Molecular Phylogenetics and Evolution 169 (2022) 107435

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

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

Evolution of the albumin protein family in reptiles

Emilie M. Broussard [a,b], Zachary B. Rodriguez [a,b], Christopher C. Austin [a,b,*]

[a] Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA
[b] Museum of Natural Science, Louisiana State University, 119 Foster Hall, Baton Rouge, LA 70803, USA

ARTICLE INFO

Keywords:
Albumin family
Alpha-fetoprotein
Gene family evolution
Serum albumin
Squamates

ABSTRACT

The albumin family of proteins consists of vitamin-D binding protein/group-specific component (GC), serum albumin (ALB), alpha-fetoprotein (AFP), and afamin (AFM), which are responsible for transporting many ligands throughout the body. The albumin family proteins are physiologically and medically important, but our understanding of their functions and applications is hindered by the dearth of information regarding these proteins' evolutionary relationships and functions in non-mammalian lineages. In this study we investigate the evolution of the albumin family proteins in reptiles, using bioinformatic methods to survey available reptile genomes and transcriptomes for albumin family proteins and phylogenetically characterize their relationships. We reinforce the established evolutionary relationships among the albumin protein family in reptiles, however, they are variable in their number of domains, overall genetic sequence, and synteny. We find a novel absence of the physiologically important ALB in squamates and identify two distinct lineages of AFP, one in mammals and another in reptiles. Our study provides a comparative genomic framework for further studies identifying lineage-specific gene expansions that may compensate for the lack of serum albumin in squamates.

1. Introduction

The albumin genes are a family of water-soluble transport proteins variably found in the blood plasma of vertebrates (Fasano et al., 2005). Albumin family proteins play important roles in vertebrate physiology, acting as versatile serum transport proteins by binding and delivering a wide range of molecules within the circulatory system of all vertebrates (Fanali et al., 2012). In tetrapods, the multigene family contains four major genes that arose from a series of duplication events (Gibbs et al., 1998; Nishio et al., 1996): Vitamin D-binding protein/group-specific component (GC), serum albumin (ALB), alpha-fetoprotein (AFP), and afamin (AFM).

The evolutionary relationships of the four albumin family genes are well established (Fig. 1). A gene duplication event of the ancestral albumin progenitor occurred approximately 580 million years ago (MYA), producing GC and a second copy of albumin (Gibbs et al., 1998; Haefliger et al., 1989). This second copy of albumin is differentially retained in fishes with Dipnoi (lungfish), Petromyzontiformes (lamprey), and lower teleost fishes (Osteoglossiformes, Esociformes, and Salmoniformes) retaining albumin, while Elasmobranchii, coelacanths, and higher teleosts lack albumin (Andreeva, 2019; Metcalf et al., 2007). However, in amniotes, this second copy underwent a duplication event around 295 MYA, producing ALB and a third copy that underwent a duplication event 250 MYA, giving rise to AFP and the mammal-specific AFM (Gibbs et al., 1998; Haefliger et al., 1989). The albumin family is variably distributed and exists in several forms across tetrapods. GC is conserved across all vertebrates and ALB is present in most tetrapods. AFP and AFP-like proteins, which have high sequence similarity to AFP, are present in amniotes (mammals and reptiles) but absent in amphibians and fish while AFM is found exclusively in mammals (Noël et al., 2010). In addition to serving as generalized transport proteins, each member of this family has high and unique specificity for different ligands that make albumin family proteins essential for the distribution of those ligands in the body and proper physiological function (Fanali et al., 2012; Terentiev and Moldogazieva, 2013; Voegele et al., 2002; White and Cooke, 2000).

ALB is the most well-studied and best-understood member of the albumin protein family, having been investigated for its physiological properties and potential biomedical and pharmaceutical applications (Fanali et al., 2012; Larsen et al., 2016). ALB acts as a multifunctional transport protein that binds insoluble endogenous compounds such as fatty acids, toxic waste products such as bile pigments, and many pharmaceuticals, including diazepam, warfarin, and penicillin (Curry, 2002; Evans, 2002; Fanali et al., 2012; Larsen et al., 2016). ALB is highly

* Corresponding author at: Museum of Natural Science, Louisiana State University, Baton Rouge, LA 70803, USA.
E-mail address: ccaustin@lsu.edu (C.C. Austin).

https://doi.org/10.1016/j.ympev.2022.107435
Received 21 July 2020; Received in revised form 21 January 2022; Accepted 25 January 2022
Available online 4 February 2022
1055-7903/Published by Elsevier Inc.

conserved across vertebrates — its size, shape, genetic sequence, and function as a transporter remains largely unchanged across tetrapod lineages (Noël et al., 2010). ALB is the most abundant plasma protein in mammals, functioning in the regulation of osmotic pressure and blood pH in most vertebrates (Fanali et al., 2012). ALB has also been shown to be an important circulating antioxidant (Roche et al., 2008). The ability of ALB to bind to many different ligands is due to its multi-domain structure; each of its three domains contains two binding sites for a total of six distinct binding sites each with their own affinities for different ligands (Fanali et al., 2012; Fasano et al., 2005). This three-domain structure is conserved across all four members of the protein family, but the binding sites differ in number and location (Li et al., 2017).

