4-2020

Evolution of the albumin protein family in reptiles

Emilie Broussard

Evolution of the albumin protein family in reptiles

by

Emilie Broussard

Undergraduate honors thesis under the direction of

Dr. Christopher Austin

Department of Biological Sciences

Submitted to the LSU Roger Hadfield Ogden Honors College in partial fulfillment of the Upper Division Honors Program.

April, 2020

Louisiana State University

& Agricultural and Mechanical College

Baton Rouge, Louisiana

Evolution of the albumin protein family in reptiles

Emilie Broussard[1,2], Zachary B. Rodriguez[1,2], Christopher C. Austin[1,2,*]

[1] Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana, 70803 USA

[2] Museum of Natural Science, Louisiana State University, 119 Foster Hall, Baton Rouge, Louisiana, 70803, USA

* Corresponding Austin: ccaustin@lsu.edu

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). These proteins are found in the blood plasma of tetrapods and are responsible for transporting a large number of ligands throughout the body. This family is well studied in mammals but poorly characterized 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 reveal that the albumin protein family in reptiles is variable in both repertoire and genetic sequence, notably identifying an apparent loss of ALB in squamates and a new type of albumin family protein 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.

Key words: albumin family, serum albumin, gene family evolution, squamates

Introduction

The albumin gene family consists of four water-soluble proteins found in blood plasma of tetrapods (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 is comprised of 4 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. 1B). A gene duplication event of the ancestral progenitor occurred approximately 580 My ago, before the evolution of tetrapods, producing GC, and a second copy (Gibbs et al., 1998; Haefliger et al., 1989). This second copy underwent a duplication event around 295 My ago, shortly after the divergence of amphibians and reptiles, producing ALB and a third copy that underwent a duplication event 250 MY ago, 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 tetrapods. However, ALB is differentially retained in fishes. While a broad survey has not been performed, individual studies have shown that ALB is absent in cartilaginous fishes, eel, Antarctic toothfish and carp, but is present in salmonid (salmon) species and lamprey (Metcalf et al., 2007; Noël et al., 2010). 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 certain ligands that make albumin family proteins essential for the distribution of these molecules in the body and proper physiological function (Fanali et al., 2012; Terentiev and Moldogazieva, 2013; Voegele et al., 2002; White and Cooke, 2000).

Serum albumin is the most well-studied and best-understood member of the family, being investigated for its physiological properties and potential biomedical and pharmaceutical applications (Fanali et al., 2012; Larsen et al., 2016). Serum albumin 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). Serum albumin is highly conserved across vertebrates — its size, shape, genetic sequence, and function as a transporter remains largely unchanged across tetrapod lineages (Noël et al., 2010). Serum albumin is the most abundant plasma protein in mammals, and also functions in the regulation of osmotic pressure and blood pH in most vertebrates (Fanali et al., 2012). Serum albumin has also been shown to be an important circulating antioxidant (Roche et al., 2008). The ability of serum albumin to bind to many different ligands is due to its multi-domain structure (Fanali et al., 2012). Serum albumin contains three domains, each containing two binding sites for a total of 6 distinct binding sites each with their own affinities for different ligands (Fasano et al., 2005). The 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 proteins are less well characterized, 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, 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 plays a major role in the activation of macrophages during immune and inflammatory responses in mammals. (Chun, 2012; White and Cooke, 2000)

Alpha-fetoprotein (AFP) is known to be expressed in the yolk of bird eggs and fetal tissue in humans (Bader et al., 2004; Cordeiro and Hincke, 2016). Expression of AFP in human fetuses peaks at 14 weeks of gestation and steadily decreases until almost ceasing at birth (Bader et al., 2004). Studies have shown that AFP is capable of binding and transporting hydrophobic ligands such as fatty acids, bilirubin, and estrogen, and 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). In mammals, the precursor to AFP underwent a duplication that gave rise to AFP and AFM, which is not found in other tetrapods. AFM is the most recently described member of the family and least understood albumin transport protein but is known to bind and transport vitamin E and Wnt proteins, important signaling molecules in many cellular and developmental pathways (Jerkovic et al., 2005; Naschberger et al., 2017).

