

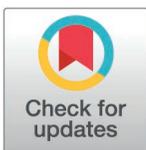
## RESEARCH ARTICLE

## Deep-time gene expression shift reveals an ancient change in avian muscle phenotypes

Christina M. Harvey<sup>1\*</sup>, Matthew J. Fuxjager<sup>2</sup>, James B. Pease<sup>3\*</sup>

**1** Department of Biology, Wake Forest University, Winston-Salem, North Carolina, United States of America, **2** Department of Ecology, Evolution, and Organismal Biology, Brown University, Providence, Rhode Island, United States of America, **3** Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, Columbus, Ohio, United States of America

\* [pease.25@osu.edu](mailto:pease.25@osu.edu) (JBP), [keatcm21@wfu.edu](mailto:keatcm21@wfu.edu) (CMH)



## Abstract

Gene duplication is an important process of molecular evolutionary change, though identifying these events and their functional implications remains challenging. Studies on gene duplication more often focus on the presence of paralogous genes within the genomes and less frequently explore shifts in expression. We investigated the evolutionary history of calsequestrin (CASQ), a crucial calcium-binding protein in the junctional sarcoplasmic reticulum of muscle tissues. CASQ exists in jawed vertebrates as subfunctionalized paralogs CASQ1 and CASQ2 expressed primarily in skeletal and cardiac muscles, respectively. We used an enhanced sequence dataset to support initial duplication of CASQ1 in a jawed fish ancestor prior to the divergence of cartilaginous fishes. Surprisingly, we find CASQ2 is the predominant skeletal muscle paralog in birds, while CASQ1 is either absent or effectively nonfunctional. Changes in the amino acid composition and electronegativity of avian CASQ2 suggest enhancement to calcium-binding properties that preceded the loss of CASQ1. We identify this phenomenon as CASQ2 “synfunctionalization,” where one paralog functionally replaces another. While additional studies are needed to fully understand the dynamics of CASQ1 and CASQ2 in bird muscles, the long and consistent history of CASQ subfunctions outside of birds indicate a substantial evolutionary pressure on calcium-cycling processes in muscle tissues, likely connected to increased avian cardiovascular and metabolic demands. Our study provides an important insight into the molecular evolution of birds and shows how gene expression patterns can be comparatively studied across phylum-scale deep time to reveal key evolutionary events.

## OPEN ACCESS

**Citation:** Harvey CM, Fuxjager MJ, Pease JB (2025) Deep-time gene expression shift reveals an ancient change in avian muscle phenotypes. *PLoS Genet* 21(4): e1011663. <https://doi.org/10.1371/journal.pgen.1011663>

**Editor:** Ingo Braasch, Michigan State University, UNITED STATES OF AMERICA

**Received:** September 6, 2024

**Accepted:** March 22, 2025

**Published:** April 11, 2025

**Copyright:** © 2025 Harvey et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data availability statement:** Original Data is available on NCBI SRA, Protein, Nucleotide databases as referenced through accession numbers in [S3](#) and [S4 Data](#) files. All other relevant data are in the manuscript and its [supporting information](#) files.

**Funding:** This work was supported by National Science Foundation grants NSF-DEB#2217117 to JBP and NSF-OISE#1952542 to MJF ([www.nsf.gov](http://www.nsf.gov)). JBP received salary support from

## Author Summary

Duplicated genes adopt different functional relationships that can directly influence physiological processes. Calsequestrin-1 and calsequestrin-2 are two similar proteins expressed from genes that bind calcium needed for strong muscle contractions in vertebrate muscle tissue. We compared gene sequences and expression across all chordates and found that CASQ1 and CASQ2 have specialized functions in the skeletal muscles and heart, respectively, in most vertebrates. However, birds only express CASQ2 in both the

NSF-DEB#2217117 and MJF received salary support from NSF-OISE#1952542. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

heart and skeletal muscles, indicating a shift in muscle function on the molecular level. Furthermore, CASQ1 had its gene silenced or deleted in birds. Our analysis discovered an example of a rarely reported gene family functional change and influences our current understanding of avian evolution and muscle physiology.