The three other albumin family proteins are less well characterized in comparison to ALB, but serve similar roles in the transport of molecules in the blood plasma. The function of GC is well characterized in humans but remains largely unknown in other animals (Bikle and Schwartz, 2019). As its name implies, vitamin-D binding protein (GC) principally acts as a transporter for vitamin D and its metabolites, however, it has also been found to act as an actin scavenger and play a major role in the activation of macrophages during immune and inflammatory responses in mammals (Chun, 2012; White and Cooke, 2000).

AFP is commonly referred to as the fetal counterpart to ALB in humans (Bader et al., 2004) and has diagnostic uses in human medicine, but its functions are otherwise poorly understood. AFP is expressed in the yolk of bird eggs and fetal tissue in humans, where expression of AFP peaks at 14 weeks of gestation and steadily decreases until almost ceasing at birth (Bader et al., 2004; Cordeiro and Hincke, 2016). Studies have shown that AFP is capable of binding and transporting hydrophobic ligands such as fatty acids, bilirubin, and estrogen, and it also acts as a dual regulator of cell proliferation and tissue growth (Terentiev and Moldogazieva, 2013). Although its function in pregnancy and fetal development is still being investigated, AFP has become a powerful diagnostic tool, notably as a prominent tumor marker (Abelev and Eraiser, 1999; Terentiev and Moldogazieva, 2013). Increased expression of AFP in adult tissue is characteristic of hepatocellular carcinoma, germ cell tumors, and several liver diseases (Abelev and Eraiser, 1999; Bader et al., 2004).

The remaining members of the albumin protein family, AFM and alpha-fetoprotein related gene (ARG), have been only recently identified in mammals and are poorly understood. In mammals, AFP underwent a duplication that gave rise to mammal-specific AFP and AFM. AFM is known to bind and transport vitamin E and Wnt proteins, which are important signaling molecules in many cellular and developmental pathways (Jerkovic et al., 2005; Naschberger et al., 2017). ARG is a recently proposed member of the albumin family (Naidu et al., 2010). While its function is unknown, ARG is known to be expressed in mouse, rat, dog, and horse genomes, but is an inactive and unexpressed gene in primates (Naidu et al., 2010).

Comparative studies on albumin family proteins have focused on mammals due to their relative abundance of whole-genome assemblies and importance to human medicine (Li et al., 2017). However, their functions and underlying genetic sequences are poorly characterized in other vertebrates. Until recently, a comprehensive examination of albumin family variation in sauropsids, the clade sister to mammals

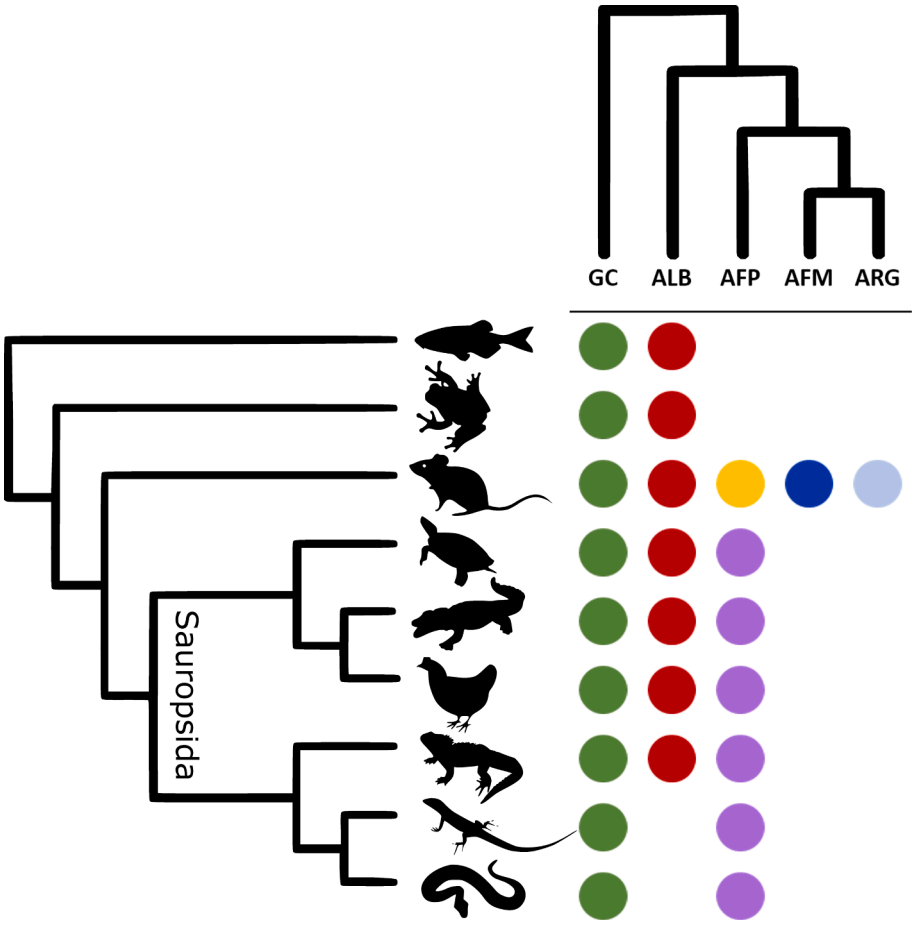

Fig. 1. The accepted vertebrate phylogeny with a corresponding matrix showing our findings of the distribution of albumin family members in each lineage. The clade Sauropsida, which contains all extant reptile lineages including birds, is labeled. Above the matrix is a representation of the previously accepted relationships of the major albumin family members.

containing all extant reptiles (including birds), was not possible due to a lack of comparative data from crocodilians, turtles, squamates (lizards and snakes), and rhynchocephalians (tuatara). This dearth of genomic data has limited our ability to decipher orthologous and paralogous relationships among the albumin genes of different vertebrates. With the recent release of multiple nonavian sauropsid genomes, we can now extend these studies to include all major groups of amniotes. Comparative studies suggest that most amniotes possess a full set of albumin proteins, several with an expanded repertoire (Li et al., 2017).

