 219720223 and a p-value < 0.001 based on 999 permutations).

In contrast, the phylogenetic histories of HBB1 and HBB2 were complex. For turtles, we found that all their HBBs were nearly identical in this lineage, which made homology assignment or distinguishing between HBB1 and HBB2 very difficult, as shown in previous studies (Rücknagel and Braunitzer 1988; Damsgaard et al. 2013). For squamates excluding snakes, we

found that non-snake reptiles (127 species) presented repeated duplications and gene conversions for HBBs. Therefore, we tested for duplication directionality (i.e., HBB1 toward HBB2 or HBB2 toward HBB1) and identified three events that supported transformations from HBB2 toward HBB1 in *Ctenotus*, *Ophisaurus*, and *Varanus* lizards (Table 1). Similarly, we found that *Varanus* and Lacertidae lizards had at least 3 and 4 duplications at different points in their evolutionary history at 39.9 and 78.0 MYA, respectively, based on Squamate+Testudinea–ToL tree (Fig. S3). Consequently, to avoid any homology assignment errors for the HBBs in our study, we did not include these subunits of turtles and lizards, and only included those of snakes in any further coevolutionary and protein evolution analysis.

In the snakes, we found that both HBB loci diverged only once and did not undergo any further gene duplication or conversion events after the snakes split from their closest iguanian ancestor at ~160 MYA (Kumar et al. 2017). This was supported in both HBBs, which formed two well-supported sister clades with 100 nonparametric bootstrap support for each one (Fig. S3). In the case of HBB1, this dataset included a total of 117 sequences of snakes that resulted in an ML tree with overall agreement with the Serpentes lineage within the Squamate+Testudinea–ToL. Such that the cophylogeny results between the HBB1 and Serpentes species trees were significant (m^2_{XY} = 131990.7, p-value < 0.001, n =1000), as were their parafit test results, with ParaFitGlobal = 425852.7 and a p-value of 0.001, based on 999 permutations. For the HBB2 of snakes, this dataset included a total of 115 sequences that resulted in an ML tree with significant cophylogeny results with the Serpentes species tree (m^2_{XY} = 147171.3, p-value < 0.001, n =1000) as well as significant parafit test results with ParaFitGlobal = 302549.1 and a p-value of 0.001 based on 999 permutations. Therefore, we considered that the alignment of both HBBs of snakes could be used in all subsequent coevolutionary and protein evolution analyses.

Hemoglobin subunits evolved in response to the environment

We used coevolutionary (CoEvol) analyses (Lartillot and Poujol 2011) to test the correlation between substitution rates of Hbs with environmental variables (Table S3). Given

that we are addressing ~ 281 MYA of evolutionary history across Reptilia, we first determined whether all four Hbs had experienced genetic saturation, which included tests for HBA1 and HBAD (lizards + snakes + turtles) and HBB1 and HBB2 (only for snakes). After inspection of the saturation plots (Figs. S6–S9), we considered that not all Hbs might be saturated, as evidenced by a near-linear correspondence between genetic distances for each Hbs. We note that these genetic distance plots are qualitative in nature, and they cannot rule out underlying saturation. For this reason, we also performed additional CoEvol analyses using the *Kr/Kc* model (Nabholz et al. 2013) to adjust for potential genetic saturation. The results were similar to those found without considering saturation (Table S4). Therefore, we inferred that any subsequent analyses using the Hbs datasets would not be affected by saturation and would not underestimate the substitution rate if we performed phylogenetic correlation analyses (Tamura and Nei 1993).

Our CoEvol analyses revealed a diversity of pairwise associations between evolutionary rates (i.e., ω = *dN/dS* and *dS*; where *dN* is the non-synonymous and *dS* is the synonymous substitution rates) with environmental variables (i.e., absolute latitude, altitude, primary production as EVI, and mean annual temperature). Given the large number of comparisons, we adjusted our interpretation of these relationships by excluding confounding effects with phylogenetic partial correlations and mapping those in an association network. Partial correlations measure the relationships between variables while controlling for the influence of additional or confounding variables in a multivariate process (Wong et al. 2003; Lartillot and Poujol 2011). With this approach, we isolated associations between molecular rates with environmental variables while determining their directionality (i.e., direct or inverse) and removing the effects of other confounding variables (e.g., ambient temperature relationship with *dS* as controlled by the influence of altitude). By mapping those relationships in a network, we could interpret each pairwise association as an individual link between environmental and molecular variables and code them by their statistical significance, directionality, and strength. We estimated one network for each Hbs gene and then compared them to find common

patterns across Hbs (e.g., altitude association ω by adjusting for ambient temperature). We interpret the findings for each of these phylogenetic partial correlation networks below, which are also summarized in Tables 1 and S3 and Figure 2.

For HBA1, we found that altitude was negatively correlated with ω (posterior probability or PP < 0.001), which suggests that diversification of this Hbs might increase at lower altitudes in land reptiles (i.e., turtles, lizards and snakes) where oxygen density is the highest and that the affinity of this gas for respiratory proteins like Hbs might be less stringent for survival. We did not observe any other significant correlation between the environmental and molecular evolution variables for HBA1. These coevolutionary results were the same even when we excluded turtles from our analyses (Table S5). Therefore, our results suggest that temperature, latitude, and biomass production are not correlated with the diversification of HBA1 in land reptiles.

For HBAD, altitude was also negatively correlated with ω (PP < 0.001), suggesting a similar pattern as seen in HBA1, such as that oxygen density might exert a significant constraint on the diversification of both subunits. We also found that temperature was negatively correlated with ω but positively correlated with *dS* for HBAD, with similar results after excluding turtles. These findings indicate that land reptiles inhabiting places with warmer temperatures might undergo less molecular diversification of HBAD. We also found a positive correlation between *dS* and primary production (EVI). However, this relationship was not significant after we excluded turtles. Like the results of HBA1, we found that latitude was not correlated with any of the rates of molecular evolution for HBAD. These results suggest that latitudinal gradients of diversification might not apply to both HBA1 and HBAD.

Our analyses of HBB1 and HBB2 were restricted to snakes (Serpentes) and revealed some specific trends for these reptiles. First, HBB1 showed no significant correlations between the environmental variables and the substitution rates. Therefore, no discernible trends are evident between the environmental variables and the evolution of HBB1, and this subunit is likely the most conserved of all adult Hbs in snakes, and perhaps reptiles. We found that EVI

was negatively correlated with *dS* of HBB2, which suggests that more productive regions, such as tropical rainforests, might be associated with fewer SNPs for this Hbs. This result is the opposite of what we found for HBAD (i.e., a positive correlation) and suggests that HBB2 is conserved in snakes. Finally, HBB2 was the only subunit with a significant negative association with absolute latitude. This relationship was with *dS*, which suggests that HBB2 in snakes tends to accumulate more SNPs as species live and diversify closer to the equator. However, no significant relationship was found between latitude ω and HBB2, which might have supported an increased diversification of encoded amino acids beyond synonymous nucleotide changes as predicted for the latitudinal gradient hypothesis.

Whole-protein properties of hemoglobins are weakly predicted by environmental variables

We performed pairwise phylogenetic regressions between physicochemical variables (Supplementary Results) for each Hbs and the environmental variables at the collection site of the corresponding reptile species. The physicochemical properties were derived from the complete amino acid sequences (i.e., whole protein) and included 17 parameters that summarized the molecular weight, charge, composition, thermostability, hydrophobicity, and isoelectric point. The environmental characteristics included absolute latitude, altitude, primary production, and mean annual temperature. Before the regression analyses, we paired the physicochemical properties as dependent variables with the corresponding environmental variables, and some were log-transformed to improve their distribution, like altitude (see Supplementary Methods). Some relationships were marginally significant (i.e., p-values < 0.05, see Table S6), but they all were regarded as non-significant after p-value recalculation using a Bonferroni criterion. Therefore, no physiological variable is related to any environmental characteristics for all hemoglobin subunits (p-values > 0.05 after Bonferroni correction, see Table S6). Regarding phylogenetic signal, related species tend to share similar physicochemical

characteristics for all subunits, except molecular weight (p-value > 0.05) in HBAD; proportions of aromatic and tiny amino acids, or for the molecular weight for HBB1 (Table S7). We conclude that whole-protein properties of Hbs show similar values based on species relationship as expected for a significant phylogenetic signal, but they show only marginal associations with environmental variables in land reptiles.

Per-site sequence analysis reveals sites under positive selection in hemoglobin subunits

We tested whether specific codon sites in the Hbs are under selection as indicated by the magnitude of the $\omega = dN/dS$ and dS metrics under different models estimated with HyPhy (FEL, FUBAR, SLAC) and CodeML (M7 and M8). As expected, most sites were under purifying selection, as detailed in Table S8. However, some codons were found to be under diversifying selection and illustrated in Figure 3 and 4. The total number of codons under diversifying selection were 3–8 sites in HBA1 (3 for FEL and FUBAR;8 sites for SLAC, and 1 site CodeML), 1–10 sites in HBAD (1 for FEL; 2 for FUBAR; 10 for SLAC, and 3 sites for CodeML), 5–6 sites in HBB1 (6 for FEL; 5 for FUBAR; 6 for SLAC, and 6 sites for CodeML), and 1–5 sites in HBB2 (3 for FEL; 3 for FUBAR; 5 for SLAC and 1 for CodeML). The results HBBs are for snakes, while HBA and HBAD1 are for turtles, snakes, and lizards.

We summarize the significant results of these per-site analyses (Table S8) to those in agreement among selection models and most likely impacting the physiological function of Hbs. To help contextualize these results, we use the amino acid numbering from the two available crystallographic structures for all reptiles derived from two turtles (PDB 1V75 and 3AT5; see supplementary methods). In the case of HBA1 (Fig. 3), we observed that the 137th site is undergoing positive selection, its consensus state being nonpolar, and this site is next to the conserved heme-binding position at 136th (Hasegawa et al. 2011). We observed a total of 69 changes in the 137th site from the consensus nonpolar to a polar state, 2 changes to an aromatic state, and 1 change to a basic state. In the case of HBAD, we observed that the 121st site is also undergoing positive selection, its consensus state being nonpolar, and this site is

only 15 sites away from a key conserved heme-binding site at position 136th (Hasegawa et al. 2011). For the 121st site, we observed a total of 74 changes from the consensus nonpolar to a polar state, and 3 changes to a basic state. For HBB1 in snakes (Fig. 4), the 43rd site is under positive selection, its consensus state being polar, and adjacent to the conserved heme-binding site at position 42nd (Hasegawa et al. 2011). For site 43rd, we observed a total 35 changes from the consensus polar to a nonpolar state, and 3 changes to an acidic state. Lastly, for HBB2 in snakes, the 140th site is also found under positive selection, with its consensus state being nonpolar, and also adjacent to the heme-binding site at position 141st (Hasegawa et al. 2011). For the 140th site, we observed a total of 31 changes from the consensus nonpolar to a polar state. Overall, the sites highlighted above are undergoing positive selection with a clear substitution trend in changing residue polarity state from consensus nonpolar to polar states. However, the allosteric implications of such amino acid variability on adjacent heme-binding sites and their impact on oxygen affinity by Hbs given environmental conditions (e.g., altitude on oxygen density) are unknown; they need to be further tested experimentally (Hasegawa et al. 2011).