## Introduction

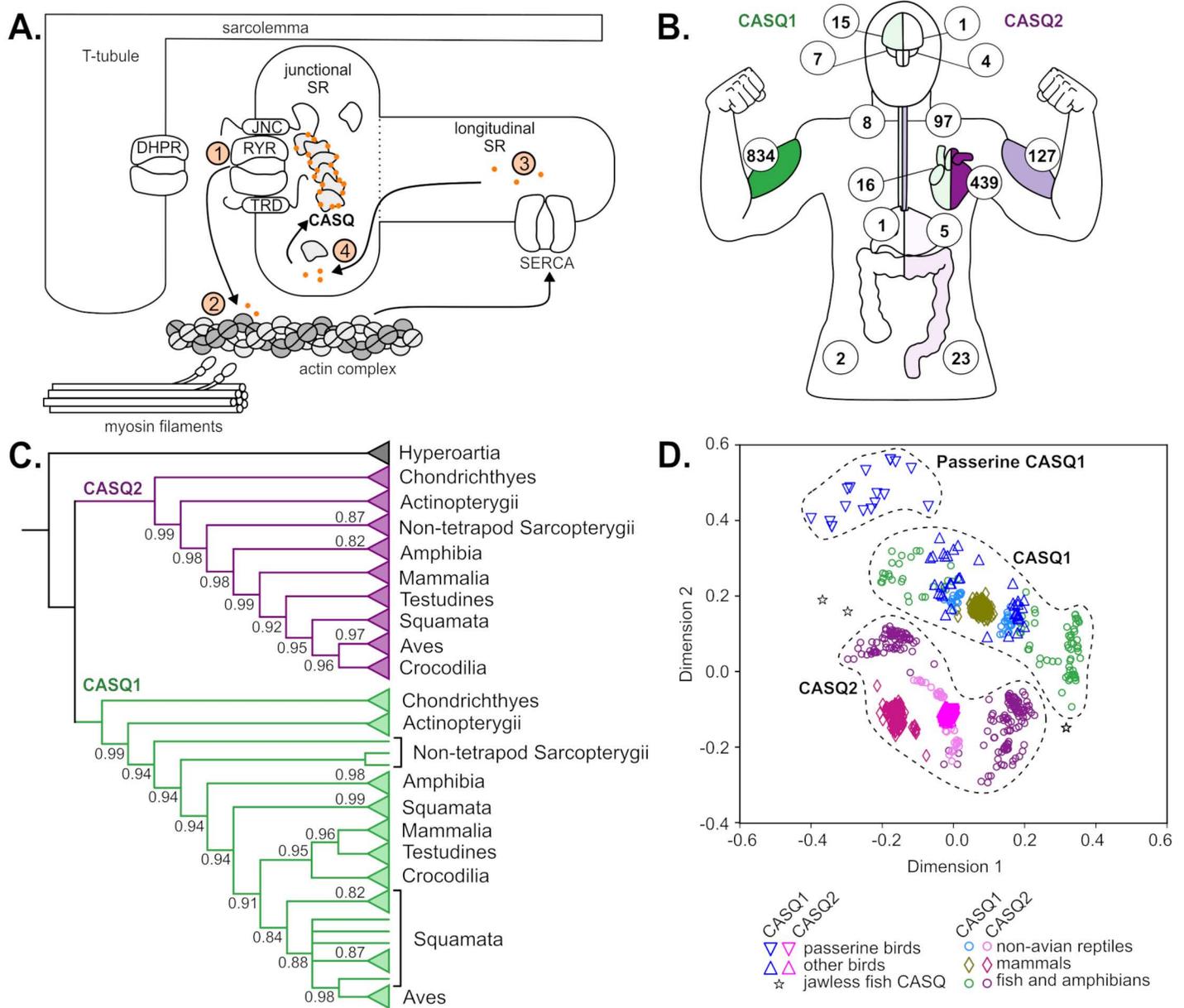
Gene duplication, loss, and dosage changes are fundamental processes of genome evolution with the potential to cause major shifts in traits and organismal functions [1–4]. Duplicated genes undergo a variety of changes in their functional roles, including specialization (subfunctionalization), evolution to new functions (neofunctionalization), and compensatory reactions to gene loss [4–7]. Duplicated genes also modulate expression as dosage balance, compensatory drift, and hypofunctionalization [6,7]. Investigations of both the presence and absence of gene duplicates and their relative expression are therefore crucial to understand the impact of these molecular evolutionary processes on key fitness-related phenotypes.

Calsequestrin (CASQ) proteins are the most abundant  $\text{Ca}^{2+}$ -binding proteins in the junctional sarcoplasmic reticulum (jSR) of vertebrate cardiac and skeletal myocytes, forming compact, high-capacity polymers when  $\text{Ca}^{2+}$  concentrations are high (Fig 1A) [8–13]. While CASQ is one of many components in myocyte calcium cycling, even minor CASQ mutations have been shown to cause lethal stress-induced myopathies in skeletal and cardiac tissues in humans, mice, and chickens [14–18]. Evidence from these and other studies directly demonstrates that calsequestrin is critical to core vertebrate muscle contraction necessary for maintaining high muscle performance [19].

Two subfunctionalized CASQ proteins are present in most vertebrates: skeletal muscle-specific CASQ1 and cardiac muscle-specific CASQ2 (Fig 1B) [20–23]. In mammals, smooth muscles (e.g., the esophagus) may also express low levels of CASQ2 and the cerebellum expresses a small amount of CASQ1 [24,25]. Mammalian CASQ1 and CASQ2 exhibit different calcium-binding capabilities, calcium-sensing thresholds, and proximal interactions with calcium-release channels [14,19,21,26,27].

Previous studies have disagreed on the evolutionary timing of CASQ gene duplication and debated the presence of CASQ1 in cartilaginous fishes, crocodylians, and birds [13,22,28]. In recent studies of molecular expression in Passerine bird muscles, we explored patterns of gene expression in core muscle function genes [29,30]. CASQ1 was among the candidate genes targeted in a PCR amplification study from zebra finch and whole transcriptome analyses of six manakins and a flycatcher, but in neither case could CASQ1 be found in the genome or transcriptomes of these species, respectively [29,30]. The absence of CASQ1 was puzzling because experimental deletion or mutation of CASQ1 has been shown to alter muscle structure and function in mammals, and presumably the loss of CASQ1 in other vertebrates would imply significant functional changes in skeletal muscle calcium-cycling processes [31,32].

Synfunctionalization is a rarely reported evolutionary process where one gene takes over the functional role of another paralogous gene [33,34]. The synfunctionalization event must necessarily be preceded by sub- or neo-functionalization where paralogs have subdivided or expanded their functional roles into separate spaces, such that one paralog can assume the function of the other later. After synfunctionalization, a paralog made redundant could later experience pseudogenization or deletion without selective consequence for the traits influenced [33,34]. The maintenance of high muscle performance in Passerine birds despite the