Comparative studies on albumin proteins have focused on mammals due to their greater availability of whole-genome assemblies and importance in 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 variation in sauropsids, the sister groups of mammals, 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 identify the absence of canonical serum albumin in squamate reptiles. Our results indicate that different sauropsid lineages have diverse repertoires of albumin proteins derived from differential retention of ancestral genes. We also find a new clade of AFP in reptiles that share conserved AFP domains, yet are phylogenetically distinct from mammalian AFP. The importance of characterizing these proteins and their relationships in understudied lineages will lead to the discovery of novel sequences and functions that could have biomedical applications. Specifically, further investigation into the loss of serum albumin in squamates may offer insight into the human condition of analbuminemia and potential treatments (Koot et al., 2004).

Materials and Methods

Bioinformatic Searches

Ancestral tetrapod albumin sequences were retrieved from the spotted gar (*Lepisosteus oculatus*), zebrafish (*Danio rerio*), and the western clawed frog (*Xenopus tropicalis*) 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 performed bioinformatic searches for albumin sequences in several

sauropsid genomes, 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. For tuatara (*Sphenodon punctatus*), because it 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 e^-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.

Phylogenetic Analyses

To determine the orthology of albumin family sequences, we performed phylogenetic analyses using Maximum Likelihood (ML) and Bayesian Inference (BI) approaches. First, nucleotide and amino acid sequences were aligned using MAFFT (Katoh and Standley, 2013) and trimmed/cleaned with trimAL (Capella-Gutiérrez et al., 2009) . We performed ML analyses in RAxML v8 (Stamatakis, 2014) with 20 searches for the best tree using the GTR GAMMA model of nucleotide evolution and PROT GAMMA model of protein evolution. Nodal support for the best ML topology was assessed using non-parametric bootstrapping with the autoMRE option in RAxML, and we reconciled the best ML tree with the bootstrap replicates in RAxML. Bayesian Inference was performed in MrBayes v3.2 (Ronquist et al., 2012) using four independent runs with four chains each for 5x10^10 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 (Rambaut et al., 2018). We discarded trees collected before the chains reached convergence and summarized results with a majority-rule consensus of trees.

Synteny Search

Sequences and gene order were obtained from the genome assemblies of the mouse *Mus musculus* (GRCg6a/galGal6), from the chicken *Gallus gallus* (GRCm38/mm10), and from the 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 Figure 2.

Results

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 genes and used phylogeny 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. In this study, 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.

ALB is absent in squamates. In our sequence searches, we did not find canonical ALB for any snake or lizard, despite extensive annotation efforts. In contrast, serum albumin is present in all other sauropsids surveyed (Fig. 1A). We identified canonical serum albumin in fish, mammals, frogs, archosaurs (birds and alligators), and turtles. Many squamate sequences that had been assigned serum albumin annotations were actually more closely related to alpha-fetoprotein. Notably, the tuatara (*Sphenodon punctatus*), sister to squamate reptiles, has a partial (degraded) serum albumin sequence. This absence of canonical serum albumin 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 being present in *M. musculus* between AFP and RASSF6 (Figure 2). 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 (Figure 2).

Vitamin D-binding protein is highly conserved across reptiles. Vitamin D-binding protein (GC) is highly conserved across all lineages and reflects the expected species topology (Figure 1). 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 physiological important functions, including acting as the principal transporter of vitamin D and its metabolites, scavenging G-actin, and as an activator of 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 development and survival and is likely under strong purifying selection.

Alpha-fetoprotein is highly variable in reptiles. All protein sequences categorized as AFP were annotated with 3 characteristic ALB domains (PF00273) from the Pfam database. However, the number of reptile AFP fingerprints per sequence were highly variable (0-6), compared to canonical mammal AFP (7 fingerprints). Putative reptile AFP sequences were similarly variable in ProDom and PROSite annotations compared to mammals. Phylogenetic analysis shows at least two distinct clades of AFP genes: one found in mammals, and another that is reptile-specific (represented in Figure 1A 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 (Figure 1A). However, there also appears to be a squamate-specific clade, with all squamate AFP sequences forming a monophyletic clade that reflects the expected species topology (Figure 1A).

Only characteristic ALB residues are conserved across reptiles. Characteristic motifs and residues in the four albumin 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. This aspect of our results is still under investigation, and further results will be included in the final draft of this manuscript to be submitted for publication in a scientific journal.