DISCUSSION

Hemoglobins comprise key proteins in the respiration physiology of vertebrates (Storz 2016). Their origin and functional differentiation into diverse subunits involved several gene duplications and conversion events during the early radiation of land vertebrates as their ancestors moved from water to land (Storz et al. 2013; Hoffman et al. 2018). For reptiles, we found that HBA1 and HBAD were monophyletic and sister clades (Fig. S3), indicating that only one gene duplication gave rise to both loci in these animals (Storz 2016). In contrast, HBB genes had multiple duplications and gene conversion events across lizards (excluding snakes and turtles). Interestingly, a previous study reported similar HBB conversions for *Anolis* lizards (Hoffmann et al. 2018). We compared our results with this study's, but we failed to find any support for such reported conversion events (i.e., p-values > 0.05 for all 21 *Anolis* species;

Table S9). If *Anolis* has experienced any HBB conversion, such results might require more species than our study's (Patel et al. 2010). Although we could not validate these previous results for *Anolis*, we found strong evidence for duplication events in other lizards, including *Ctenotus*, *Ophisaurus*, *Varanus*, and lacertids. Consequently, we consider that many HBBs submitted to NCBI as HBB1 or HBB2 might need further validation using phylogenetic analyses before submission, such as those from automatic annotation of genomes (e.g., RefSeq; Gnomon), transcriptomics, exome capture experiments, and sequences resulting from PCR-based amplifications.

Our results revealed the conversion of HBBs in at least two separate instances from HBB2 to HBB1 in *Varanus* and between *Ophisaurus gracilis* and *Ctenotus robustus* at approximately 8.8 and 174 MYA, respectively. For *Varanus* taxa, we hypothesize that their HBB gene conversion was due to recent divergence within the same genus, as supported by their close phylogenetic relationship. In contrast, the results between *Ophisaurus gracilis* and *Ctenotus robustus* are more difficult to explain because both shared a last common ancestor at ~170 MYA. One possibility is that an old conversion event has been maintained without modification in either genus. However, *Ophisaurus* and *Ctenotus* live in different continental areas (North America and Oceania, respectively) and inhabit different altitude ranges (8 - 25 and 3 - 28 MASL, respectively, in our dataset); thus, this gene conversion event requires further study. Alternatively, the species size of reptiles might be smaller than required to recover *Ophisaurus* and *Ctenotus,* more closely related lineages involved in their gene conversion events between HBBs.

For snakes, HBB1 and HBB2 formed two sister monophyletic groups, and these results suggest that both loci diverged only once, just before or during the initial radiation of snakes. Furthermore, we did not observe any gene conversion events afterwards between the HBBs of snakes (Table S9). For this reason, we considered that any further analyses that included HBBs with protein-environmental coevolution tests would be focused only on snakes. This approach

served to avoid the inclusion of any bias or incorrect HBB subunit assignment that might have occurred in the other groups of land reptiles, such as lizards or turtles.

Related reptile lineages present a strong phylogenetic signal for almost all the environmental variables (Table S1). Still, these should not be considered phenotypic traits, but as a proxy for the underlying eco-physiological adaptations each species must have to survive in a given environment. These results suggested that phylogenetically related species tend to live and have similar adaptations (i.e., present homologous trait states) in similar environmental conditions. In contrast, we found that species do not present a phylogenetic signal for some physicochemical variables (Table S1), such as the molecular weights of HBA1 and HBB1. These results could be explained by these subunits having little variation in their total weight, even if they have different amino acid compositions across species. We also did not observe statistically significant associations between some environmental variables and physicochemical properties for Hbs (Table S6, S7), suggesting that changes in environmental variables do exist; still, they tend to have a small influence (or less predictive power) on Hbs (See supplementary materials and results). An explanation of these weak associations is that we are looking at summary physicochemical characteristics derived from the whole Hbs proteins, not sections of each that might be diversifying. For example, we found many substitutions between amino acids with similar characteristics (i.e., polar-to-polar or nonpolar-to-nonpolar); as a group all of these substitutions they can collectively cancel each other over the whole Hbs by keeping the overall polarity the same.

To our knowledge, the relationship between latitude and the molecular evolution of Hbs has not been studied across reptiles. Previous studies in other groups have reported higher rates of molecular evolution at low latitudes based on data from endotherms (mammals and birds) and aquatic ectotherms such as fish (Gillman et al. 2009; Wright et al. 2011; Gillman et al. 2012; Lourenço et al. 2013). However, limited evidence exists for the association between latitude or other variables (e.g., temperature) with rates of evolution in ectotherms such as land

reptiles and glass frogs (Rolland et al. 2016; Dugo-Cota et al. 2015). In the first study, the authors tested the associations of latitude and temperature on the molecular evolution of reptiles using a large tree of 1,651 squamates (Rolland et al. 2016). This phylogeny was estimated using mostly mitochondrial genes with a few nuclear markers (Pyron et al. 2013) , and the branch lengths were used as a proxy for the number of substitutions per site. This study found no associations between temperature or latitude with evolutionary rates in squamate reptiles. In the second study, Dugo-Cota et al. (2015) also failed to establish a significant correlation between latitude and substitution rates in glass frogs after analyzing three mitochondrial and seven single-copy nuclear gene fragments. However, these authors reported a significant relationship between substitution rates and ambient temperature derived from the collection sites of the frog species in this study. Both studies agreed on the limited effect of latitude on the substitution rates in land ectotherms (i.e., squamates and glass frogs), but they disagreed on the effect of temperature on these rates.

Our results also reject latitudinal gradients for the molecular evolution of all Hbs except for HBB2 in snakes. The fact that absolute latitude is not related to ω in all analyses implies that factors beyond that geographic coordinate or latitudinal diversity gradient, such as metabolic rates or environmental variables such as biomass productivity and altitude, may exert a more substantial influence on Hbs evolution. For example, lower latitudes might include complex geographic features such as tropical rainforests with high productivity areas. Conversely, areas near the equator might also include high mountain ranges that present lower productivity and hypoxic environments, affecting ω. Moreover, many lizards and snakes live in extremely warm and low-productivity regions (e.g., the Sahara, Namib, and Sonoran Deserts) at mid-latitudes. These environmental conditions can impact hemoglobin physiology, which might favor adaptation to increase the oxygen affinity of Hbs under extreme ambient heat (Gangloff and Telemeco 2018).

In our partial correlation analyses, we found evidence of a relationship between environmental temperature and the evolution of some Hbs after controlling for the effect of the other confounding variables, like altitude. For HBAD, our results show an inverse relationship with temperature, such that higher ambient temperatures result in lower substitution rates that result in amino acid changes. More efficient DNA repair and reactive oxygen species (ROS) neutralizing mechanisms are expected to evolve in reptiles exposed to warmer climates. For example, the lizard *Ctenophorus pictus* shows modest but significant telomere repair during the mating season, which also occurs in a warm climate due to increased ROS production (Olsson et al. 2018). However, it is not clear whether such DNA repair is a direct effect of temperature on ROS production for gamete production (Fitzpatrick et al. 2019). In contrast, greater synonymous changes, as evidenced by the positive relationship between *dS* and higher temperature for HBAD, agree with the results reported for glass frogs (Dugo-Cota et al. 2015). The impact of temperature on molecular evolution has been attributed to the metabolic theory of ecology (Gillooly et al. 2005; Allen et al. 2006). For instance, warmer environments might increase rates of molecular evolution because faster cellular metabolism might increase mutation rates across the genome. Consequently, more ROS are produced, increasing oxidative stress and greater DNA damage per unit of time (Santos 2012). For HBAD, if the substitutions are synonymous, warmer temperatures might mean more variation in SNPs in this gene. In contrast, HBA1 and HBB1 variables were not correlated regarding temperature or their substitution rate. For these Hbs, our results seem to agree with the reported lack of correlation between the temperature and branch length of the squamate tree study (Rolland et al. 2016).

For HBAD, only the *dS* showed a significant positive relationship with respect to the EVI. This association could be explained by increased substitution rates for Hbs in reptile species that inhabit areas with high biomass production, such as tropical rainforests. However, a direct relationship between faster rates and biomass production is difficult to explain other than as a byproduct of increases in ROS production during aerobic respiration associated with biomass

production (Gillman and Wright 2013). We also found an inverse relationship between the HBB2 dataset of snakes and the *dS* at relatively high EVI values. We propose that snakes might have more efficient DNA repair or better ROS metabolism in more productive and warmer equatorial environments, such that the folding and stability of their Hbs could be less affected by increasing synonymous mutations (Jena 2012). For HBA1 and HBB1, we did not find any relationship with the EVI.

Altitude was found to have a negative relationship with ω HBA1 and HBAD across reptiles. These results might present constrained sequence evolution (i.e., lower values of ω) at high altitudes. Previous research has shown that physiological constraints might restrict hemoglobin variability to maintain optimal oxygen metabolism (Weber 2007; Storz and Moriyama 2008). Hbs are expected to be under stronger negative selection for those species living at high altitudes because most mutations affect sites relevant to the oxygen affinity of the heme group, and amino acid changes might significantly decrease adaptability to hypoxic conditions. In other words, mutations or changes that reduce oxygen affinity at high altitudes might cause constraints in activities dependent on aerobic capacity and performance (e.g., capturing prey and escaping predators). Evidence of such a dependence on the Hbs sequence and aerobic performance has been reported in other high-altitude vertebrates. For example, a few amino acid replacements seem to be associated with energetically demanding activities that require efficient aerobic performance, such as flight, in high-altitude hummingbirds (Trochilidae), including members of several groups, including Lesbiini (*Adelomyia*, *Oreotrochilus*); Heliantheini (*Aglaeactis*, *Coeligena*); *Patagona*, Trochilini (*Amazilia*), and Phaethornithinae (*Phaethornis*) (Projecto-Garcia et al. 2013). Many of these birds present convergent amino acid replacements, which are restricted to specific sites of the alpha and beta chains of Hbs that seem to increase the affinity for oxygen (Storz et al. 2009; Storz et al. 2010).