**Fig 1. Calsequestrin duplicated in the jawed vertebrate ancestor into two distinct proteins.** A) Calsequestrins bind  $Ca^{2+}$  ions (orange dots) in the junctional sarcoplasmic reticulum, either in monomeric, dimeric, or polymerized (as shown; DHPR: dihydropyridine receptor; JNC: junctin; TRD: triadin; RYR: ryanodine receptor; SR: sarcoplasmic reticulum; SERCA; sarcoendoplasmic reticulum ATPase). B) Average CASQ1 and CASQ2 expression in human tissues from the Genotype-Tissue Expression (GTEx) dataset. Expression of tissue-specific paralogs is represented in transcripts per million reads (TPM) and by the saturation of green (left, CASQ1) or purple (right, CASQ2) relative to blue (no CASQ expression). Values (from top to bottom) depict brain cortex, cerebellum, esophagus muscularis, heart, stomach, skeletal muscle, and transverse colon. C) Gene family cladogram of preduplication CASQ (black), CASQ1 (green), and CASQ2 (purple) amino acid sequences supports a duplication in a jawed vertebrate ancestor. Polytomies were created for branches with a support value of 0.8 or lower. Bootstrap values of 1.0 are not shown. Triangles depict collapsed clades. D) A multidimensional scaling plot of calsequestrin amino acid pairwise sequence distances shows separate, non-intercalated CASQ1 and CASQ2 clusters.

<https://doi.org/10.1371/journal.pgen.1011663.g001>

apparent loss of genetic inactivation of CASQ1 led us to hypothesize the potential for synfunctionalization in this case. Observing additional cases of synfunctionalization, especially in a core protein family with the functional impact of calsequestrin, would greatly contribute to our understanding of this gene family’s evolutionary process [34].

In this study, we conduct a deep-time analysis spanning a comprehensive set of chordate calsequestrin sequences and relative expression levels from hundreds of muscle expression profiles. Our results support the presence of *CASQ1* and *CASQ2* genes in all jawed vertebrate groups and a highly conserved pattern of *CASQ1* expression in skeletal muscles and *CASQ2* in cardiac muscles. Birds are an exception to this pattern by expressing a functionally enhanced *CASQ2* as their primary jSR calcium binder in both cardiac and skeletal muscles. We confirm that avian *CASQ1* genes exist, but sequence analysis indicates they are functionally inadequate, pseudogenized, or deleted. We conclude that the combined sequence and expression evidence show *CASQ2* has displaced its paralog *CASQ1* in an avian ancestor, under a model of paralog synfunctionalization. We also discuss the implications of *CASQ* synfunctionalization in the context of underlying evolutionary processes and in relation to changes in cell function and physiology in avian ancestors. Our results reinforce that deep-time comparative expression analysis at the class or phylum scope can reveal a markedly different sequence of evolutionary steps than gene presence-absence analyses alone and yield valuable functional insights [35–39]. More broadly, our results show a complex series of evolutionary events that integrate genomic, transcriptomic, structural, and organismal biology connecting changes in gene duplications and gene family expression to core animal physiological traits.

## Results and discussion

### Calsequestrin duplicated in the jawed vertebrate ancestor

We confirmed that calsequestrin is an ancient metazoan protein, distantly related to proteins with calcium-related functions in the endoplasmic reticulum. We identified 1248 calsequestrin genes (1137 chordate, 111 others) from 770 animal species (679 chordate, 91 others), including chordates, placozoans, hydra and corals, nematodes, arachnids, a sponge, and a tardigrade (S1 Fig, and Tables A and B in S1 Data). A phylogeny inferred from 1071 chordate *CASQ* sequences shows *CASQ1* and *CASQ2* are each monophyletic, cluster separately in a protein sequence-based multidimensional scaling plot (MDS) and are roughly equivalent amino acid distance from the jawless fish single *CASQ* sequences (Fig 1C and 1D). These features are consistent with a model of a single duplication event generating *CASQ1* and *CASQ2* as separate paralogs that each have remained orthologous through all extant vertebrates. All bony fish have at least three or four *CASQ* paralogs and salmonids have eight, congruent with inferred whole-genome duplication (WGD) events in their evolutionary history [40,41]. Another phylogeny from 111 non-chordate *CASQ*-like sequences indicates no duplications outside chordates (S2 Fig).

The presence of a single calsequestrin in all jawless chordates and two copies in all non-avian jawed vertebrates establish evidence that the most likely duplication timing of *CASQ* is in a gnathostome ancestor, potentially synchronous with the “2R” WGD event (Fig 1C) [42]. *CASQ1* and *CASQ2* both have 11 exons with nearly identical exon lengths and intron-exon boundaries for nearly all annotated genes, which further supports a DNA-based duplication. Additionally, *CASQ1* and *CASQ2* are each closely physically linked to several other genes that also exist in a few copies in gnathostomes but only a single copy in jawless chordates (*VANGL1/VANGL2*, *NHLH1/NHLH2*, *DCAF6/DCAF8*, and *ATP1A1/ATP1A2*). The presence of multiple linked duplicated genes in proximity to each copy of calsequestrin is further suggestive that calsequestrin was duplicated as part of a large-scale duplication or WGD event. However, *CASQ* remains chromosomally unplaced in lampreys and its neighboring genes in tunicates, lancelets, and hagfish are disparately located in gnathostome genomes. *CASQ1* is not placed in a clear chromosomal duplication block in reconstructions of the ancestral vertebrate linkage groups, preventing a more definitive phylogenetic placement of its duplication [43–45].