In this study, we characterize the evolutionary relationships of albumin family proteins in sauropsids and describe the absence of ALB in squamate reptiles. We also find two phylogenetically distinct AFP proteins, one in mammals and one in reptiles, that share conserved AFP domains. Our results indicate that different sauropsid lineages have diverse repertoires of albumin proteins derived from differential retention of ancestral genes.

2. Materials and methods

2.1. Bioinformatic searches

Tetrapod albumin sequences were retrieved from the spotted gar (*Lepisosteus oculatus*), zebrafish (*Danio rerio*), and the western clawed frog (*Xenopus tropicalis*), which was chosen (vs. *X. laevis*) for its diploid genome. Albumin family reference sequences were retrieved from the UniProtKB/Swiss-Prot database (Acids research, 2017) for house mouse (*Mus musculus*) and chicken (*Gallus gallus*). We queried several sauropsid genomes for albumin sequences, with emphasis on squamate reptiles. We searched the NCBI genomic database (refseq_genomic) using a low stringency BLASTx algorithm, with match/mismatch scores of 1 and 1, and gap existence and extension costs of 2 and 1, respectively, using our reference sequences as input. Because tuatara (*Sphenodon punctatus*) is the sole representative of the order Rhynchocephalia (sister to Squamata), we included four transcriptome-derived sequences in the analyses of genome-derived sequences (Tzika et al., 2015). We obtained multiple best hits (e-value cutoff of eˆ-6 and identical bit scores) for several species, likely representing paralogous sequences and/or isoforms. Thus, we performed functional annotation using InterProScan 5 (Jones et al., 2014) and removed sequences that lacked conserved protein domain annotations for each gene. If multiple sequences remained, we then chose the longest sequence as a representative. The full set of species and genes is listed in Supplementary Table S1.

2.2. Phylogenetic analyses

To determine the orthology of albumin family sequences, we reconstructed an albumin family tree under Maximum Likelihood (ML) and Bayesian Inference (BI). First, nucleotide and amino acid sequences were aligned using MAFFT (Katoh and Standley, 2013) and trimmed/cleaned with trimAL (Capella-Gutierrez et al., 2009). ML estimation was completed in IQTREE (Nguyen et al., 2015), allowing the program to assign WAG + R4 as the best-fitting model of molecular evolution using ModelFinder (Kalyaanamoorthy et al., 2017). Nodal support for the best ML topology was assessed using 1000 non-parametric bootstrap replicates. Bayesian Inference (BI) was performed in MrBayes v3.2 (Ronquist et al., 2012) using four independent runs with four chains each for 5×10^{10} generations, sampling trees every 5000 generations, and using default priors. Runs and chains were considered to have reached stationarity if the average standard deviation of split frequencies was <0.01. Convergence was assessed by checking that the effective sample size of all priors/variables was >200 and manually examined in Tracer v1.7.1 (Rambaut et al., 2018). We discarded trees collected before the chains reached convergence and summarized results with a majority-rule consensus of trees.

2.3. Synteny search

Sequences and gene order were obtained from the genome assemblies of the mouse *Mus musculus* (mm10/GRCm38), chicken *Gallus gallus* (galGal6/GRCg6a), and Green Anole *Anolis carolinensis* (AnoCar2.0) available through the NCBI (https://www.ncbi.nlm.nih.gov/genome/gdv/) and UCSC (https://genome.ucsc.edu/) websites (last accessed: 15 March 2020). Gene order was manually recorded and represented in Fig. 3.

3. Results

3.1. Bioinformatic surveys

We integrated synteny, phylogenetic, and genomic information to reconstruct the evolutionary history of the albumin family of proteins in sauropsids. We queried the genome databases to locate genes with sequence similarity to albumin family proteins using mammals, birds, frogs, and fish as references. We functionally annotated albumin family genes and used phylogenetic reconstructions to resolve orthology and paralogy. These surveys spanned all orders of sauropsids, including 15 species of squamates (lizards and snakes), 4 turtles, 2 crocodilians, plus zebra finch and chicken as representative birds. A total of 24 species covering several major squamate lineages were examined for albumin-like sequences. We searched for putative albumins out of the available overall protein sequences, including turtles, birds, alligators, and squamates (lizards and snakes) using mouse, chicken, frog, and fish albumins as reference. Our findings indicate that they are variably present and exist in several forms. Topological relationships for the albumin family were concordant between maximum likelihood and Bayesian approaches with high support (>70 bootstrap and >0.70 posterior probability).