Discussion

Given its physiological importance and conservation across vertebrates, ALB should be retained in all reptiles. Surprisingly, we did not detect serum albumin for any of the 15 squamate species spanning 8 families in this study, despite exhaustive annotation efforts. Most amniotes possess a full set of albumin proteins (GC, ALB, AFP, and AFM), several with an expanded repertoire. However, ALB is generally conserved across vertebrates. Thus, our findings suggest the loss or modification beyond recognition of serum albumin in squamates.

The clade Sauropsida consists of extant reptiles, a diverse set of organisms with a wide variety of adaptations that are largely understudied when compared to their mammalian counterparts. Within Sauropsids, squamates (lizards and snakes of the order Squamata) represent the largest order of Sauropsids and the second largest order of vertebrates, with over 10,000 described species. The loss of a major physiological protein such as serum albumin is highly surprising and presumably leads to major physiological shifts in order to account for this absence.

Although ALB serves vital physiological transport roles, there is a human medical condition characterized by the absence, or abnormally low levels, of ALB circulating in the blood serum known as congenital analbuminemia (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 mutation that causes premature truncations in the ALB molecule (Caridi et al., 2019; Minchiotti et al., 2019). As of 2019 there are 27 known genetic variants of the ALB gene that result in the pathogenic condition, including a variant of the start codon, frameshift/insertions, frameshift/deletions, nonsense mutations, and mutations affecting splicing (Caridi et al., 2019; Minchiotti et al., 2019). The majority of these variants are unique to each identified affected family, but one particular frameshift deletion called Kayseri is known to cause one third of all known cases (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 oligohydramnions and placental abnormalities, and increases the chance of death in early childhood, being associated with low birth weight, oedema, fluid retention, and lower respiratory tract infections (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). 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. Exciting future challenges involve identifying what compensatory mechanisms have evolved in snakes and lizards, and if these mechanisms could be used to develop biomedical therapies for congenital analbuminemia.

AFP has long been known as the fetal counterpart to albumin, but this function has only been characterized in mammals (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 reptiles the role of AFP in development as well as its expression and function in adulthood remains unknown. The lack of knowledge of the function of this protein in non-mammalian lineages limits our understanding of this protein and its potential uses. The divergence of reptilian and squamate AFP could indicate novel functions in these lineages that warrant rigorous investigation.

Comparative studies on reptiles, especially lizards and snakes, suffer from a dearth of genomic resources. Additionally, the resources available are incomplete and plagued with assembly artifacts (Hoffmann et al., 2018)). Future studies with more well-annotated genomes for reptiles could effectively explore the expansion and contraction of gene families more thoroughly. However, our study provides a comparative genomic framework for identifying any lineage-specific gene expansions that may compensate for the lack of the physiologically important serum albumin in squamates.

Acknowledgments

We thank Maheshi Dassanayake and Vinson Doyle 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 the National Science Foundation grant DEB-1146033 to CCA.

References

Abelev, G.I., Eraiser, T.L., 1999. Cellular aspects of alpha-fetoprotein reexpression in tumors. Semin. Cancer Biol. 9, 95–107.

Acids research, N., 2017, 2017. UniProt: the universal protein knowledgebase. Nucleic Acids Res. 45, D158–D169.

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, 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-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 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, 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, 862–867.

Curry, S., 2002. Beyond expansion: structural studies on the transport roles of human serum albumin. Vox Sang. 83 Suppl 1, 315–319.

307 Evans, T.W., 2002. albumin as a drug—biological effects of albumin unrelated to oncotic
308 pressure. Aliment. Pharmacol. Ther. 16, 6–11.
309 Fanali, G., di Masi, A., Trezza, V., Marino, M., Fasano, M., Ascenzi, P., 2012. Human serum
310 albumin: from bench to bedside. Mol. Aspects Med. 33, 209–290.
311 Fasano, M., Curry, S., Terreno, E., Galliano, M., Fanali, G., Narciso, P., Notari, S., Ascenzi, P.,
312 2005. The extraordinary ligand binding properties of human serum albumin. IUBMB Life
313 57, 787–796.
314 Gibbs, P.E., Witke, W.F., Dugaiczyk, A., 1998. The molecular clock runs at different rates
315 among closely related members of a gene family. J. Mol. Evol. 46, 552–561.
316 Haefliger, D.N., Moskaitis, J.E., Schoenberg, D.R., Wahli, W., 1989. Amphibian albumins as
317 members of the albumin, alpha-fetoprotein, vitamin D-binding protein multigene family. J.
318 Mol. Evol. 29, 344–354.
319 Hoffmann, F.G., Vandewege, M.W., Storz, J.F., Opazo, J.C., 2018. Gene Turnover and
320 Diversification of the α- and β-Globin Gene Families in Sauropsid Vertebrates. Genome
321 Biol. Evol. 10, 344–358.
322 Jerkovic, L., Voegele, A.F., Chwatal, S., Kronenberg, F., Radcliffe, C.M., Wormald, M.R.,
323 Lobentanz, E.M., Ezeh, B., Eller, P., Dejori, N., Dieplinger, B., Lottspeich, F., Sattler, W.,
324 Uhr, M., Mechtler, K., Dwek, R.A., Rudd, P.M., Baier, G., Dieplinger, H., 2005. Afamin is
325 a novel human vitamin E-binding glycoprotein characterization and in vitro expression. J.
326 Proteome Res. 4, 889–899.
327 Jones, P., Binns, D., Chang, H.-Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J.,
328 Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M.,
329 Yong, S.-Y., Lopez, R., Hunter, S., 2014. InterProScan 5: genome-scale protein function