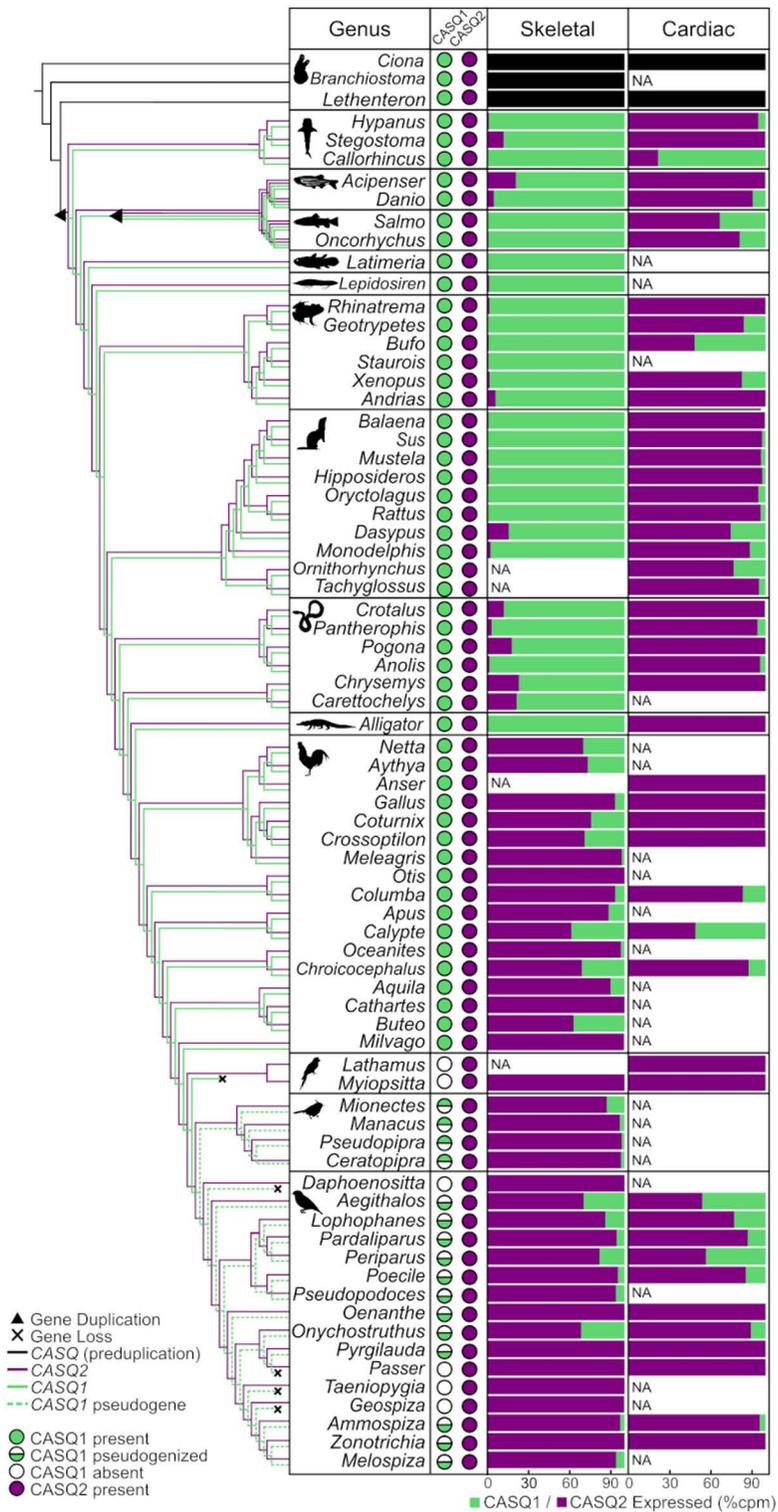
## Cardiac calsequestrin has replaced skeletal muscle calsequestrin in birds

CASQ expression levels from 414 RNA-seq profiles from 93 diverse chordate species show that skeletal muscles predominantly expressed CASQ1 and cardiac tissues CASQ2 in all jawed vertebrate lineages, except birds (Fig 2). The surprising exception is that all bird transcriptomes we sampled expressed CASQ2 as the major form of calsequestrin in both skeletal and cardiac muscle tissue. CASQ1 transcripts are present in many bird species, though the average skeletal muscle expression of CASQ1 in those species is low (83 counts per million; cpm). Levels of CASQ2 expression in skeletal muscles in birds are robust (up to 6322 cpm) and 2.5 times higher than CASQ2 levels in cardiac tissue, on average (Tables C and D in S1 Data). For comparison, the average mammalian CASQ1 expression is 3836 cpm in skeletal muscle tissue, representing over 99% of the total calsequestrin expressed in those tissues. We also confirmed that other calcium binding proteins found commonly in the SR do not noticeably increase or decrease in expression in birds compared to other tetrapods (e.g., CALM, HRC, CALR, PVALB). Furthermore, birds, non-avian reptiles, and mammals all express ryanodine receptor-1 in skeletal muscles and ryanodine receptor-2 in cardiac muscle, meaning other key molecular components of the jSR do not appear to have noticeably changed their expression patterns.

The expression of CASQ2 in chicken skeletal muscles was noted in a few early protein assays of CASQ [24], but we establish here that CASQ2 expression in skeletal muscles is consistent across all 42 bird species sampled. The finding that birds switched their calsequestrin is surprising because expression specificity of CASQ1 in skeletal muscle and CASQ2 in cardiac muscle has been consistent across all other lineages in the entire ~450 Myr evolutionary history of jawed vertebrates (Fig 2). The high conservation of sequence, clear split in sequence identity between CASQ1 and CASQ2, and lack of variation in expression patterns of these paralogs across ~300 Myr of evolutionary history between the jawed vertebrate and avian ancestors all indicate an ancient subfunctionalization that is not genetically labile. Based on these facts, we conclude that the switch in use of CASQ2 in the skeletal muscle in birds is a biologically significant molecular evolutionary event.