At low altitudes, reptiles present more nonsynonymous substitutions for their Hbs, including evidence of positive selection at some sites. Such high rates result in amino acid

variability and more SNPs since such species are already evolving in oxygen-rich environments. Therefore, selection in those taxa living at low altitudes might be less stringent in causing physiological stress, where their Hbs might have more variants, and some might be less physiologically efficient in binding oxygen. It is also possible that other factors might be influencing such higher sequence variability of Hbs at lower altitudes, such as aerobic performance. For instance, some reptiles might be more active as they look for prey (Hillman et al. 2013).

HBB1 and HBB2 in snakes did not show such a tendency for higher ω values regardless of altitude. We consider that HBB subunits in snakes could bear a lesser physiological burden regarding oxygen affinity than HBA1 and HBAD do (Storz 2018). Alternatively, it is possible that HBBs are more conserved, as is HBB1, and less flexible to vary in response to environmental changes. In any case, such changes in selective pressure in the HBBs of snakes imposed by low oxygen densities at higher altitudes are not reflected in their ω parameter. We emphasize that we were unable to explore such relationships with other reptiles. Additionally, most reptiles could also have other physiological adaptations for efficient oxygen metabolism that could compensate for less efficient HBBs. These compensatory mechanisms might include increased erythrocyte numbers, higher hemoglobin concentrations, and dense hematocrit, which have been observed in some lizards, such as *Phrynocephalus erythrurus* (Lu et al. 2015). However, different results have been reported in *Sceloporus grammicus* lizards, which support a nonlinear relationship between erythrocyte count and hematocrit with changes in altitude, such that the oxygen binding capacity of these lizards is more complex than the mere number of red blood cells (González-Morales et al. 2017).

We found the relationships between temperature and altitude with evolutionary rates of Hbs covary in reptiles. These results might be a consequence of our use of precision matrices during analyses [i.e., the network of relationships between variables is estimated via partial correlations between them (Wong et al. 2003)]; thus, the effect of altitude on temperature was

already accounted for when their covariation was estimated. Alternatively, the relationship between altitude and temperature might be close enough to confound their individual contributions to the associations with ω and *dS*. Similarly, the tight association between altitude and temperature is explained by adiabatic cooling, such that the process of reducing heat (i.e., temperature) through a change in air pressure caused by air volume expansion with increasing altitude. Additional biological factors, such as behavioral responses to overheating or thermoconformer versus heliotherm habits, can also influence the effects of altitude on the evolutionary rates of Hbs. For example, reptiles living at high altitudes could develop strategies to address temperature, such as basking and seeking refuge, in response to thermal variation. Such behaviors in response to environmental temperature, as in most ectotherms, can influence metabolic rates in reptiles and be reflected in the observed substitution rates of their Hbs (Navas 2002; Clarke and Pörtner 2010).

Our analyses of the per-individual amino acid site evolution (i.e., HyPhy and CodeML results) revealed that most residues that interact in the heme-binding regions are under strong purifying selection (Fig. 3). This result is expected, as most amino acids in the pocket of each Hbs are already efficient in their heme/oxygen-affinity interactions (Storz 2018). This is exemplified by HBAD, which did not present sites under positive selection next to or at any of its heme-binding pocket regions. In contrast, we found that 137th in HBA1, next to a binding site, is under positive selection. This position could suggest that substitutions in this site can influence allosterically the adjacent heme binding in the reptile crystallographic models (PDB: 1V75 and 3AT5 for turtles) and the oxygen affinity of this subunit. In HBA1, we found that at site 137^{th,} the consensus state is a valine (a nonpolar amino acid); still, many species have a substitution for another nonpolar amino acid (alanine) in the same location. For species that have adapted to high altitudes, such as those living in areas with low oxygen levels, we expect that they present changes that increase the polarity of the heme pocket, such as those seen at site 137th. Such adaptations would be in the form of residue changes that could contribute to the stability of the

ligand-affinity threshold for oxygen (Storz 2018). However, we found that most species that experienced substitutions from nonpolar to a polar amino acid at site 137th live in lowland environments and might already be exposed to high oxygen levels. For example, 13 of 14 turtle species living between 0 and 622 meters above sea level (MASL) presented a change between the consensus amino acid (valine) and a polar amino acid (threonine). Similarly, 53 species of lizards that lived between 2 and ~2649 MASL presented the same change to a polar amino acid. Only one snake species with a low altitudinal distribution between 3 and 207 MASL presented a change to a polar amino acid (i.e., *Erythrolamprus poecilogyrus*).

Our results suggest that other snakes that live less than 1000 MASL also presented substitutions near a conserved binding site in other Hbs, but not HBA1. For HBBs in snakes, we found that the site 140th in HBB2 (near the binding site) is under positive selection, and lowland distributed snakes presented changes from nonpolar to polar amino acids. For HBB1, we found a similar site near a heme binding location at position 43rd. The consensus state for this site is a polar amino acid, and most snakes with nonpolar substitutions tend to reach an altitude of 2797 MASL. Such changes can increase the hydrophobicity and polarity of the whole heme pocket and increase the oxygen affinity even if the site near the binding site is nonpolar (Storz 2018). However, it has been suggested that mutations that change polarity could increase oxygen affinity but simultaneously could have pleotropic and deleterious implications and reduce structural stability with potential increases in spontaneous heme oxidation (Bonaventura et al. 2013; Varnado et al. 2013; Tufts et al. 2015) . We predict that all adjacent sites we found to be under positive selection might also modulate the oxygen affinity of HBBs of snakes. Still, their pleotropic effect has yet to be determined experimentally (Maillett et al. 2008; Projecto-Garcia et al. 2013; Varnado et al. 2013).

Amino acid substitutions that might affect any distal histidine in the heme pocket could also increase oxygen binding but affect the stability of the iron–oxygen binding of heme (Storz 2018). In HBA1, we found that at least 10 species (2 snakes, 1 turtle and 7 lizards) presented a

glutamine residue instead of such a conserved histidine. Similarly, 130 species (122 snakes, 18 lizards) also presented a glutamine at this position in HBAD. However, the presence of such a glutamine change has been demonstrated to have a minimal functional impact on the heme pocket in terms of its oxygen affinity capacity (Birukou et al. 2010). However, we found that a single species (*Varanus niloticus*) has an arginine replacing the histidine in HBAD. This substitution could increase CO_2 binding in the heme pocket, as has been reported in a human blood disorder in HBB subunits (Wallace et al. 1976). However, such substitutions might be adaptive if they impact their diving capacity (Branch 1998), as *Varanus niloticus* is a species that regularly dives for 15–60 minutes to find food (Wood and Johansen 1974). However, most other reptilian species with similar diving capacities do not present arginine changes in the HBAD. Therefore, physiological experiments using the blood of *Varanus niloticus* are necessary to determine the consequences of arginine changes in the heme pocket of the HBAD of this species.

Our study reveals how environmental variables such as altitude, temperature, and primary productivity, but not latitude, influence the molecular evolution of hemoglobin in reptiles. Additionally, our findings highlight the role of environmental pressures and the selection process in amino acid changes near heme-binding sites, shaping respiratory hemoglobin function. Thus, our findings offer broader insights into the convergent evolution of physiological traits across globally distributed taxa.

MATERIALS AND METHODS

Species selection for harvesting of hemoglobin sequences: We searched and downloaded RNA-seq data after searching for squamata reptiles and terrestrial turtles taxa from the SRA database in NCBI (until January 2022; See Table S2). However, we excluded some species from our analyses that might fall within this 'land reptile' category, which included crocodilians, endothermic taxa (birds), and aquatic/fossorial reptiles. Our rationale to remove such taxonomic groups is as follows: (1) The hemoglobin subunits (Hbs) in crocodilians are difficult to assign to

a correct homology with other Hbs found in other land reptiles as expected for its closer relationship to birds (Weber et al. 2013); (2) endothermy in birds makes their underlying physiology nonhomologous to ectothermic reptiles and less dependent on environmental conditions such as ambient temperature; (3) fossorial reptiles have an extreme aerobic physiology due to hypoxia adaptations to life underground that are less dependent on aboveground environmental variables that we used in our comparative analyses including altitude [i.e., air oxygen density], plant productivity, and ambient temperature (Kooyman 2012; Grigg 2015; Scanes 2015); and (4) marine and many semiaquatic reptiles (e.g., marine turtles) might have a near global distribution that could not be geolocated over land raster grids to retrieve corresponding environmental variables. After filtering out taxa based on the above criteria, our final selection comprised >270 reptile species. These species were utilized to reconstruct hemoglobins (Hbs) using transcriptomes, genomes, or PCR-based sequences available in GenBank for subsequent evolutionary analyses.

Confirming hemoglobin homology and identity: We reconstructed Hbs via the *in silico* target capture software MITObim, which uses the MIRA aligner (Chevreux et al. 2004; Hahn et al. 2013). We summarize this methodology as follows. To maximize our chance of reconstructing a correct hemoglobin gene, we used different sequences of HBA1, HBAD, HBB1, and HBB2 derived from the NCBI 'nuccore' collection to find the closest seed sequences to each target SRA experiment (e.g., same family or genus if possible). Additionally, we retrieved the full coding sequence (CDS) with Hbs for each reptile species available in NCBI via BLASTn with an e-value of $< 10^{-50}$ (Table S2). For the MITObim procedure, we used the recovered Hbs as bait/seed sequences and the raw demultiplexed SRA as target data. We included the following MITObim parameters: "−quick" (starts process with initial baiting using a fasta reference), "−mismatch" (12–17% as a range of allowed nucleotide mismatches), and "−kbait" (a set of 10–15 kmer for baiting stringency). This approach allowed MITObim to harvest and reconstruct Hbs using the gene bait/seed as a reference. Once all potential Hbs were reconstructed, some

species had multiple alleles for a given subunit, and we selected only one of those as representative of that species. Our criteria to select this allele were as follows: (1) the most common allele was selected if multiple SRAs were available; (2) the allele selection should have the least amount of nonsynonymous changes and ambiguities if compared with others if they were at the same frequency; and (3) alleles that have long insertions were excluded if closely related species lack such features, and we assumed that those insertions were likely erroneous reconstructions. With the final sequence matrices, we performed global alignments assuming codons (i.e., translation guided) for each gene using the R-package *DECIPHER* v3.2 (Wright 2016). Then, we inspected this alignment by eye and removed sequences that had long gaps compared to congeneric sequences. With the final alignments, we obtained base phylogenetic reconstructions for each Hbs based on their best molecular model of evolution using IQ-TREE 2 (Minh et al. 2020). The IQ-TREE 2 parameters for the initial search include the input model parameter "MFP+MERGE" with a partition file indicating codon positions. The support of each node of this tree was also calculated using IQ-TREE 2 with the following parameters for ultrafast bootstrap "UFBoot" search with 1000 replicates and optimized with "−bnni" parameter and "alrt = 0".