3.2. ALB is absent in squamates

While we found many albumin-like proteins in our sequence searches, we were not able to ascribe any of these sequences to ALB for any snake or lizard, despite extensive annotation efforts. In contrast, typical ALB is present in all other sauropsids surveyed (Fig. 2). We identified well-described ALB in fish, mammals, frogs, archosaurs (birds and alligators), and turtles. Many squamate sequences that had been assigned ALB annotations were more closely related to AFP. Notably, the tuatara (*Sphenodon punctatus*), sister to squamate reptiles, has a partial (degraded) ALB sequence. This absence of typical ALB is further supported by synteny analysis of a representative of the order Squamata, the Green Anole (*Anolis carolinensis)*. In mammals and birds (represented by *Mus musculus* and *Gallus gallus,* respectively), the gene order of serum albumin and the neighboring genes is conserved and placed on a single chromosome with the gene order: COX18-ANKRD17-ALB-AFP-RASSF6, with AFM and ALBFM1 being present in *M. musculus* between AFP and RASSF6 (Fig. 3). In *A. carolinensis*, however, this region has been split, with COX18 and ANKRD17 found on chromosome 6, AFP found on an uncharacterized chromosome, and ALB and RASSF6 unable to be recovered (Fig. 3).

3.3. Vitamin D-binding protein is highly conserved across reptiles

GC is highly conserved across all lineages and reflects the expected species topology (Fig. 2). All protein sequences identified as GC contained 6–9 GC-specific fingerprints (PR00804 from the PRINTS database) and most (17/23) sequences were annotated with the Pfam profile model PF09164. GC serves several physiologically important functions, including acting as the principal transporter of vitamin D and its metabolites, scavenging G-actin, and activating macrophages in immune responses (Chun, 2012; White and Cooke, 2000), and no cases of complete loss have been reported. This suggests GC is essential for normal

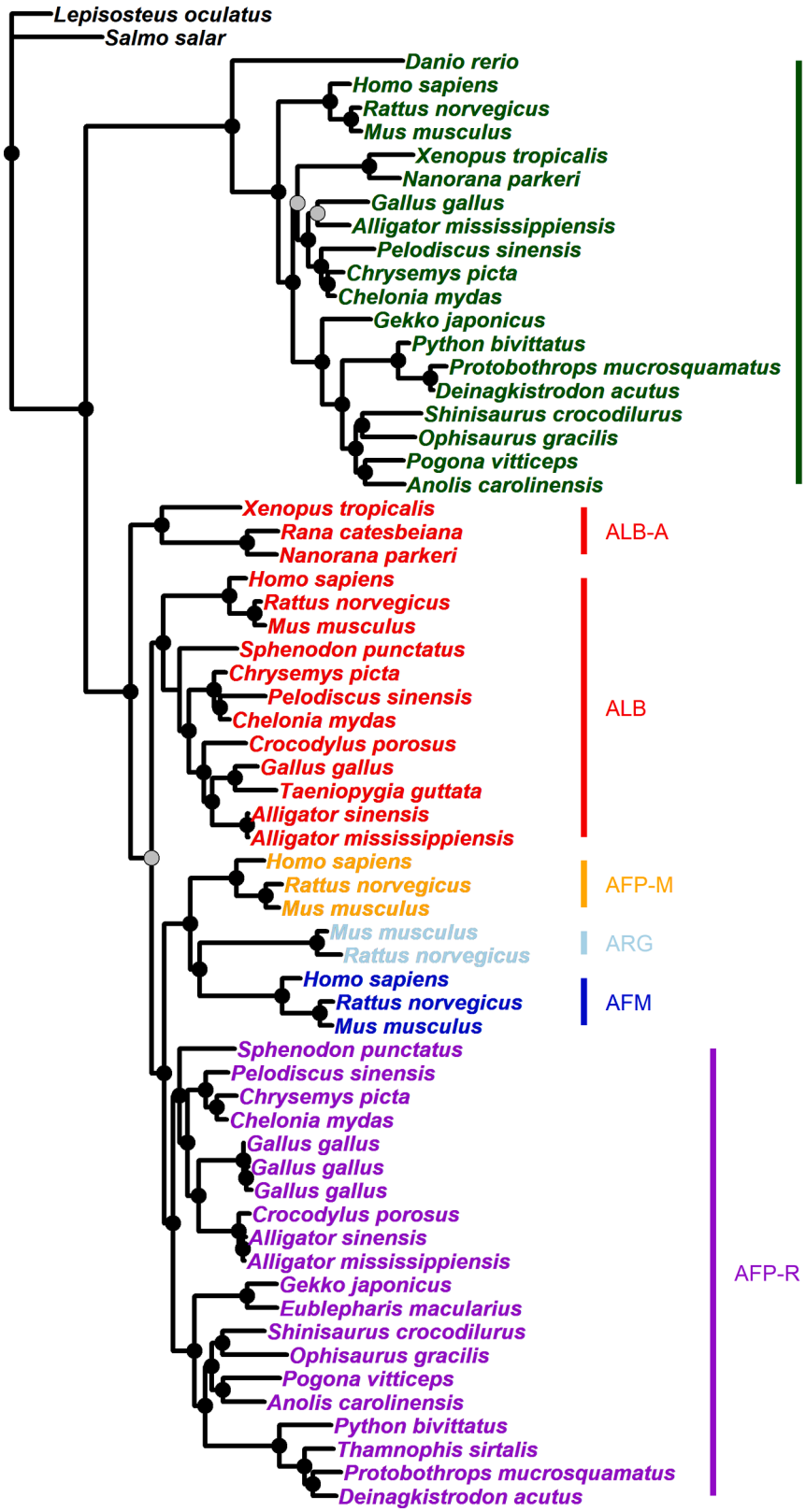

Fig. 2. Maximum Likelihood phylogeny of albumin family amino acid sequences. Bootstrap support is represented at nodes with black circles representing bootstrap values of >70 and gray circles representing values of <70. Protein clades are labeled with the protein name and a colored bar. Fish albumin predates the duplication event that occurred in tetrapods and was used to root the tree. Colors correspond to each family protein: green for GC, red for ALB, yellow for AFP, purple for AFP-R, blue for AFM, and light blue for ARG. AFP-R represents reptilian AFP, a reptile specific AFP distinct from mammal AFP. ALB-A represents amphibian serum albumin, which predates the duplication event that occurred in amniotes to give rise to ALB and AFP. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

development and survival and is likely under strong purifying selection.