classification. Bioinformatics 30, 1236–1240.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 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, 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.

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, 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, 190–200.

Nishio, H., Heiskanen, M., Palotie, A., Bélanger, L., Dugaiczyk, A., 1996. Tandem arrangement

353 of the human serum albumin multigene family in the sub-centromeric region of 4q:
354 evolution and chromosomal direction of transcription. J. Mol. Biol. 259, 113–119.
355 Noël, E.S., Reis, M. dos, Arain, Z., Ober, E.A., 2010. Analysis of the Albumin/α-
356 Fetoprotein/Afamin/Group specific component gene family in the context of zebrafish liver
357 differentiation. Gene Expr. Patterns 10, 237–243.
358 Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A., 2018. Posterior
359 Summarization in Bayesian Phylogenetics Using Tracer 1.7. Syst. Biol. 67, 901–904.
360 Roche, M., Rondeau, P., Singh, N.R., Tarnus, E., Bourdon, E., 2008. The antioxidant properties
361 of serum albumin. FEBS Lett. 582, 1783–1787.
362 Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B.,
363 Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian
364 phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–
365 542.
366 Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
367 large phylogenies. Bioinformatics 30, 1312–1313.
368 Terentiev, A.A., Moldogazieva, N.T., 2013. Alpha-fetoprotein: a renaissance. Tumour Biol. 34,
369 2075–2091.
370 Tzika, A.C., Ullate-Agote, A., Grbic, D., Milinkovitch, M.C., 2015. Reptilian Transcriptomes
371 v2.0: An Extensive Resource for Sauropsida Genomics and Transcriptomics. Genome Biol.
372 Evol. 7, 1827–1841.
373 Voegele, A.F., Jerković, L., Wellenzohn, B., Eller, P., Kronenberg, F., Liedl, K.R., Dieplinger,
374 H., 2002. Characterization of the vitamin E-binding properties of human plasma afamin.
375 Biochemistry 41, 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. U. S. A. 91, 9417–9421.

White, P., Cooke, N., 2000. The multifunctional properties and characteristics of vitamin D-binding protein. Trends Endocrinol. Metab. 11, 320–327.

FIGURE CAPTIONS

Figure 1.

(A.) Bayesian phylogeny of amino acid sequences. Values at nodes indicate posterior probabilities. Protein clades are labeled with the protein name and a colored bar. Colors correspond to each family protein: green for GC, red for ALB, yellow for AFP, purple for AFP-R, and blue for AFM. AFP-R represents reptilian AFP, a reptile specific AFP distinct from mammal AFP. (B.) Representation of the accepted relationships of the albumin family members. (C.) The accepted vertebrate phylogeny with a corresponding matrix showing the distribution of albumin family members in each lineage. Lineages that comprise sauropsida are highlighted.

Figure 2. Synteny of the albumin gene family chromosomal regions in the genomes of house mouse (*Mus musculus*), chicken (*Gallus gallus*), and green anole (*Anolis carolinensis*). Members of the albumin gene family are shown in black, while neighboring genes are shown in white. The highlighted blue portion of each chromosome corresponds to the 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 represents a distance of 4 million base pairs.