For phylogenetic comparative analyses, we use the largest multigene Squamata-Tree of Life (Zheng and Wiens 2016). We augmented this chronogram to include turtles (i.e., Squamate+Testudinea–ToL) by harmonizing the age node of Squamata-Tree of Life with a chronogram of the turtles derived from a time-calibrated Tree of Life (Kumar et al. 2017). This required selecting a subtree of testudines in the Squamata-Tree of Life (Zheng and Wiens 2016), and it was replaced with the chronogram turtles from the Tree of Life, and the node age of Testudinea was adjusted to 203.0478 MYA. This approach preserved the topology of Squamate+Testudinea–ToL.

After comparing the Squamate+Testudinea–ToL topology with the placement of our Hbs phylogenies, we detected several mislabeled samples with incorrect species information,

chimeric reconstructions, and putative gene duplication events (e.g., due to gene conversion). We removed contaminants (i.e., mislabeled sequences) and labeled those Hbs that might be gene duplications. We repeated the above process 5–6 times iteratively to allow the progressive removal of incorrect Hbs reconstructions until we converged to Hbs matrices congruent with the Squamate+Testudinea–ToL topology reference.

We further validated our sequence identity using a derivation of cophylogenetic analyses, where we considered that each Hbs tree topology should mostly mirror the phylogenetic positions of its source taxon in the Squamate+Testudinea–ToL reference. For this purpose, we tested for cophylogenetic signal between the time-calibrated phylogeny and each of the Hbs phylogeny in our study using the R package paco using the "PACo" function (Balbuena et al. 2013) under 1000 permutations; and the 'parafit' test in the R package *ape* (Paradis et al. 2004) with 999 permutations. The parafit test was introduced (Legendre et al. 2002) to study the host-parasite evolution comparing their evolutionary relationships. Here, we used this approach to compare two phylogenetic trees: a gene tree for each hemoglobin and a species tree to check that the gene tree has similar topologies. If the cophylogeny tests were significant, they would support that the taxon placement corresponds to a similar/equal source taxon position in the reptile reference phylogeny. We further validated each Hbs sequence by comparing each amino acid sequence with those from its congenerics. We expected that they should be similar to its congenerics (see Fig. S4 for a schematic). If any Hbs sequence that did not mirror the phylogenetic placement of its source taxon and showed an extremely misplaced phylogenetic position relative to congenerics, we considered this Hbs reconstruction of this taxon as erroneous, and it was dropped from the alignment before further analyses. Lastly, it is possible that true Hbs variants might have been excluded in our stringent filtering process. Yet, we emphasize that our goal is to maximize Hbs homology of squamates and turtles within the broad Reptilia class level.

***Connecting taxon metadata with environmental variables for hemoglobin evolution analyses*:** For the 278 reptiles species with transcriptomic data, we used their SRX/SRR number reported in NCBI-SRA repository to search for the geographic localization of each tissue sample with its associated metadata (SRA-BioSample). When the SRX/SRR numbers lack geographic coordinates or georeferenceable information (e.g., country, province/state, locality, etc.), we used voucher-validated species localities and/or records in the GBIF with biological collections (VertNet: www.vertnet.org) that have that include geographic coordinates or georeferenceable metadata (e.g., country, state/province, locality).

Once we obtained geographic coordinates, we derived the following environmental values from each location where a study species was found: (1) mean annual temperature calculated as the daily mean air temperatures during 1 year in °C at 1 km of resolution and extracted from the 'Climatologies at high resolution for the earth's land surface areas' (CHELSA) dataset ((Karger et al. 2017); https://chelsa-climate.org/bioclim/); (2) altitude as the elevation measured in meters over sea level at a 90 m of resolution as it was summarized in the SRTM 90 m Digital Elevation Data for the entire world provided by the CGIAR-CSI GeoPortal at https://cmr.earthdata.nasa.gov/search/concepts/C1214622194-SCIOPS. Additionally, we collected all georeferenced occurrences from the Global Biodiversity Information Facility (GBIF). For each point, we extracted its altitude information and plotted it to obtain an altitudinal range. With this range, we visually inspected whether our samples are within the altitudinal range of each species. ~95% of the samples in our dataset are within the altitudinal range (Fig. S5). (3) the absolute value of latitude from the corresponding geographic coordinates; (4) potential evapotranspiration (PET) as a proxy of the water transference between the land and the atmosphere for the entire planet at the given geographic coordinate (Zomer et al. 2022); (5) aridity index or AWI as a measurement of dryness in the climate at a specific geographic location, considering temperature and precipitation (Zomer et al. 2022); (6) net primary production (NPP) at the given geographic location from a NPP global

raster (https://neo.gsfc.nasa.gov); and (7) Enhanced Vegetation Index (EVI) as an improved measurement of primary productivity at a given geographic location. For all comparative analyses based on primary production, we chose the EVI measurement instead of other alternatives including the normalized difference vegetation index (NDVI), PET, NPP, and AWI. Our rationale is based on the consideration that EVI makes more corrections to some atmospheric conditions and canopy backgrounds, and the EVI is more sensible in areas with dense vegetation, such as those locations in the tropics and temperate forests where many reptiles reside (Matsushita et al. 2007). The EVI measurements are also unitless with a 1 km resolution and extracted from the MOD13A3 - MODIS/Terra Vegetation Indices Monthly L3 Global 1 km SIN Grid (https://ladsweb.modaps.eosdis.nasa.gov/missions-and-measurements/products/MOD13A3/). The EVI data are measured monthly, so we averaged these values annually to obtain annual mean values for EVI from the records corresponding to 2000 – 2021. To obtain values for each geographic coordinates, all global environmental grids were preloaded into R and location estimates were extracted using the function extract with each geographic coordinates of the study species as input into the R-package raster (Hijmans et al. 2015). Finally, we average the values of all the coordinates to obtain a species value for each environmental variable.

Phylogenetic signal and regressions of hemoglobin subunits and the environment: We calculated the phylogenetic signal of all environmental variables to identify whether related species of reptiles presented similar trait values related to their adaptation to such external conditions due to shared ancestry (Pagel 1997; Pagel 1999). We assessed four different signal statistics: Pagel's λ (Pagel 1997), Blomberg's K (Blomberg et al. 2003), Abouheif's C_{mean} (Abouheif 1999), and Moran's I (Gittleman and Kot 1990). For phylogenetic regressions, we evaluated each environmental variable under three models of phylogenetic covariance. These models include Brownian motion (BM), the Ornstein–Uhlenbeck (OU) model and Pagel's λ (Pagel 1997; Pagel 1999; Ho and Ané 2013). We recorded the Akaike information criterion

corrected (AICc) score from each data point and used it to select the best model. Models with the lowest AICc values were chosen as the best supported based on pairwise comparisons such that ΔAICc values ≤ 2 (Burnham and Anderson 2001). The significance of the phylogenetic regression coefficients was evaluated with a threshold of alpha < 0.05. We adjusted P-values for Bonferroni correction using the function 'p.adjust' in R. We also calculated the phylogenetic signal with these methods for whole-protein physicochemical properties of each of the Hbs, which are presented in the Supplementary Results (Table S1).

Gene duplication and gene conversion tests of hemoglobins: We searched for evidence of duplication events in our Hbs dataset by exploring their ML phylogenetic trees, which have strong bootstrap support (i.e., >90) after 1000 replicates. Briefly, HBA1 and HBAD did not present any duplication or gene conversion events, and they were retrieved as monophyletic. In contrast, HBB1 and HBB2 alignments revealed several gene duplication events across the complete dataset of land reptiles (i.e., turtles, lizards, and snakes). These results are not unexpected, as most *b*-globin genes (including these Hbs) have evolved independently, and such loci are prone to duplication events across vertebrates, including reptiles (Hoffmann et al. 2018). Therefore, we considered further testing for gene duplication and gene conversion events in the HBB datasets. For this purpose, we translated the HBB1 and HBB2 alignments into amino acids via a universal genetic code in Mesquite v3.70 (Maddison and Maddison 2023). We subsequently tested gene conversion between HBBs via GENECONV with 10,000 permutations (Sawyer 1989; Sawyer 1999). This approach looks for the most likely candidate segment of DNA that could be copied into another segment, causing a gene conversion of this recipient locus. The results of these analyses include a global p-value and a set of pairwise p-values that were Bonferroni corrected based on 10,000 permutations. Significant positive gene conversion events were determined if $p < 0.05$.

Phylogenetic correlations between substitution rates and environmental variables: We used CoEvol v.1.6 (Lartillot and Poujol 2011) to derive pairwise correlations between

substitution rates and environmental variables. CoEvol provides a robust approach, as it calculates Bayesian estimated correlations between quantitative traits and variation in substitution rates of the gene alignment in a phylogenetic context (Lartillot and Poujol 2011). Additionally, CoEvol allows us to analyze each variable being controlled by the other variables in the study by implementing partial correlations and precision matrices. For this approach, we used each Hbs alignment as molecular data entry, environmental variables for each species, and the Squamate+Testudinea–ToL as a guide phylogeny. We note that environmental variables represent trait and physiological adaptations to such abiotic conditions, such as ambient temperature, which represents the thermal tolerance construct of each species to such conditions. The CoEvol algorithm does not accept input variable values equal to zero; we adjusted such environmental values by adding a unit to such zero values, as this does not affect our analyses, given that all nonzero values were >10 units. During the CoEvol analysis, we ran two Markov chains for each Hbs gene under the 'custom' model; under this, traits are correlated with the synonymous substitution rate (*dS*) and the ratio of nonsynonymous (*dN*) to synonymous (*dS*) substitution rates as $\omega = dN/dS$ for each Hbs (Massingham and Goldman 2005). The ω (*dN*/*dS*) ratio provides insights into these evolutionary pathways of protein-coding genes and can be used as evidence of adaptive evolution. Therefore, an excess of nonsynonymous substitutions suggests adaptive evolution, with $\omega > 1$ indicating increased amino acid variability and positive selection, whereas $\omega < 1$ indicates conservation and purifying selection.

We checked for convergence via the 'tracecomp' function to determine if the runs were of sufficient quality. Since each gene analysis converged differently, we applied different burn-ins to each Hbs analysis. Still, we always conserved an effect size > 300 and a real difference between chains < 0.1, as suggested by Lartillot and Poujol 2011 (Table S3 for details). We considered a significant posterior probability if the values were < 0.025 and > 0.975. Mutational saturation can be present in several genes (Paradis et al. 2004), creating artifacts or stationary

estimates that can be difficult to interpret. To check for these artifacts, we visually inspected each Hbs gene, comparing its raw genetic distances versus its genetic distances under the TrN model via the R package ape v5.8 (Tamura and Nei 1993; Paradis et al. 2004). We also use the CoEvol *Kr/Kc* model, which has been proposed to overcome the saturation problem with diverse genetic datasets (Nabholz et al. 2013). This model was applied using the '*-charge*' option in CoEvol for each Hbs gene in additional analyses with the same recommendations as those for the 'custom' model. The results of the *Kr/Kc* model were similar and are provided in Table S4.