3.4. Alpha-fetoprotein is highly variable in reptiles

All protein sequences categorized as AFP were annotated with 3 characteristic ALB domains (PFAM profile PF00273) from the Pfam database. However, the number of reptile AFP fingerprint motifs (PRINTS PR00803) per sequence was highly variable among reptiles, compared to 7-fingerprint mammal AFP. Most turtle, crocodilian, and bird AFP sequences contained 6 AFP fingerprints, while squamate AFP

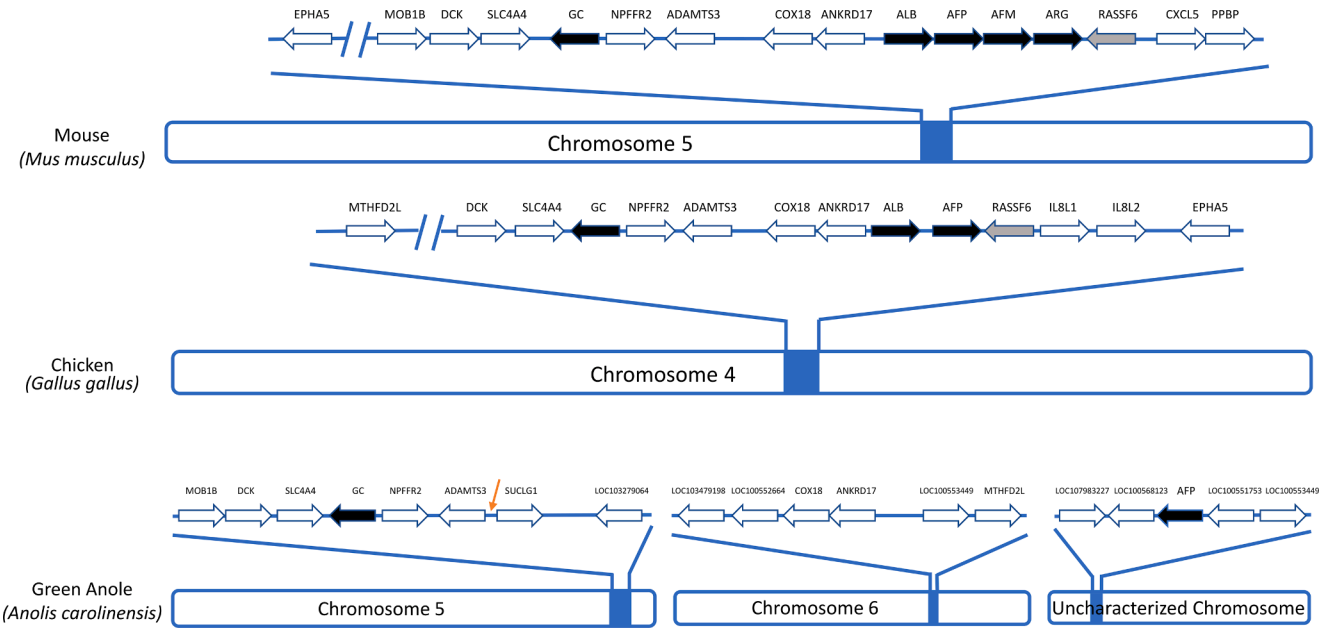

Fig. 3. Synteny of albumin gene family chromosomal regions in the house mouse (*Mus musculus*), chicken (*Gallus gallus*), and green anole (*Anolis carolinensis*) genomes. Members of the albumin gene family are shown in black while neighboring genes are shown in white. The gene RASSF6 in *M. musculus* and *G. gallus* is shown in gray to emphasize its absence in *A. carolinensis*. The highlighted blue portion of each chromosome corresponds to the approximate region containing the shown genes. The orange arrow in *A. carolinensis* chromosome 5 represents a point of interest where a translocation may have occurred. The line breaks in *M. musculus* chromosome 5 and *G. gallus* chromosome 4 represent a distance of 4 million base pairs. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

sequences had 3–5 AFP fingerprints. Putative reptile AFP sequences were similarly variable in ProDom and PROSite annotations when compared to mammals. Phylogenetic analysis shows that AFP is paraphyletic, with at least two distinct clades of AFP genes: one found in mammals (Fig. 2 as AFP-M), and another that is reptile-specific (Fig. 2 as AFP-R). Within the reptile-specific AFP clade (AFP-R), we find evidence for a major split between squamates and other reptiles. Tuatara, the sister group to squamates, has an AFP that is sister to turtles, crocodilians, and birds (Fig. 2), while all squamate AFP sequences form a monophyletic clade sister to the rest of sauropsids (Fig. 2). Maximum Likelihood and Bayesian Inference strongly support Sphenodon AFP as sister to Archosaurs and turtles (Supplementary Fig. 1). However, topology tests forcing tuatara AFP as sister to squamate AFP cannot be significantly rejected for the Shimodaira-Hasegawa test ($p = 0.174$) or approximately unbiased test ($p = 0.162$), where $p < 0.05$ would be significant.