## Calsequestrin-1 is present in many (but not all) bird genomes

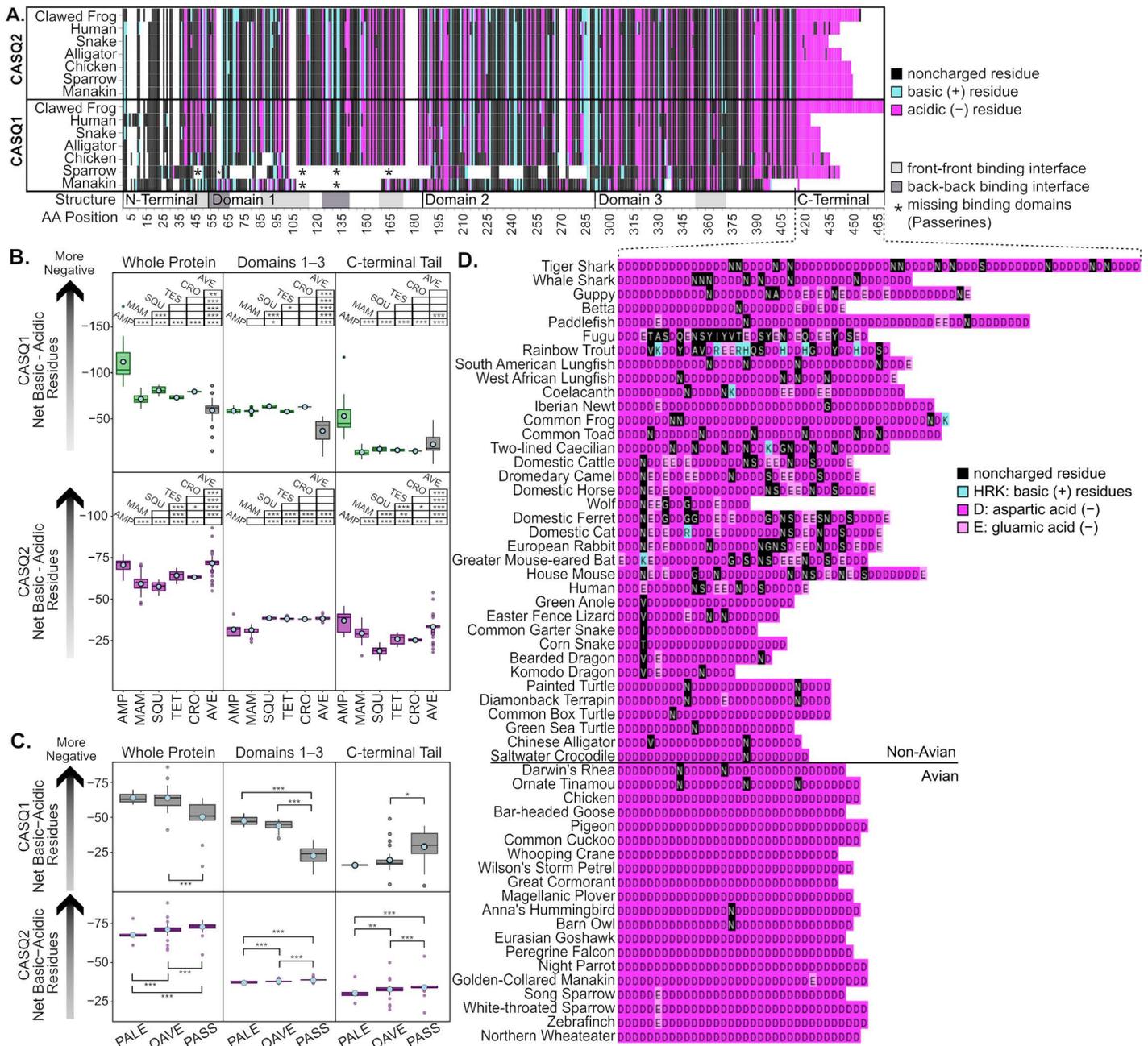
The switch to CASQ2 expression in avian skeletal muscles is furthermore surprising because we find that a full-length CASQ1 gene with no obvious mutational defects is present in the genomes of many bird species and is sometimes even expressed at low levels (Fig 2). We also find CASQ1 is present in all cartilaginous fish and crocodilian genomes that were previously reported as gene losses. We found these CASQ1 sequences through intensive surveys of new and updated genome assemblies and *de novo* assembly of transcriptomes. CASQ1 and CASQ2 each have the same immediately neighboring genes in all tetrapod genomes surveyed, and highly repetitive content of CASQ gene sequences made gene order often more reliable than sequence similarity searches to locate CASQ1 in many bird genomes. We determined that all birds have a CASQ1 gene present in their genome, except Psittaciformes (parrots) and some Passeriformes (perching birds). Multiple complete Psittaciformes genome assemblies show a complete deletion of CASQ1 between its two syntenic neighboring genes (S3 Fig). In Passeriformes, CASQ1 has undergone multiple independent partial or complete deletion events. Both genomic annotation and *de novo* assembly show that exons 2 and 3 are deleted in suboscines, while exons 4 and 5 are deleted in Passeridae (Fig 3A). Other Passerine groups (including zebra finch) appear to have complete deletion of CASQ1. Reports of CASQ1 loss in chicken are likely due to inconsistent assembly at the CASQ1 locus on microchromosome 25. Vocal learners exhibit high rates of genomic reorganization and nucleotide loss at the CASQ1



**Fig 2. Deep-time comparison of CASQ1 and CASQ2 relative expression shows a shift in birds.** RNA expression values of pre-duplication CASQ (black), CASQ1 (green), and CASQ2 (purple) as the percentage of counts per million reads out of the total calsequestrin expression for the skeletal muscle (left column) and cardiac muscle (right column). “NA” is shown for groups where RNA expression data for the given tissue sample was not available. Circle icons indicate the

presence (filled), absence (empty), or functionally pseudogenization (half-filled) for each gene. Triangles on the parallel cladograms for CASQ1 (green) and CASQ2 (purple) indicate gene duplication events, and X's indicate gene loss.