Per-site sequence evolutionary analyses of hemoglobins:

We mapped amino acid positions under positive or purifying selection on each Hbs. To perform these calculations, we used HyPhy v2.5.33 (Pond and Muse 2005) with these algorithms: fixed effects likelihood (FEL); fast, unconstrained Bayesian approximation (FUBAR); and single-likelihood ancestor counting (SLAC). The FEL method is a maximum likelihood (ML) approach that estimates *dS* and *dN* substitutions per site and employs likelihood ratio tests to compare values in determining whether ω exceeds or falls below 1 (Kosakovsky Pond and Frost 2005). A ω estimate close to 1 indicates pervasive positive selection, whereas values less than 1 suggest negative or purifying selection. The FUBAR method is similar to FEL, but it is calculated via a hierarchical Bayesian approach to estimate the *dS* and *dN* for each site (Murrell et al. 2013). The FUBAR is also considered more sensitive than the FEL when ω is close to 1. The FUBAR parameters included grid points per dimension equal to 20, the number of MCMC chains equal to 5, the length of each chain equal to $2*10^6$, and the burn-in equal to 1000. Our last positive selection test was SLAC (also an ML method), which incorporates ancestral reconstructions to test whether different sites are under different selection regimes and enables the detection of positive selection (Kosakovsky Pond and Frost 2005). For all three methods, we employed aligned sequence matrices for each Hbs dataset, a guide tree derived from the Squamate+Testudinea–ToL, and the universal genetic code as inputs. The significance

thresholds were set at a p-value ≤ 0.01 for both the FEL and SLAC approaches and a posterior probability ≥ 0.99 for FUBAR.

For a complementary method, we utilized the CODEML software from the PAML package v4 (Yang 2007), with this analysis further facilitated through the EasyCodeML interface (Gao et al. 2019). The CODEML approach has been used extensively to study protein-coding genes, estimate synonymous (*dS*) and nonsynonymous (*dN*) rates, and identify positive selection in protein evolution (Álvarez-Carretero et al. 2023). For these analyses, we used five site models implemented in CODEML: M0 (one-ratio), M1a (nearly neutral), M2a (positive selection), M3 (discrete), M7 (beta), M8 (beta and $\omega > 1$) and M8a (beta and $\omega = 1$) to identify the potential signatures of positive selection in each Hbs. Some of these models allowed ω to vary among codon sites and a fixed ω ratio in the branches. After model estimation, the fit and estimates per codon site were evaluated for positive selection and compared using a likelihood ratio test (Yang and Nielsen 2002). Bayes Empirical Bayes (BEB) inference was also used to compare the models M2 and M8 and then estimate their posterior probabilities (Yang 2005). For this last analysis, we considered that codon sites underwent positive selection if their posterior probability was > 0.95.

Data Availability Statement

All data is included in the text and supplementary materials and can be found here: https://dx.doi.org/10.6084/m9.figshare.26888569

Acknowledgments

This work was supported by the National Science Foundation [IOS CAREER # 2443460] to JCS. We are grateful to two anonymous reviewers and the editor for their help in greatly improving previous versions of this manuscript with their suggestions. Finally, thanks to St. John's University for logistical support.

REFERENCES

Abouheif E. 1999. A method for testing the assumption of phylogenetic independence in comparative data. *Evol. Ecol. Res.* 1:895–909.

Allen AP, Gillooly JF, Savage VM, Brown JH. 2006. Kinetic effects of temperature on rates of genetic divergence and speciation. *Proc. Natl. Acad. Sci.* 103:9130–9135.

Álvarez-Carretero S, Kapli P, Yang Z. 2023. Beginner's Guide on the Use of PAML to Detect Positive Selection. *Mol. Biol. Evol.* 40:msad041.

Andersen Ø, Wetten OF, De Rosa MC, Andre C, Carelli Alinovi C, Colafranceschi M, Brix O, Colosimo A. 2009. Haemoglobin polymorphisms affect the oxygen-binding properties in Atlantic cod populations. *Proc. R. Soc. B Biol. Sci.* 276:833–841.

Angilletta MJ. 2009. Thermal Adaptation: A Theoretical and Empirical Synthesis. Oxford, UK: OUP Oxford

Atkinson D, Sibly RM. 1997. Why are organisms usually bigger in colder environments? Making sense of a life history puzzle. *Trends Ecol. Evol.* 12:235–239.

Baalsrud HT, Voje KL, Tørresen OK, Solbakken MH, Matschiner M, Malmstrøm M, Hanel R, Salzburger W, Jakobsen KS, Jentoft S. 2017. Evolution of Hemoglobin Genes in Codfishes Influenced by Ocean Depth. *Sci. Rep.* 7:7956.

Balbuena JA, Míguez-Lozano R, Blasco-Costa I. 2013. PACo: A Novel Procrustes Application to Cophylogenetic Analysis. *PLOS ONE* 8:e61048.

Birukou I, Schweers RL, Olson JS. 2010. Distal histidine stabilizes bound O2 and acts as a gate for ligand entry in both subunits of adult human hemoglobin. *J. Biol. Chem.* 285:8840–8854.

Blanchet S, Fargeot L, Raffard A. 2023. Phylogenetically-conserved candidate genes unify biodiversity–ecosystem function relationships and eco-evolutionary dynamics across biological scales. *Mol. Ecol.* 32:4467–4481.

Blomberg SP, Garland JR. T, Ives AR. 2003. Testing for Phylogenetic Signal in Comparative Data: Behavioral Traits Are More Labile. *Evolution* 57:717–745.

Bonaventura C, Henkens R, Alayash AI, Banerjee S, Crumbliss AL. 2013. Molecular Controls of the Oxygenation and Redox Reactions of Hemoglobin. *Antioxid. Redox Signal.* 18:2298–2313.

Branch B. 1998. Field guide to snakes and other reptiles of South Africa. *Sanibel Isl. FL Ralph Curtis Books*.

Burnham KP, Anderson DR. 2001. Kullback-Leibler information as a basis for strong inference in ecological studies. *Wildl. Res.* 28:111–119.

Chevreux B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, Wetter T, Suhai S. 2004. Using the miraEST Assembler for Reliable and Automated mRNA Transcript Assembly and SNP Detection in Sequenced ESTs. *Genome Res.* 14:1147–1159.

Clarke A, Pörtner H-O. 2010. Temperature, metabolic power and the evolution of endothermy. *Biol. Rev.* 85:703–727.

Damsgaard C, Storz JF, Hoffmann FG, Fago A. 2013. Hemoglobin isoform differentiation and allosteric regulation of oxygen binding in the turtle, Trachemys scripta. *Am. J. Physiol.-Regul. Integr. Comp. Physiol.* 305:R961–R967.

Dugo-Cota Á, Castroviejo-Fisher S, Vilà C, Gonzalez-Voyer A. 2015. A test of the integrated evolutionary speed hypothesis in a Neotropical amphibian radiation. *Glob. Ecol. Biogeogr.* 24:804–813.

Fitzpatrick LJ, Olsson M, Parsley LM, Pauliny A, Pinfold TL, Pirtle T, While GM, Wapstra E. 2019. Temperature and telomeres: thermal treatment influences telomere dynamics through a complex interplay of cellular processes in a cold-climate skink. *Oecologia* 191:767–776.

Gangloff EJ, Telemeco RS. 2018. High Temperature, Oxygen, and Performance: Insights from Reptiles and Amphibians. *Integr. Comp. Biol.* 58:9–24.

Gao F, Chen C, Arab DA, Du Z, He Y, Ho SYW. 2019. EasyCodeML: A visual tool for analysis of selection using CodeML. *Ecol. Evol.* 9:3891–3898.

Gaston KJ. 2000. Global patterns in biodiversity. *Nature* 405:220–227.

Gillman LN, Keeling DJ, Ross HA, Wright SD. 2009. Latitude, elevation and the tempo of molecular evolution in mammals. *Proc. R. Soc. B Biol. Sci.* 276:3353–3359.

Gillman LN, McCowan LSC, Wright SD. 2012. The tempo of genetic evolution in birds: body mass and climate effects. *J. Biogeogr.* 39:1567–1572.

Gillman LN, Wright SD. 2013. Patterns of Evolutionary Speed: In Search of a Causal Mechanism. *Diversity* 5:811–823.

Gillman LN, Wright SD. 2014. Species richness and evolutionary speed: the influence of temperature, water and area. *J. Biogeogr.* 41:39–51.

Gillooly JF, Allen AP, West GB, Brown JH. 2005. The rate of DNA evolution: Effects of body size and temperature on the molecular clock. *Proc. Natl. Acad. Sci.* 102:140–145.

Gittleman JL, Kot M. 1990. Adaptation: statistics and a null model for estimating phylogenetic effects. *Syst. Zool.* 39:227–241.

Giulivi C, Davies K. 1990. A novel antioxidant role for hemoglobin. The comproportionation of ferrylhemoglobin with oxyhemoglobin. *J. Biol. Chem.* 265:19453–19460.

González-Morales JC, Beamonte-Barrientos R, Bastiaans E, Guevara-Fiore P, Quintana E, Fajardo V. 2017. A Mountain or a Plateau? Hematological Traits Vary Nonlinearly with

Altitude in a Highland Lizard. *Physiol. Biochem. Zool. Ecol. Evol. Approaches* 90:638–645.

Grigg G. 2015. Biology and Evolution of Crocodylians. Csiro Publishing

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

Hancock AM, Witonsky DB, Gordon AS, Eshel G, Pritchard JK, Coop G, Di Rienzo A. 2008. Adaptations to climate in candidate genes for common metabolic disorders. *PLoS Genet.* 4:e32.

Hardison R. 1998. Hemoglobins from bacteria to man: evolution of different patterns of gene expression. *J. Exp. Biol.* 201:1099–1117.

Hasegawa T, Shishikura F, Kuwada T. 2011. Side-necked turtle (Pleurodira, Chelonia, reptilia) hemoglobin: cDNA-derived primary structures and X-ray crystal structures of Hb A. *IUBMB Life* 63:188–196.

Hijmans RJ, Van Etten J, Cheng J, Mattiuzzi M, Sumner M, Greenberg JA, Lamigueiro OP, Bevan A, Racine EB, Shortridge A, et al. 2015. Package 'raster.' *R Package* 734.