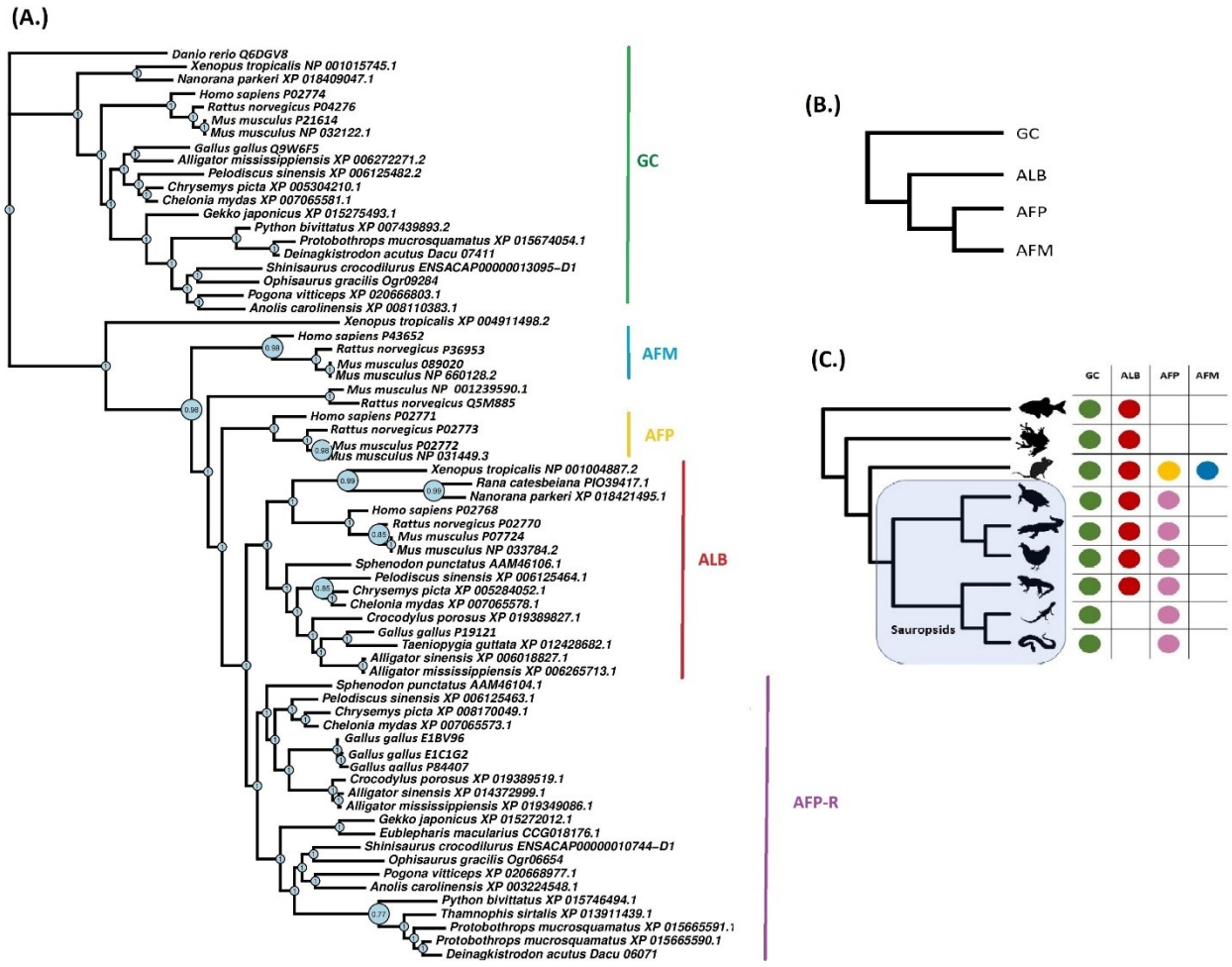
(A.)
Danio rerio Q6DGV8
Xenopus tropicalis NP 001015745.1
Nanorana parkeri XP 018409047.1
Homo sapiens P02774
Rattus norvegicus P04276
Mus musculus P21614
Mus musculus NP 032122.1
Gallus gallus Q9W6F5
Alligator mississippiensis XP 006272271.2
Pelodiscus sinensis XP 006125482.2
Chrysemys picta XP 005304210.1
Chelonia mydas XP 007065581.1
Gekko japonicus XP 015275493.1
Python bivittatus XP 007439893.2
Protobothrops mucrosquamatus XP 015674054.1
Deinagkistrodon acutus Dacu 07411
Shinisaurus crocodilurus ENSACAP00000013095-D1
Ophisaurus gracilis Ogr09284
Pogona vitticeps XP 020666803.1
Anolis carolinensis XP 008110383.1
Xenopus tropicalis XP 004911498.2
Homo sapiens P43652
Rattus norvegicus P36953
Mus musculus 089020
Mus musculus NP 660128.2
Mus musculus NP 001239590.1
Rattus norvegicus Q5M885
Homo sapiens P02771
Rattus norvegicus P02773
Mus musculus P02772
Mus musculus NP 031449.3
Xenopus tropicalis NP 001004887.2
Rana catesbeiana PIO39417.1
Nanorana parkeri XP 018421495.1
Homo sapiens P02768
Rattus norvegicus P02770
Mus musculus P07724
Mus musculus NP 033784.2
Sphenodon punctatus AAM46106.1
Pelodiscus sinensis XP 006125464.1
Chrysemys picta XP 005284052.1
Chelonia mydas XP 007065578.1
Crocodylus porosus XP 019389827.1
Gallus gallus P19121
Taeniopygia guttata XP 012428682.1
Alligator sinensis XP 006018827.1
Alligator mississippiensis XP 006265713.1
Sphenodon punctatus AAM46104.1
Pelodiscus sinensis XP 006125463.1
Chrysemys picta XP 008170049.1
Chelonia mydas XP 007065573.1
Gallus gallus E1BV96
Gallus gallus E1C1G2
Gallus gallus P84407
Crocodylus porosus XP 019389519.1
Alligator sinensis XP 014372999.1
Alligator mississippiensis XP 019349086.1
Gekko japonicus XP 015272012.1
Eublepharis macularius CCG018176.1
Shinisaurus crocodilurus ENSACAP00000010744-D1
Ophisaurus gracilis Ogr06654
Pogona vitticeps XP 020668977.1
Anolis carolinensis XP 003224548.1
Python bivittatus XP 015746494.1
Thamnophis sirtalis XP 013911439.1
Protobothrops mucrosquamatus XP 015665591.1
Protobothrops mucrosquamatus XP 015665590.1
Deinagkistrodon acutus Dacu 06071
GC
AFM
AFP
ALB
AFP-R
(B.)
GC
ALB
AFP
AFM
(C.)
GC
ALB
AFP
AFM
Sauropsids