3.5. Only characteristic ALB residues are conserved across reptiles

Characteristic motifs and residues in the four albumin family genes were identified. Only the most conserved domains (e.g., PFAM profile PR00802) were conserved across putative ALB sequences in our study. All other domains and motifs were highly variable in both sequence and number. It is clear the overall sequence identities of albumin characteristic domains are very low across vertebrates.

4. Discussion

Given its physiological importance and relative conservation across tetrapods, we expected ALB to be retained in all reptiles. Surprisingly, we did not detect ALB for any of the 15 squamate species spanning 8 families in this study, despite exhaustive annotation efforts. Since all known ALB proteins retain highly conserved domains, it is detectable even in those vertebrates that have an expanded albumin repertoire. Thus, our findings indicate the loss of or modification beyond recognition of ALB in squamates.

The clade Sauropsida consists of extant reptiles, a diverse set of organisms. Squamates (lizards and snakes) represent the largest order of Sauropsida and are the second largest order of vertebrates, with over 10,000 described species (Uetz and Etzold 1996) and a wide variety of adaptations. The loss of a major physiological protein such as ALB is surprising given its important function and sequence conservation in other lineages. This lack of ALB presumably leads to major physiological shifts to account for this absence.

Modifications in other blood plasma proteins may compensate for the lack of ALB in Squamates. Although ALB serves vital physiological transport roles, congenital analbuminemia is a human medical condition characterized by the absence or abnormally low levels of ALB circulating in the blood serum (Koot et al., 2004; Minchiotti et al., 2019). This condition is an inherited autosomal recessive disorder with an incidence of 1:1,000,000 live births, arising from a number of possible mutations that cause premature truncations in the ALB molecule (Caridi et al., 2019; Minchiotti et al., 2019). Surprisingly, this condition has long believed to be largely asymptomatic and there are only 90 cases recorded worldwide (Cormode et al., 1975; Watkins et al., 1994), but recent studies have shown that congenital analbuminemia can lead to the miscarriage and preterm birth of fetuses with the condition by means of low amniotic fluid levels and placental abnormalities, and increases the chance of death in early childhood (Minchiotti et al., 2019). The few reported cases of patients with congenital analbuminemia surviving until adulthood still suffer several mild detrimental effects such as fatigue and reduced blood pressure, however they survive due to compensatory measures of other plasma proteins (Minchiotti et al., 2019). While most bony fishes lack the albumin gene, this is due to the divergent evolutionary paths that blood plasma took in bony fish and other vertebrates. Fish that lack albumin have an expanded repertoire of multifunctional lipoproteins that function in osmoregulation and lipid transport, while in other vertebrates those functions have been split between albumin proteins and lipoproteins (Andreeva, 2019). The fact that other plasma proteins counterpoise the lack of albumin function in most fish and in humans with nonfunctional ALB is intriguing and suggests that modifications in other plasma proteins have evolved in

squamates to compensate for their lack of ALB. By characterizing differences in albumin composition derived from gene duplications, losses, and retention of ancestral genes in non-mammalian lineages, we may gain insight into how these genetic changes affect physiological adaptations in these lineages, and better understand human pathologies and potential therapies (Nurdiansyah et al., 2016). Exciting future challenges are to identify what compensatory mechanisms have evolved in snakes and lizards and if these mechanisms could be used to develop biomedical therapies for congenital analbuminemia.

Until recently the availability of squamate genomes was extremely limited, and even now squamate genome assemblies are notoriously low-quality, limiting the scope of our bioinformatic searches and observation of synteny. Our inability to recover ALB in our searches may be due to assembly artifacts masking the true status of ALB in squamates. In our synteny comparison Anolis carolinensis was chosen to represent squamates (Fig. 3), however cursory searches of other published squamate genomes (*Gekko japonicus*, *Podarcis muralis*, *Pogona vitticeps,* and *Crotalus tigris*) revealed that there is a substantial amount of variation in the gene order of the region containing the albumin family genes in squamates (Supplementary Figure S3). While this implies that this region is dynamic in squamates, it is difficult to conclude whether this is a case of genuine dynamic genome architecture or due to the quality of these assemblies. Our findings suggest that ALB has been lost in squamates, but there is a possibility that ALB is present in the squamate genome but has undergone a translocation to a different chromosome and has not been recovered in our search due to the current poor quality of squamate genomes.

Although GC protein sequence remains conserved across reptiles and mammals, we find high variability in AFP proteins. We found two major clades of AFP in tetrapods, one in mammals and one in sauropsids (Fig. 2). Additionally, there is a major split in reptilian AFP (AFP-R): squamate AFP forms a distinct clade from other sauropsids (tuatara, turtles, crocodilians, and birds) (Fig. 2). AFP-R could have a single origin in the ancestral Lepidosauria, which is suggested to be the most parsimonious origin by the topology tests which found low support for tuatara as sister to turtles, crocodilians, and birds. This pattern could have resulted from multiple different models of retention of duplicates. For example, we cannot exclude the possibility of a duplication in the common ancestor of Sauropsids, with reciprocal losses in squamate on one side, and tuatara, turtles, crocodilians, and birds on the other. But narrowing down more complex models of gene retention and loss is challenging given the highly fragmented nature of reptile genomes and lack of experimentally verified AFP proteins. For this study, we used strict filtering parameters to select albumin family proteins that were sufficiently annotated. As gene annotation is improved, future work should include all identifiable albumin family variations to discover paralogs, and determine gene gains and losses in reptiles.