<https://doi.org/10.1371/journal.pgen.1011663.g002>



**Fig 3. Calsequestrin conservation, amino acid composition, and expression differs across tetrapod groups.** A) Amino acid composition and conservation of CASQ2 and CASQ1. The bottom track denotes structurally important binding sites as described in the literature. The tracks above depict acidic residues (cyan) and positive residues (magenta) for 7 species representing major tetrapod clades. Black and white represent non-charged residues and gaps, respectively. B) Relative charge for the entire protein, the structural domains, and the C-terminal for tetrapod (B) and avian (C) CASQ1 (top three graphs) and CASQ2 (bottom three graphs). Dots represent the mean and boxes are the IQR with median center line. In B, each graph contains a diagram to illustrate significant relationships between groups where the group listed on the diagram corresponds to the row of boxes beside it and the columns correspond to the boxplot. C depicts significant relationships between avian clades with brackets (\*  $P < 0.05$ , \*\*  $P < 0.01$ , and \*\*\*  $P < 0.001$ ). Nonsignificant relationships are not marked. Abbreviations: AMP = amphibians, MAM = mammals, SQU = squamates, TES = Testudines, CRO = crocodilians, PALE = Palaeognathae (ratites), OAVE = other Aves, PASS = passerine birds. D) Amino acid residues of CASQ2 tails from 56 vertebrate species.

<https://doi.org/10.1371/journal.pgen.1011663.g003>

locus, which may have contributed to this gene truncation. Additionally, *CASQ1* is located on microchromosomes in both non-avian reptiles and birds, which potentially affects their dynamics of gene rearrangement and loss [46–49].

The retention of a full-length *CASQ1* coding sequence in several birds does leave the possibility that this gene might be functioning in some other context. We quantified expression in chicken lungs, intestines, brains, kidneys, and testes and *CASQ1* was either lowly expressed or exhibited much lower expression than *CASQ2* in all cases (Tables E and F in [S1 Data](#)). While it is possible that *CASQ1* has neofunctionalized or maintains a minor undiscovered function in avian tissues, the dramatic drop in skeletal muscle *CASQ1* expression across all birds sampled still supports a synfunctionalization by *CASQ2* with regards to its primary skeletal  $\text{jSR}$  role. More comprehensive RNA sampling will be necessary to discover any additional *CASQ1* functions in birds.

### Avian calsequestrin-2 has enhancements to calcium binding

Calsequestrin is one of the most negatively charged vertebrate proteins with negatively charged amino acids distributed throughout the protein that facilitate the binding of  $\text{Ca}^{2+}$  ions [9,10,27,50]. *CASQ* has three structural domains that arrange triangularly and are flanked by unstructured N- and C-terminal tails [9,15]. As the concentration of calcium rises, both calsequestrin proteins stack into dimers, oligomers, helical polymers, and finally a lattice of polymer fibers [12,19]. The N-terminal domain carries a slightly positive charge, and the C-terminal tail is mostly or completely aspartic acid and has a highly negative charge. The negative charge of the molecule as a whole—particularly the C-terminal tail—creates a highly electronegative surface for  $\text{Ca}^{2+}$  ions to bind [10,11,28].

We analyzed 299 *CASQ1* and 507 *CASQ2* tetrapod sequences and determined that avian *CASQ2* has greater overall net negativity than most other amniotes (Fig 3). While birds and other reptiles do not significantly differ in negativity of the primary domains, birds are significantly more negative in their C-terminal ( $P < 0.001$ ). Thus, increased negativity in avian *CASQ2* compared to other amniotes is driven by increased acidic residues in the C-terminal tail. To better understand where this shift occurred, we also compared amino acid sequences within birds (Fig 3C). The C-terminal tail of all non-ratite birds is nearly always completely aspartic acid residues (Fig 3D). Within birds, ratite *CASQ2* has a significantly less negative C-terminal compared to non-Passerines ( $P < 0.001$ ) and Passerines have a significantly more negative net charge compared to all other non-ratite birds ( $P < 0.001$ ). Both amphibian *CASQ1* and *CASQ2* exhibit greater negativity on average than all other tetrapod groups. Amphibian calsequestrins often have longer tails, and this often increases the total number of negative amino acids within the protein. Because this trend is observed in both *CASQ* paralogs, future research is needed to understand the functional role of calsequestrin in clades diverging before amniotes.

We also tested an alignment of 69 representative vertebrate *CASQ2* coding sequences with aBSREL and found evidence of significant  $d_N/d_S$  increases on the ancestral branch of Sauropsida (LRT,  $P < 0.05$  with Holm-Bonferroni correction) [51]. This evidence indicates a possible selective event on the molecule as a whole in a Sauropsida ancestor, followed by an increase in the net number of negatively charged amino acids in the C-terminal tail in birds.

The C-terminal is the major calcium-binding site ion calsequestrin. While high and low-affinity calcium-binding sites can be found across the entire protein, the total negativity of the unstructured C-terminal correlates more closely with the protein's calcium-binding capacity than the negativity of the protein as a whole [10]. As an example, the mutation of only three charged aspartic acids in the C-terminal tail to neutral alanine was sufficient to reduce *CASQ*

calcium binding by 45% [11]. High calcium concentrations in chicken and quail directly support the increased  $\text{Ca}^{2+}$  binding of avian CASQ [52,53]. We conclude that CASQ2 likely increased calcium binding capacity dramatically in an avian ancestor, and that additional cumulative boosts to calcium binding appear to have occurred sequentially in ancestors of Neognathae and Passeriformes.

In contrast, avian CASQ1 has a significantly less negative charge compared to all other tetrapod groups, both overall and in the C-terminal tail (all  $P < 0.05$ ; Fig 3B). Oscine and Suboscine passerine CASQ1 sequences have separate deletions that remove large portions of the first and second domains, including the entirety of the intermolecular binding region needed for polymerization (Fig 3A). Based on existing structural models, we conclude that there are no passerine CASQ1 proteins with the structural potential to be capable of any substantial calcium binding or formation of multimers. Additionally, the generally lower net negativity of avian CASQ1 further reinforces the expression data that indicates a severely reduced, if not completely nonfunctional, role in myocyte calcium binding.