Hillman SS, Hancock TV, Hedrick MS. 2013. A comparative meta-analysis of maximal aerobic metabolism of vertebrates: implications for respiratory and cardiovascular limits to gas exchange. *J. Comp. Physiol. B* 183:167–179.

Hindle AG. 2020. Diving deep: understanding the genetic components of hypoxia tolerance in marine mammals. *J. Appl. Physiol.* 128:1439–1446.

Ho LST, Ané C. 2013. Asymptotic theory with hierarchical autocorrelation: Ornstein–Uhlenbeck tree models. *Ann. Stat.* 41:957–981.

Hoffmann FG, Vandewege MW, Storz JF, Opazo JC. 2018. Gene Turnover and Diversification of the α- and β-Globin Gene Families in Sauropsid Vertebrates. *Genome Biol. Evol.* 10:344–358.

Hsia CC. 1998. Respiratory function of hemoglobin. *N. Engl. J. Med.* 338:239–248.

Huntingford F. 2012. The study of animal behaviour. Springer Science & Business Media

Jena NR. 2012. DNA damage by reactive species: Mechanisms, mutation and repair. *J. Biosci.* 37:503–517.

Jennions MD, Petrie M. 1997. Variation in mate choice and mating preferences: a review of causes and consequences. *Biol. Rev.* 72:283–327.

Karger DN, Conrad O, Böhner J, Kawohl T, Kreft H, Soria-Auza RW, Zimmermann NE, Linder HP, Kessler M. 2017. Climatologies at high resolution for the earth's land surface areas. *Sci. Data* 4:1–20.

Karpov A, Novikov G. 1980. Hemoglobin alloforms in cod, Gadus morhua (Gadiformes, Gadidae), their functional characteristics and occurrence in populations. *J Ichthyol* 20:45–50.

Kooyman GL. 2012. Diverse divers: physiology and behavior. Springer Science & Business Media

Kosakovsky Pond SL, Frost SDW. 2005. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. *Mol. Biol. Evol.* 22:1208–1222.

Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol. Biol. Evol.* 34:1812–1819.

Lartillot N, Poujol R. 2011. A phylogenetic model for investigating correlated evolution of substitution rates and continuous phenotypic characters. *Mol. Biol. Evol.* 28:729–744.

Lasky JR, Josephs EB, Morris GP. 2023. Genotype–environment associations to reveal the molecular basis of environmental adaptation. *Plant Cell* 35:125–138.

Legendre P, Desdevises Y, Bazin E. 2002. A Statistical Test for Host–Parasite Coevolution.Page RDM, editor. *Syst. Biol.* 51:217–234.

Ligon RA, Diaz CD, Morano JL, Troscianko J, Stevens M, Moskeland A, Laman TG, Scholes III E. 2018. Evolution of correlated complexity in the radically different courtship signals of birds-of-paradise. *PLoS Biol.* 16:e2006962.

LourençO JM, Glémin S, Chiari Y, Galtier N. 2013. The determinants of the molecular substitution process in turtles. *J. Evol. Biol.* 26:38–50.

Lu S, Xin Y, Tang X, Yue F, Wang H, Bai Y, Niu Y, Chen Q. 2015. Differences in Hematological Traits between High- and Low-Altitude Lizards (Genus Phrynocephalus). *PLOS ONE* 10:e0125751.

Maddison WP. 2007. Mesquite: a modular system for evolutionary analysis. Version 2.0. *Httpmesquiteproject Org*.

Maillett DH, Simplaceanu V, Shen T-J, Ho NT, Olson JS, Ho C. 2008. Interfacial and Distal-Heme Pocket Mutations Exhibit Additive Effects on the Structure and Function of Hemoglobin. *Biochemistry* 47:10551–10563.

Massingham T, Goldman N. 2005. Detecting Amino Acid Sites Under Positive Selection and Purifying Selection. *Genetics* 169:1753–1762.

Matsushita B, Yang W, Chen J, Onda Y, Qiu G. 2007. Sensitivity of the Enhanced Vegetation Index (EVI) and Normalized Difference Vegetation Index (NDVI) to Topographic Effects: A Case Study in High-density Cypress Forest. *Sensors* 7:2636–2651.

Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37:1530–1534.

Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K. 2013. FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. *Mol. Biol. Evol.* 30:1196–1205.

Nabholz B, Uwimana N, Lartillot N. 2013. Reconstructing the phylogenetic history of long-term effective population size and life-history traits using patterns of amino acid replacement in mitochondrial genomes of mammals and birds. *Genome Biol. Evol.* 5:1273–1290.

Natarajan C, Jendroszek A, Kumar A, Weber RE, Tame JRH, Fago A, Storz JF. 2018. Molecular basis of hemoglobin adaptation in the high-flying bar-headed goose. *PLOS Genet.* 14:e1007331.

Natarajan C, Projecto-Garcia J, Moriyama H, Weber RE, Muñoz-Fuentes V, Green AJ, Kopuchian C, Tubaro PL, Alza L, Bulgarella M, et al. 2015. Convergent Evolution of Hemoglobin Function in High-Altitude Andean Waterfowl Involves Limited Parallelism at the Molecular Sequence Level. *PLOS Genet.* 11:e1005681.

Navas CA. 2002. Herpetological diversity along Andean elevational gradients: links with physiological ecology and evolutionary physiology. *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.* 133:469–485.

Olsson M, Friesen CR, Rollings N, Sudyka J, Lindsay W, Whittington CM, Wilson M. 2018. Long-term effects of superoxide and DNA repair on lizard telomeres. *Mol. Ecol.* 27:5154–5164.

Pagel M. 1997. Inferring evolutionary processes from phylogenies. *Zool. Scr.* 26:331–348.

Pagel M. 1999. Inferring the historical patterns of biological evolution. *Nature* 401:877–884.

Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.

Patel VS, Ezaz T, Deakin JE, Marshall Graves JA. 2010. Globin gene structure in a reptile supports the transpositional model for amniote α- and β-globin gene evolution. *Chromosome Res.* 18:897–907.

Pond SLK, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies. In: Statistical methods in molecular evolution. Springer. p. 125–181.

Pough FH, Andrews RM, Cadle JE, Crump ML, Savitzky AH, Wells KD. 2016. Herpetology. Sinauer Associates, Incorporated, Publishers

Projecto-Garcia J, Natarajan C, Moriyama H, Weber RE, Fago A, Cheviron ZA, Dudley R, McGuire JA, Witt CC, Storz JF. 2013. Repeated elevational transitions in hemoglobin function during the evolution of Andean hummingbirds. *Proc. Natl. Acad. Sci.* 110:20669–20674.

Pyron RA, Burbrink FT, Wiens JJ. 2013. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evol. Biol.* 13:93.

Revell LJ, Harmon LJ, Collar DC. 2008. Phylogenetic Signal, Evolutionary Process, and Rate. *Syst. Biol.* 57:591–601.

Rolland J, Loiseau O, Romiguier J, Salamin N. 2016. Molecular evolutionary rates are not correlated with temperature and latitude in Squamata: an exception to the metabolic theory of ecology? *BMC Evol. Biol.* 16:95.

Rücknagel KP, Braunitzer G. 1988. Hemoglobins of Reptiles. The Primary Structure of the Major and Minor Hemoglobin Component of Adult Western Painted Turtle (Chrysemys picta bellii). 369:123–132.

Santos JC. 2012. Fast Molecular Evolution Associated with High Active Metabolic Rates in Poison Frogs. *Mol. Biol. Evol.* 29:2001–2018.

Sawyer S. 1989. Statistical tests for detecting gene conversion. *Mol. Biol. Evol.* 6:526–538.

Sawyer S. 1999. GENECONV: A computer package for the statistical detection of gene conversion.

Scanes CG. 2015. Chapter 10 - Blood. In: Scanes CG, editor. Sturkie's Avian Physiology (Sixth Edition). Sixth Edition. San Diego: Academic Press. p. 167–191. Available from: https://www.sciencedirect.com/science/article/pii/B9780124071605000105

Scott GR, Schulte PM, Egginton S, Scott ALM, Richards JG, Milsom WK. 2011. Molecular Evolution of Cytochrome c Oxidase Underlies High-Altitude Adaptation in the Bar-Headed Goose. *Mol. Biol. Evol.* 28:351–363.

Searle S. 1961. Phenotypic, genetic and environmental correlations. *Biometrics* 17:474–480.

Storz JF. 2007. Hemoglobin Function and Physiological Adaptation to Hypoxia in High-Altitude Mammals. *J. Mammal.* 88:24–31.

Storz JF. 2018. Hemoglobin: insights into protein structure, function, and evolution. Oxford University Press

Storz JF, Moriyama H. 2008. Mechanisms of Hemoglobin Adaptation to High Altitude Hypoxia. *High Alt. Med. Biol.* 9:148–157.

Storz JF, Natarajan C, Signore AV, Witt CC, McCandlish DM, Stoltzfus A. 2019. The role of mutation bias in adaptive molecular evolution: insights from convergent changes in protein function. *Philos. Trans. R. Soc. B Biol. Sci.* 374:20180238.

Storz JF, Opazo JC, Hoffmann FG. 2013. Gene duplication, genome duplication, and the functional diversification of vertebrate globins. *Mol. Phylogenet. Evol.* 66:469–478.

Storz JF, Runck AM, Moriyama H, Weber RE, Fago A. 2010. Genetic differences in hemoglobin function between highland and lowland deer mice. *J. Exp. Biol.* 213:2565–2574.

Storz JF, Runck AM, Sabatino SJ, Kelly JK, Ferrand N, Moriyama H, Weber RE, Fago A. 2009. Evolutionary and functional insights into the mechanism underlying high-altitude adaptation of deer mouse hemoglobin. *Proc. Natl. Acad. Sci.* 106:14450–14455.

Storz JF, Weber RE, Fago A. 2012. Oxygenation properties and oxidation rates of mouse hemoglobins that differ in reactive cysteine content. *Comp. Biochem. Physiol. A. Mol. Integr. Physiol.* 161:265–270.

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.* 10:512–526.

Tilman D. 1999. The ecological consequences of changes in biodiversity: a search for general principles. *Ecology* 80:1455–1474.

Tufts DM, Natarajan C, Revsbech IG, Projecto-Garcia J, Hoffmann FG, Weber RE, Fago A, Moriyama H, Storz JF. 2015. Epistasis Constrains Mutational Pathways of Hemoglobin Adaptation in High-Altitude Pikas. *Mol. Biol. Evol.* 32:287–298.

Varnado CL, Mollan TL, Birukou I, Smith BJZ, Henderson DP, Olson JS. 2013. Development of Recombinant Hemoglobin-Based Oxygen Carriers. *Antioxid. Redox Signal.* 18:2314–2328.

de Villota ED, Carmona MG, Rubio J, de Andres SR. 1981. Equality of the in vivo and in vitro oxygen-binding capacity of haemoglobin in patients with severe respiratory disease. *Br. J. Anaesth.* 53:1325–1328.