The function of AFP has only been characterized in mammals, where it has long been known as the fetal counterpart to ALB (Bader et al., 2004). AFP is an important transporter in mammals during fetal development due to its ability to cross the placental barrier, unlike ALB (Newby et al., 2005). In sauropsids, the role of AFP in development, as well as its expression and function in adulthood, remains unknown. The modified squamate AFP could have adapted as a compensatory mechanism for the loss of ALB, but the absence of knowledge of the function of AFP in non-mammalian lineages limits our understanding of this protein and its potential function. The divergence of reptilian and squamate AFP suggests it has evolved novel functions in these lineages that warrant rigorous investigation.

We provide a comparative genomic framework for identifying any lineage-specific expansions in gene repertoire that may compensate for the lack of the physiologically important serum albumin in squamates. Comparative studies on reptiles, especially lizards and snakes, suffer from a dearth of genomic resources, and the resources that are available are incomplete and plagued with assembly artifacts (Hoffmann et al., 2018). Future studies with increased sampling of well-annotated genomes for reptiles could effectively explore the expansion and contraction of gene families in squamates more thoroughly. Characterizing these proteins and their relationships in understudied lineages will result in the discovery of novel sequences and functions and may lead to biomedical innovations and nontraditional applications to human health problems.

5. Data availability

All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. All sequences used in this study were already publicly available (Supplementary Table S1). Sequence alignments, command scripts, and tree files related to this paper may be requested from the authors.

CRediT authorship contribution statement

Emilie M. Broussard: Data curation, Methodology, Writing – review & editing, Validation. **Zachary B. Rodriguez:** Data curation, Methodology, Writing – review & editing, Validation. **Christopher C. Austin:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing, Visualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank Maheshi Dassanayake, Vinson Doyle, and Darin Rokyta for providing helpful comments on previous drafts of this manuscript. The high-performance computing resources at LSU were used for all computational analyses. This work was funded in part by National Science Foundation grants DEB-1146033 and DEB-1926783 to CCA.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2022.107435.

References

Abelev, G.I., Eraiser, T.L., 1999. Cellular aspects of alpha-fetoprotein reexpression in tumors. Semin. Cancer Biol. 9 (2), 95–107.
Acids research N, 2017. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 45, D158–D169.
Andreeva, A.M., 2019. The strategies of organization of the fish plasma proteome: with and without albumin. Russ. J. Mar. Biol. 45 (4), 263–274.
Bader, D., Riskin, A., Vafsi, O., Tamir, A., Peskin, B., Israel, N., Merksamer, R., Dar, H., David, M., 2004. Alpha-fetoprotein in the early neonatal period—a large study and review of the literature. Clin. Chim. Acta 349 (1-2), 15–23.
Bikle, D.D., Schwartz, J., 2019. Vitamin D Binding Protein, Total and Free Vitamin D Levels in Different Physiological and Pathophysiological Conditions. Front. Endocrinol. 10, 317.
Capella-Gutierrez, S., Silla-Martinez, J.M., Gabaldon, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25 (15), 1972–1973.
Caridi, G., Lugani, F., Rigat, B., Van den Abeele, P., Layet, V., Gavard, M.-S., Campagnoli, M., Galliano, M., Minchiotti, L., 2019. A novel insertion (c.1098dupT) in the albumin gene causes analbuminemia in a consanguineous family. Eur. J. Med. Genet. 62, 144–148.
Chun, R.F., 2012. New perspectives on the vitamin D binding protein. Cell Biochem. Funct. 30 (6), 445–456.
Cordeiro, C.M.M., Hincke, M.T., 2016. Quantitative proteomics analysis of eggshell membrane proteins during chick embryonic development. J. Proteomics 130, 11–25.
Cormode, E.J., Lyster, D.M., Israels, S., 1975. Analbuminemia in a neonate. J. Pediatr. 86 (6), 862–867.
Curry, S., 2002. Beyond expansion: structural studies on the transport roles of human serum albumin. Vox Sang. 83 (Suppl 1), 315–319.