### Avian CASQ2 synfunctionalization indicates a shift in cardiac function

Based on the expression, sequence, and structural data we propose a model of avian calsequestrin with an unusual series of gene duplication and loss events. First, in a jawed vertebrate ancestor, the ancestral calsequestrin duplicated and subfunctionalized. Later, in an avian ancestor, changes to the coding sequence or regulation (possibly both) of CASQ2 caused it to functionally replace the activity of CASQ1 in the jSR, a process previously defined as synfunctionalization [33]. Once CASQ2 became synfunctionalized, the displaced CASQ1 decreased in expression in an ancestor of all birds and was subsequently lost, truncated, or pseudogenized in Psittaciformes and Passeriformes lineage. In this synfunctionalization model, we propose that positive selection on CASQ2 C-terminal tail and calcium-binding capacity in cardiomyocytes drove CASQ2 to also become functionally superior to CASQ1 in skeletal muscles. Once rendered effectively redundant in its putative ancestral role, CASQ1 was apparently downregulated and eventually decayed (pseudogenization and loss) in many lineages.

Why would CASQ2 have been under positive selection in avian ancestors? CASQ2 is the predominant calsequestrin observed in chordate cardiomyocytes. Mammals exhibit improved cardiac function compared to ectothermic species due to changes in cardiomyocyte structure [54–56]. These changes include the development of thicker, shorter cardiomyocytes, often with t-tubules that allow for calcium-induced-calcium release to occur simultaneously across the cell [57]. Compared to mammals, avian hearts exhibit further improved performance, as evident by greater cardiac output and stroke volume [56,57]. Birds lack t-tubules and have elongated myocytes more typical of ectothermic vertebrates [57]. Investigations on the cellular structure and calcium-cycling processes of cardiomyocytes in birds have demonstrated that increased cardiac performance is instead linked to changes in SR calcium content and cycling [53,57].

The conclusion from these lines of evidence seems to be that, on a cellular level, avian ancestors developed higher-performing cardiomyocytes in a completely separate way from mammals. Because birds have maintained roughly the same cardiomyocyte structure as reptiles, this performance may have resulted from changes in SR calcium cycling at the molecular level. Studies investigating the relationship between CASQ1 content, calcium leakage, and free calcium content have demonstrated that decreased calcium-binding results in increased calcium leakage via SERCA pumps [20]. Increased CASQ1 in mammalian fast-twitch muscles reduces leakage and prevents additional ATP waste because of improved calcium-binding [20]. As avian cardiomyocytes exhibit increased SR calcium content compared to mammals,

improved calcium-binding in the SR likely further contributes to decreased calcium leakage needed to sustain frequent, rapid contractions [57]. Thus, alteration of CASQ2's  $\text{Ca}^{2+}$  binding capacity and polymerization capabilities is potentially a major factor contributing to the improved cardiomyocyte performance and subsequent greater cardiovascular function in birds.

While our data clearly supports that CASQ2 has adopted CASQ1's role in avian skeletal muscles, the primary driver of this shift is less clear, and we can only comment on possibilities. A possible triggering event could have been selection for increased cardiac performance resulting in better calcium-binding capabilities in CASQ2 that could have shifted the dynamics of SR  $\text{Ca}^{2+}$  flux and led to a reduction of CASQ1 expression in skeletal myocytes. Alternatively, CASQ2 could have been under selection for skeletal muscle performance, potentially related to one of the many physiological differences between modern birds and crocodylians with endothermy and flight as possible candidates. Nonshivering thermogenesis in birds is partially achieved by uncoupling SERCA activity and skeletal muscle contraction [58]. The result of this decoupling is decreased efficiency of SERCA pumps leading to an increase in the amount of heat produced in skeletal muscle tissues. Evidence from mammals has demonstrated both that CASQ mutations can result in hyperthermia-like myopathies and that CASQ expression influences SERCA expression [59]. If positive selection for cardiac performance also resulted in changes to calcium-cycling processes in skeletal muscle tissues, positive selection for CASQ2 in skeletal muscle tissues may be related to the development of avian endothermy, improved cardiovascular and skeletal muscle calcium cycling as a component of flight, or both. Further empirical work to characterize the function of calsequestrins in active avian skeletal muscles will be necessary to test these hypotheses.

## Conclusions

A more complete understanding of genome-trait connections over the long evolutionary histories of animals requires knowledge of both the changing mutational nature of the genomic sequences and their changing context of expression. Our model of calsequestrin synfunctionalization offers new perspectives on the post-duplication relationships of genes and the layered processes of adaptation, pseudogenization, and loss that affect gene families. Defining the intricacies of calcium-cycling in avian muscle tissue requires further empirical investigation, but the evidence we present clearly indicates a substantial event in the molecular evolution of birds. The altered cardiac and skeletal junctional sarcoplasmic reticulum on a molecular level has potentially broad implications for the evolution of key avian cardiovascular, musculoskeletal, and biomechanical traits. Given the many large and small eukaryotic gene families, we suspect many other cases of synfunctionalization are likely to be discovered and hope this work provokes additional investigations of the deep-time evolutionary expression dynamics of animal gene families.