Vitt LJ, Caldwell JP. 2013. Herpetology: an introductory biology of amphibians and reptiles. San Diego: Academic press

Wallace WJ, Volpe JA, Maxwell JC, Caughey WS, Charache S. 1976. Properties of hemoglobin a and hemoglobin Zurich (β63 histidine→ arginine): Quantitative evaluation of functional abnormalities in hemoglobins. *Biochem. Biophys. Res. Commun.* 68:1379–1386.

Weber RE. 2007. High-altitude adaptations in vertebrate hemoglobins. *Respir. Physiol. Neurobiol.* 158:132–142.

Weber RE, Fago A, Malte H, Storz JF, Gorr TA. 2013. Lack of conventional oxygen-linked proton and anion binding sites does not impair allosteric regulation of oxygen binding in dwarf caiman hemoglobin. *Am. J. Physiol.-Regul. Integr. Comp. Physiol.* 305:R300–R312.

Winslow RM. 2007. The role of hemoglobin oxygen affinity in oxygen transport at high altitude. *Respir. Physiol. Neurobiol.* 158:121–127.

Womack MC, Bell RC. 2020. Two-hundred million years of anuran body-size evolution in relation to geography, ecology and life history. *J. Evol. Biol.* 33:1417–1432.

Wong F, Carter CK, Kohn R. 2003. Efficient estimation of covariance selection models. *Biometrika* 90:809–830.

Wood SC, Johansen K. 1974. Respiratory adaptations to diving in the nile monitor lizard,Varanus niloticus. *J. Comp. Physiol.* 89:145–158.

Wright ES. 2015. DECIPHER: harnessing local sequence context to improve protein multiple sequence alignment. *BMC Bioinformatics* 16:322.

Wright SD, Gillman LN, Ross HA, Keeling DJ. 2010. Energy and the tempo of evolution in amphibians. *Glob. Ecol. Biogeogr.* 19:733–740.

Wright SD, Ross HA, Jeanette Keeling D, McBride P, Gillman LN. 2011. Thermal energy and the rate of genetic evolution in marine fishes. *Evol. Ecol.* 25:525–530.

Yang Z. 2005. Bayes Empirical Bayes Inference of Amino Acid Sites Under Positive Selection. *Mol. Biol. Evol.* 22:1107–1118.

Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* 24:1586–1591.

Yang Z, Nielsen R. 2002. Codon-Substitution Models for Detecting Molecular Adaptation at Individual Sites Along Specific Lineages. *Mol. Biol. Evol.* 19:908–917.

Zheng Y, Wiens JJ. 2016. Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Mol. Phylogenet. Evol.* 94:537–547.

Zomer RJ, Xu J, Trabucco A. 2022. Version 3 of the global aridity index and potential evapotranspiration database. *Sci. Data* 9:409.

Table 1. Gene conversion events in the HBB1 and HBB2 genes of lizards. BC KA stands for Bonferroni-corrected Karlin-Altschul.

SPECIES	GENE 1-GENE 2	BC KA P-VALUE	PROTEIN BEGIN	PROTEIN END	LENGTH (BP)	TOTAL POLY (BP)	NUM DIFFS (BP)	TOTAL DIFFS (BP)
VARANUS BECCARII	HBB2-HBB1	0.01022	187	314	128	99	0	53
VARANUS ACANTHURUS	HBB2-HBB1	0.01442	184	320	137	108	0	48
VARANUS BECCARII/VARANUS MACRAEI	HBB2-HBB1	0.01526	187	314	128	99	0	52
VARANUS MACRAEI/VARANUS	HBB2-HBB1	0.01526	187	314	128	99	0	52

BECCARII								
OPHISAURUS GRACILIS/CTENOTUS ROBUSTUS	HBB2-HBB1	0.0215	256	333	78	60	0	81
VARANUS MACRAEI	HBB2-HBB1	0.02275	187	314	128	99	0	51

1

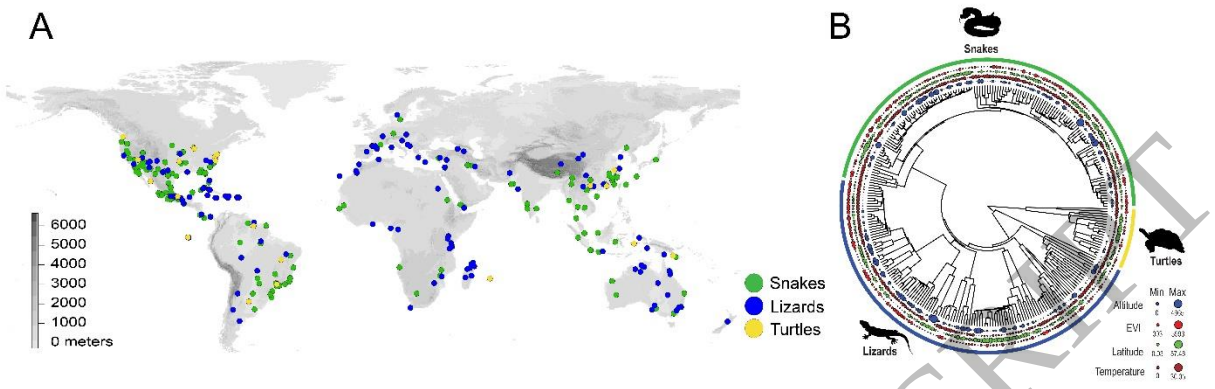

Figure 1. Spatial distribution and phylogeny of terrestrial lizards, snakes, and turtles with adult hemoglobin subunit (Hbs) and the estimates of environmental variables influencing oxygen transport by these proteins. (A) Global distribution of land reptiles used in the Hbs coevolution analyses. (B) Ecological variables affecting respiratory proteins evolution including altitude (in meters), primary biomass production as the mean annual enhanced vegetation index (EVI), absolute latitude (in degrees), and mean annual ambient temperature (in °C). Environmental estimates were extracted from global grids using the tissue sample metadata and mean distribution estimates. The circle diameters are scaled according to the minimum and maximum values of the environmental variables.

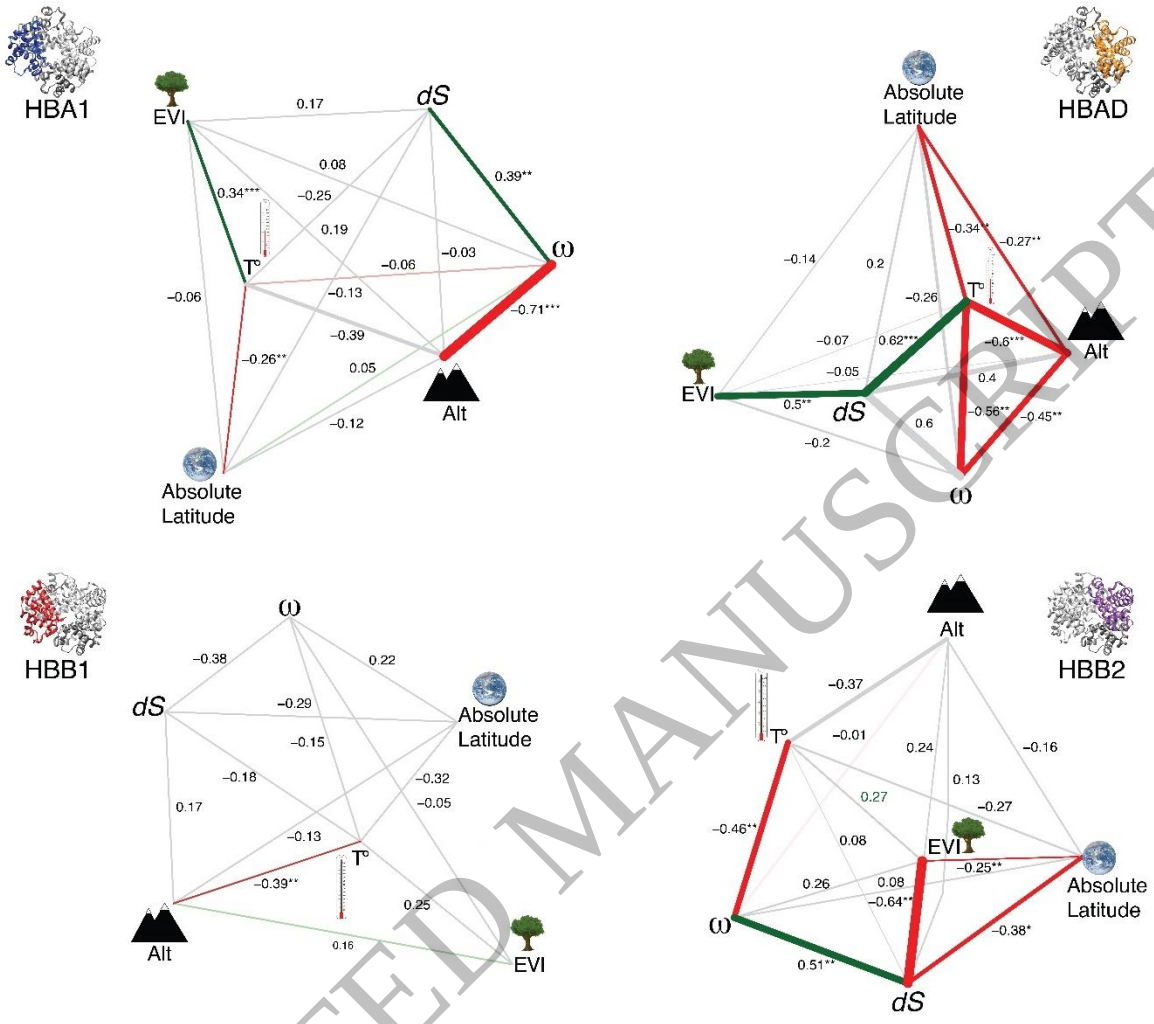

Figure 2. Partial phylogenetic correlations between the evolutionary rates of each adult hemoglobin subunit (Hbs: HBA1, HBAD, HBB1, and HBB2) and the environmental variables affecting physiological traits relevant to oxygen metabolism and transport. HBA1 and HBAD include data from lizards, snakes, and turtles. HBB1 and HBB2 are only for snakes that did not experience recurrent gene duplications to confound their homology assignation. The molecular parameters include synonymous substitution (*dS*), nonsynonymous (*dN*) substitution rates, and their ratio as $\omega = dN/dS$. The environmental variables included altitude (in meters, represented by Alt), primary biomass production as the mean annual enhanced vegetation index (unitless EVI), absolute latitude (in degrees), and mean annual ambient temperature (in °C, represented

by °T). The green lines indicate positive relationships, the red lines indicate negative relationships, and the gray lines indicate nonsignificant relationships. Numeric values of the partial correlation coefficient were derived via a precision matrix to remove confounding effects such altitude on temperature. Significance is provided as * (pp < 0.050 or pp > 0.950), ** (pp < 0.025 or > 0.975), and *** (pp < 0.001 or pp > 0.999).