Figure 1.

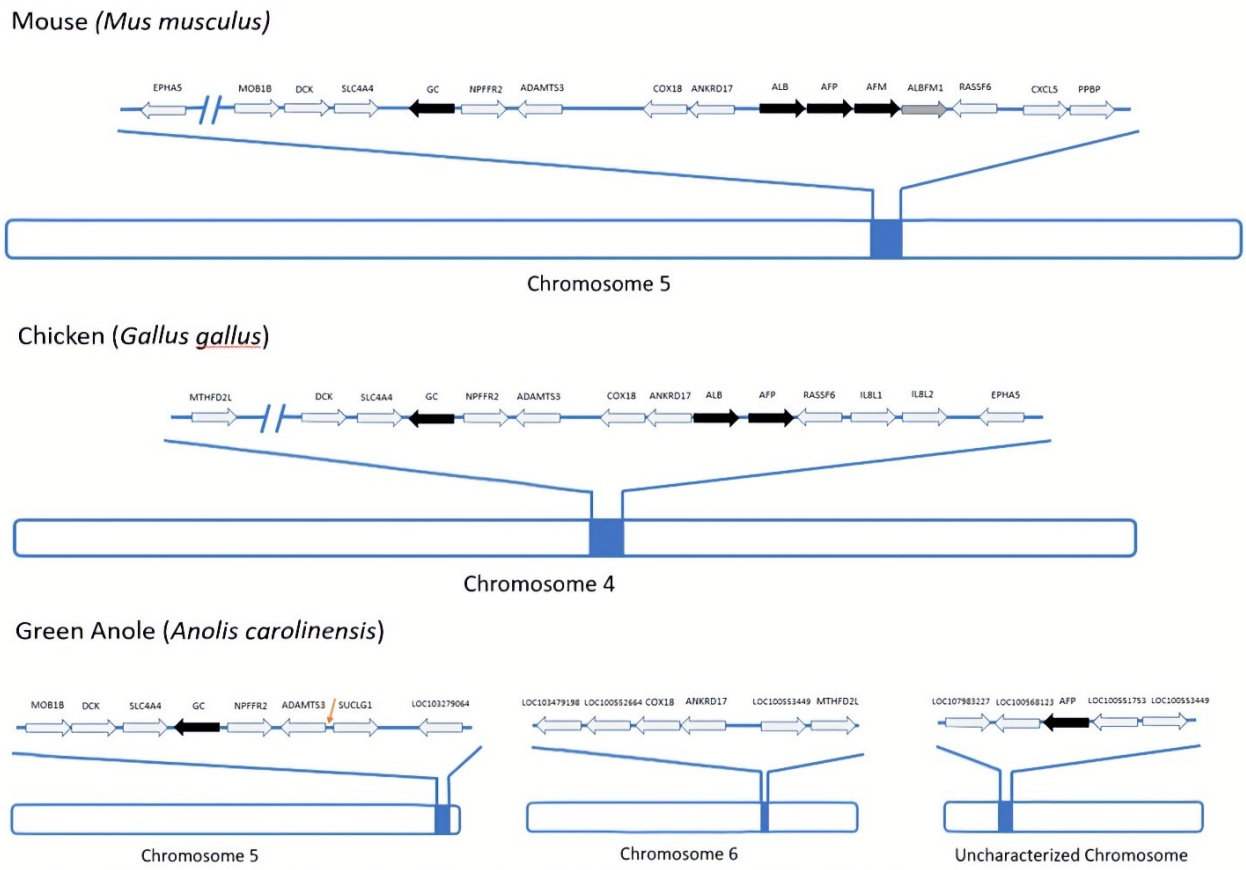
Mouse (Mus musculus)
EPHA5
MOB1B
DCK
SLC4A4
GC
NPFFR2
ADAMTS3
COX18
ANKRD17
ALB
AFP
AFM
ALBFM1
RASSF6
CXCL5
PPBP
Chromosome 5
Chicken (Gallus gallus)
MTHFD2L
DCK
SLC4A4
GC
NPFFR2
ADAMTS3
COX18
ANKRD17
ALB
AFP
RASSF6
ILBL1
ILBL2
EPHA5
Chromosome 4
Green Anole (Anolis carolinensis)
MOB1B
DCK
SLC4A4
GC
NPFFR2
ADAMTS3
SUCLG1
LOC103279064
Chromosome 5
LOC103479198
LOC100552664
COX18
ANKRD17
LOC100553449
MTHFD2L
Chromosome 6
LOC107983227
LOC100568123
AFP
LOC100551753
LOC100553449
Uncharacterized Chromosome

Figure 2.

Supplementary Table 1. Gene list. Protein name, gene ID, organism, and source sequence was retrieved for all amino acid sequences used in bioinformatic and phylogenetic analysis.

Higher Taxonomy	Genus	species	Protein Name	ID	Source
Actinopterygii	Danio	rerio	GC	Q6DGV8	Phylome
Amphibia	Xenopus	tropicalis	GC	NP_001015745.1	NCBI
Amphibia	Xenopus	tropicalis	ALB	NP_001004887.2	Refseq
Amphibia	Xenopus	tropicalis	Ancestral albumin	XP_004911498.2	NCBI
Amphibia	Nanorana	parkeri	GC	XP_018409047.1	NCBI
Amphibia	Nanorana	parkeri	ALB	XP_018421495.1	NCBI