Evans, T.W., 2002. albumin as a drug—biological effects of albumin unrelated to oncotic pressure. Aliment. Pharmacol. Ther. 16, 6–11.
Fanali, G., di Masi, A., Trezza, V., Marino, M., Fasano, M., Ascenzi, P., 2012. Human serum albumin: from bench to bedside. Mol. Aspects Med. 33 (3), 209–290.
Fasano, M., Curry, S., Terreno, E., Galliano, M., Fanali, G., Narciso, P., Notari, S., Ascenzi, P., 2005. The extraordinary ligand binding properties of human serum albumin. IUBMB Life 57 (12), 787–796.
Gibbs, P.E.M., Witke, W.F., Dugaiczyk, A., 1998. The molecular clock runs at different rates among closely related members of a gene family. J. Mol. Evol. 46 (5), 552–561.
Haefliger, D.N., Moskaitis, J.E., Schoenberg, D.R., Wahli, W., 1989. Amphibian albumins as members of the albumin, alpha-fetoprotein, vitamin D-binding protein multigene family. J. Mol. Evol. 29 (4), 344–354.
Hoffmann, F.G., Vandewege, M.W., Storz, J.F., Opazo, J.C., 2018. Gene Turnover and Diversification of the α- and β-Globin Gene Families in Sauropsid Vertebrates. Genome Biol. Evol. 10, 344–358.
Jerkovic, L., Voegele, A.F., Chwatal, S., Kronenberg, F., Radcliffe, C.M., Wormald, M.R., Lobentanz, E.M., Ezeh, B., Eller, P., Dejori, N., Dieplinger, B., Lottspeich, F., Sattler, W., Uhr, M., Mechtler, K., Dwek, R.A., Rudd, P.M., Baier, G., Dieplinger, H., 2005. Afamin is a novel human vitamin E-binding glycoprotein characterization and in vitro expression. J. Proteome Res. 4 (3), 889–899.
Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.-Y., Lopez, R., Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30 (9), 1236–1240.
Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14 (6), 587–589.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30 (4), 772–780.
Koot, B.G.P., Houwen, R., Pot, D.-J., Nauta, J., 2004. Congenital analbuminaemia: biochemical and clinical implications. A case report and literature review. Eur. J. Pediatr. 163, 664–670.
Larsen, M.T., Kuhlmann, M., Hvam, M.L., Howard, K.A., 2016. Albumin-based drug delivery: harnessing nature to cure disease. Mol. Cell Ther. 4, 3.
Li, S., Cao, Y., Geng, F., 2017. Genome-Wide Identification and Comparative Analysis of Albumin Family in Vertebrates. Evol. Bioinform. Online 13, 1176934317716089.
Metcalf, V.J., George, P.M., Brennan, S.O., 2007. Lungfish albumin is more similar to tetrapod than to teleost albumins: purification and characterisation of albumin from the Australian lungfish, Neoceratodus forsteri. Comp. Biochem. Physiol. B Biochem. Mol. Biol. 147 (3), 428–437.
Minchiotti, L., Caridi, G., Campagnoli, M., Lugani, F., Galliano, M., Kragh-Hansen, U., 2019. Diagnosis, Phenotype, and Molecular Genetics of Congenital Analbuminemia. Front. Genet. 10, 336.
Naidu, S., Peterson, M.L., Spear, B.T., 2010. Alpha-fetoprotein related gene (ARG): a new member of the albumin gene family that is no longer functional in primates. Gene 449 (1-2), 95–102.
Naschberger, A., Orry, A., Lechner, S., Bowler, M.W., Nurizzo, D., Novokmet, M., Keller, M.A., Oemer, G., Seppi, D., Haslbeck, M., Pansi, K., Dieplinger, H., Rupp, B., 2017. Structural Evidence for a Role of the Multi-functional Human Glycoprotein Afamin in Wnt Transport. Structure 25 (12), 1907–1915.e5.
Newby, D., Dalgliesh, G., Lyall, F., Aitken, D.A., 2005. Alphafetoprotein and alphafetoprotein receptor expression in the normal human placenta at term. Placenta 26 (2-3), 190–200.
Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274.
Nishio, H., Heiskanen, M., Palotie, A., Bélanger, L., Dugaiczyk, A., 1996. Tandem arrangement of the human serum albumin multigene family in the sub-centromeric region of 4q: evolution and chromosomal direction of transcription. J. Mol. Biol. 259 (1), 113–119.
Noël, E.S., Reis, M.D., Arain, Z., Ober, E.A., 2010. Analysis of the Albumin/α-Fetoprotein/Afamin/Group specific component gene family in the context of zebrafish liver differentiation. Gene Expr. Patterns 10 (6), 237–243.
Nurdiansyah, R., Rifa'i, M., Widodo, 2016. A comparative analysis of serum albumin from different species to determine a natural source of albumin that might be useful for human therapy. J. Taibah Univ. Med. Sci. 11 (3), 243–249.
Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst. Biol. 67, 901–904.
Roche, M., Rondeau, P., Singh, N.R., Tarnus, E., Bourdon, E., 2008. The antioxidant properties of serum albumin. FEBS Lett. 582, 1783–1787.
Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542.
Terentiev, A.A., Moldogazieva, N.T., 2013. Alpha-fetoprotein: a renaissance. Tumour Biol. 34 (4), 2075–2091.
Tzika, A.C., Ullate-Agote, A., Grbic, D., Milinkovitch, M.C., 2015. Reptilian Transcriptomes v2.0: An Extensive Resource for Sauropsida Genomics and Transcriptomics. Genome Biol. Evol. 7, 1827–1841.
Voegele, A.F., Jerković, L., Wellenzohn, B., Eller, P., Kronenberg, F., Liedl, K.R., Dieplinger, H., 2002. Characterization of the vitamin E-binding properties of human plasma afamin. Biochemistry 41 (49), 14532–14538.
Watkins, S., Madison, J., Galliano, M., Minchiotti, L., Putnam, F.W., 1994. Analbuminemia: three cases resulting from different point mutations in the albumin gene. Proc. Natl. Acad. Sci. USA 91 (20), 9417–9421.
White, P., Cooke, N., 2000. The multifunctional properties and characteristics of vitamin D-binding protein. Trends Endocrinol. Metab. 11 (8), 320–327.