Table 2. Molecular correlations between adult hemoglobin subunits (Hbs) of reptiles and environmental variables as predictors. The protein evolution parameters include the synonymous substitution (*dS*) rates and the ratio of nonsynonymous (*dN*) to synonymous (d*S*) substitution rates as $\omega = dN/dS$. All values were adjusted via a precision matrix, and their significance is * (pp < 0.050 or pp > 0.950), ** (pp < 0.025 or > 0.975), and *** (pp < 0.001 or pp > 0.999).

	Gene		***dS***	ω	**Altitude**	**Temperature**	**EVI**	**Latitude**
All reptiles	**HBA1**	*dS*	-	0.331*	-0.035	-0.165	0.239	-0.094
		ω	0.331*	-	-0.715***	0.203	0.196	0.153
	HBAD	*dS*	-	0.605***	0.404	0.619***	0.496**	0.199
		ω	0.605***	-	-0.450**	-0.556**	-0.198	-0.262
Only Snakes	**HBB1**	*dS*	-	-0.378	0.175	-0.180	0.007	-0.287
		ω	-0.378	-	0.004	-0.149	-0.051	0.216
	HBB2	*dS*	-	0.376	0.120	-0.328	-0.710***	-0.530**
		ω	0.376	-	-0.167	-0.421*	0.275	0.029

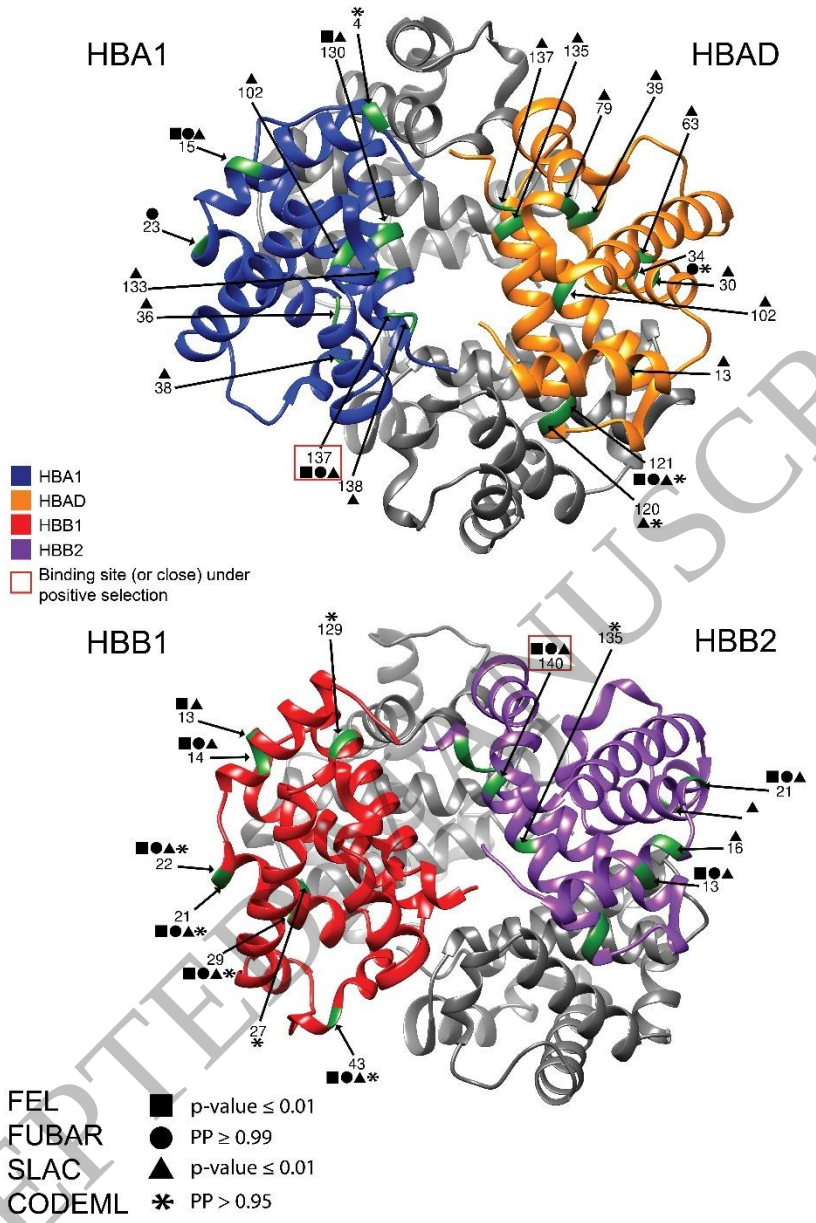

Figure 3. Predicted 3D-models of the adult hemoglobin subunits (Hbs) of reptiles: HBA1 (blue), HBAD (orange), HBB1 (red), and HBB2 (purple). Sites found under diversifying selection are indicated by a green overlaid with a line indicating the amino acid (AA) position number and analysis that supported such result (geometric shape). This numbering sequence follows the order in the only available crystallographic models for reptiles (both turtles): *Geochelone gigantea* (PDB accession number: 1V75) and *Podocnemis unifilis* (PDB accession number:

3AT5). The protein evolution analyses were done using the following models: Fixed Effects Likelihood (HyPhy-FEL, square); Fast, Unconstrained Bayesian Approximation (HyPhy-FUBAR, circle); Single-Likelihood Ancestor Counting (HyPhy-SLAC, triangle); and PAML-CODEML (asterisk). Note that sites 137th and 140th are marked within a red box as both sites are adjacent to two key heme-binding sites 136th and 141st, which both are conserved.

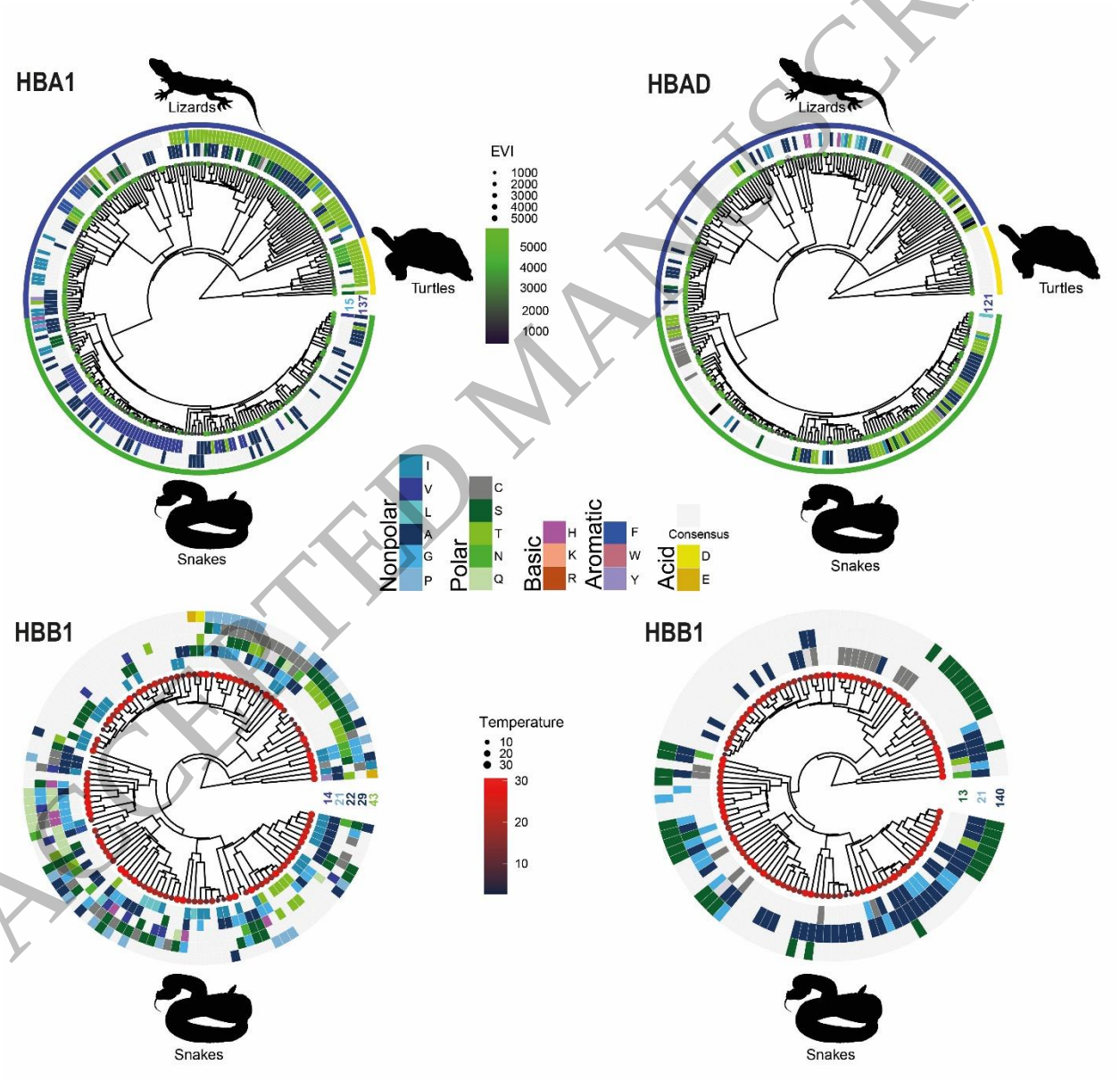

Figure 4. Pattern of changes in amino acid classes for sites under positive selection on each adult hemoglobin subunit (Hbs: HBA1, HBAD, HBB1, and HBB2). Consensus state (gray cells) represents that site's dominant and most frequent amino acid class. Amino acid changes different from the consensus state are colored by amino acid class (i.e., hydrophobicity, presence of aromatic groups, charge, and polarity). The numbers on the side of each Hbs (e.g., HBB1 sites 14, 21, 22, 29, and 43) indicate codon/amino acid positions under positive selection, which have resulted in a change from the consensus amino acid class (e.g., from polar to non-polar). The silhouettes under each plot are the reptile groups included in the analyses with verifiable gene homology. The amino acid numbering sequence follows the two available crystallographic models for reptiles (both turtles): *Geochelone gigantea* (PDB accession number: 1V75) and *Podocnemis unifilis* (PDB accession number: 3AT5).