Amphibia	Rana	catesbeiana	ALB	PIO39417.1	NCBI
Aves	Gallus	gallus	GC	Q9W6F5	TrEMBL
Aves	Gallus	gallus	ALB	P19121	SwissProt
Aves	Gallus	gallus	AFP	E1BV96	Phylome
Aves	Gallus	gallus	AFP	E1C1G2	Phylome
Aves	Gallus	gallus	AFP	P84407	SwissProt
Crocodilia	Alligator	mississippiensis	GC	XP_006272271.2	TrEMBL
Crocodilia	Alligator	mississippiensis	ALB	XP_006265713.1	TrEMBL
Crocodilia	Alligator	mississippiensis	AFP	XP_019349086.1	TrEMBL
Crocodilia	Alligator	sinensis	ALB	XP_006018827.1	TrEMBL
Crocodilia	Alligator	sinensis	AFP	XP_014372999.1	TrEMBL
Crocodilia	Crocodylus	porosus	ALB	XP_019389827.1	Blastp
Crocodilia	Crocodylus	porosus	AFP	XP_019389519.1	Blastp
Tetsudinae	Chrysemys	picta	GC	XP_005304210.1	NCBI
Mammalia	Homo	sapiens	GC	P02774	SwissProt
Mammalia	Rattus	norvegicus	GC	P04276	Phylome
Mammalia	Mus	musculus	GC	P21614	SwissProt
Mammalia	Mus	musculus	GC	NP_032122.1	NCBI
Tetsudinae	Pelodiscus	sinensis	GC	XP_006125482.2	TrEMBL
Tetsudinae	Chelonia	mydas	GC	XP_007065581.1	NCBI
Squamata	Gekko	japonicus	GC	XP_015275493.1	NCBI
Squamata	Python	bivittatus	GC	XP_007439893.2	NCBI
Squamata	Protobothrops	muscrosquamatus	GC	XP_015674054.1	NCBI

Squamata	Deinagkistrodon	acutus	GC	Dacu_07411	Blastp
Squamata	Shinisaurus	crocodilurus	GC	ENSACAP00000013095-D1	Blastp
Squamata	Ophisaurus	gracilis	GC	Ogr09284	Blastp
Squamata	Pogona	vitticeps	GC	XP_020666803.1	NCBI
Squamata	Anolis	carolinensis	GC	XP_008110383.1	TrEMBL
Mammalia	Homo	sapiens	AFM	P43652	SwissProt
Mammalia	Rattus	norvegicus	AFM	P36953	Phylome
Mammalia	Mus	musculus	AFM	89020	SwissProt
Mammalia	Mus	musculus	AFM	NP_660128.2	SwissProt
Mammalia	Mus	musculus	ALBFM1	NP_001239590.1	Phylome
Mammalia	Rattus	norvegicus	ALBFM1	Q5M885	Phylome
Mammalia	Homo	sapiens	AFP	P02771	SwissProt
Mammalia	Rattus	norvegicus	AFP	P02773	Phylome
Mammalia	Mus	musculus	AFP	P02772	SwissProt
Mammalia	Mus	musculus	AFP	NP_031449.3	Phylome
Mammalia	Homo	sapiens	ALB	P02768	SwissProt
Mammalia	Rattus	norvegicus	ALB	P02770	Phylome
Mammalia	Mus	musculus	ALB	P07724	SwissProt
Mammalia	Mus	musculus	ALB	NP_033784.2	Phylome
Squamata	Sphenodon	punctatus	ALB	AAM46106.1	NCBI
Tetsudinae	Pelodiscus	sinensis	ALB	XP_006125464.1	Blastp
Tetsudinae	Chrysemys	picta	ALB	XP_005284052.1	Blastp
Tetsudinae	Chelonia	mydas	ALB	XP_007065578.1	NCBI
Aves	Taeniopygia	guttata	ALB	XP_012428682.1	NCBI
Rhynchocephalia	Sphenodon	punctatus	AFP	AAM46104.1	NCBI
Squamata	Pelodiscus	sinensis	AFP	XP_006125463.1	NCBI
Tetsudinae	Chrysemys	picta	AFP	XP_008170049.1	NCBI
Tetsudinae	Chelonia	mydas	AFP	XP_007065573.1	NCBI
Squamata	Gekko	japonicus	AFP	XP_015272012.1	NCBI
Squamata	Eublepharis	macularius	AFP	CCG018176.1	Blastp
Squamata	Shinisaurus	crocodilurus	AFP	ENSACAP00000010744-D1	Blastp

Squamata	Ophisaurus	gracilis	AFP	Ogr06654	Blastp
Squamata	Pogona	vitticeps	AFP	XP_020668977.1	NCBI
Squamata	Anolis	carolinensis	AFP	XP_003224548.1	NCBI
Squamata	Python	bivittatus	AFP	XP_015746494.1	NCBI
Squamata	Thmanophis	sirtalis	AFP	XP_013911439.1	Blastp
Squamata	Protobothrops	muscrosquamatus	AFP	XP_015665591.1	NCBI
Squamata	Protobothrops	muscrosquamatus	AFP	XP_015665590.1	NCBI
Squamata	Deinagkistrodon	acutus	AFP	DACU_06071	Blastp