## Materials and methods

### Calsequestrin sequence set construction

We queried NCBI's GenBank ortholog database for chordate protein sequences labeled as either CASQ, CASQ1, or CASQ2. While this produced many sequences, we feared it left out homologous genes from species with little or poor genome annotation. To overcome this limitation, we used NCBI's Basic Local Alignment Search Tool (BLAST) [60] protein-protein search option (blastp). BLAST locates homologous sequences by comparing an initial query sequence against NCBI's GenBank databases. Because these searches may be sensitive to given query parameters, we frequently changed the starting sequence, algorithm, database, and

returnable taxa to prevent relying on too broad of a search query. We also used nucleotide sequences from calsequestrin-encoding DNA regions to find CASQ-like DNA sequences using tblastn and blastx. This provided us with additional species that were not found with protein-protein searches alone. We used these same methods to search for homologous sequences in the University of California Santa Cruz's (UCSC) Genome database.

Because BLAST compares sequences and not gene or protein descriptions to locate homologs, it can return hypothetical or unknown proteins. To confirm that these proteins were truly homologous and not similar (yet unrelated) genes, we viewed the sequences within the genome to identify retained syntenic features. *CASQ1* does not appear in the GRCg7b (GCF\_016699485.2) version of the chicken genome, but does appear in the previous GRCq7w\_WZ version (GCF\_016700215.2). We located a homologous nucleotide sequence for CASQ in the inshore hagfish (*Eptatretus burgeri*) using tblastn and annotated a full length CASQ coding sequence with homologous intron/exon boundaries using the great white shark (*Chacharodon carcharias*) CASQ2 sequence.

### Alignment, quality control, and phylogenetic inference

We aligned our sequences with MAFFT's L-INS-i progressive method [61] and made small manual edits to sequences when needed (Tables B, G, and H in [S1 Data](#), and [S3](#) and [S4 Data](#) files). Some sequences possessed highly unique regions that shared little to no homology with other sequences in the alignment. If a unique motif appeared to be erroneous from an annotation error, we either removed the erroneous portion or removed the entire sequence from the alignment. Calsequestrin possesses a long tail of repeating acidic residues, though the length and exact composition of the tail varies greatly between species. We suspected that the highly variable tails would introduce a large amount of noise in our phylogenetic reconstruction, and so we truncated our sequences by removing the acidic tail.

We inferred our phylogenies using RAxML-ng with WAG [62,63] as our model of protein evolution and used bootstrapping to generate branch supports. We generated bootstrap values using the transfer bootstrap expectation [64]. Our final phylogeny had multiple near-zero branches due to the large number of species included in the alignment with highly similar or identical sequences. iTol's online tree editor (v6) [65] was used to create the phylogeny in [Fig 1C](#).

### Transcriptome assembly

Sequences were downloaded from the NCBI Sequence Read Archive (<https://www.ncbi.nlm.nih.gov/sra>) using fastq-dump and assembled using Trinity (v.2.15.1) [66]. If single-end or multiple FASTQ files were available for an individual, the command line was altered to incorporate these files or the correct single-end options, but all other options remained the same. See [Table B](#) in [S1 Data](#) for details on individual assemblies and [S2 Data](#) for assembled and newly annotated sequences.

### Identification of calsequestrin genes

Calsequestrin genes were identified in *de novo* and reference transcriptomes by using BLAST to compare transcript sequences against nucleotide sequences (tBLASTn) using known, closely related CASQ1 and CASQ2 protein sequences as the queries. The high conservation of sequence in calsequestrin, its distinctness from other genes, and clear sequence identity between CASQ1 and CASQ2 made identification of these genes generally unambiguous. To exhaustively look for CASQ genes in assembled transcriptomes, we searched down to tBLASTn *E*-value cutoff  $E < 10$ , but all successful matches had an of  $E < 10^{-20}$ . *De novo* sequences are provided in [S2 Data](#).

## Multidimensional scaling plot

The MDS was calculated from a pairwise distance matrix using simple amino acid distances (all differences = 1; [S5 Data](#)). Pairwise distances were calculated using a Python3 script. The MDS was calculated based on Euclidean distances with 10000 iterations using the *mds* function from the *scikit-learn* Python3 module and plotted using *matplotlib*.

## Transcript quantification

Transcript expression levels were quantified using kallisto (v. 0.46.1; <https://github.com/pachterlab/kallisto>) [67] for species with a reference transcriptome (Table I in [S1 Data](#)) and species with de novo Trinity assembly transcriptomes (v.2.15.1). We conducted multiple tests on the effect of additional quality trimming schema and different quantification methods. Quantification with salmon [68], RSEM [69], and STAR [70], and the use of CutAdapt [71] found expression rates reported comparable to kallisto with no additional trimming beyond fastq-dump adapter removal. Based on these tests, we were satisfied that the relative expression numbers reported here are robust to methodology. Expression values are available in Tables C and D in [S1 Data](#). Counts per million (cpm) expression values shown in the main text were calculated as the estimated counts from kallisto divided by the total number of mapped reads by kallisto times  $10^6$ .

## Calculation of electronegative charge and statistical methods

We quantified an overall net negativity of CASQ proteins, their domains, and the c-terminal by counting the number of positively and negatively charged residues ([S5 Data](#)). Aspartic and glutamic acids were counted as a single negative charge (-1) while arginine, histidine, and lysine were each counted as a single positive charge (+1). All other residues were given a value of 0. We then tested for significant differences between groups of tetrapods and separately between birds grouped roughly by their evolutionary relatedness (Paleognathe, Psittaciformes + Passerines, and all other avian species). We calculated *P*-values using pairwise *t*-tests with Bonferroni corrections. Amino acid sequence in [Fig 3C](#) was created with Jalview (V2) [72]. To test for selection in CASQ2, we downloaded nucleotide sequences for 69 representative vertebrate sequences and used aBSREL [51] via Datamonkey's web interface ([S7](#) and [S8 Data](#) files) [73].

## Supporting Information

**S1 Fig. Breakdown of 1,071 sequences used in phylogenetic construction.** Color corresponds to the number of sequences attributed to each taxonomic grouping (PRED = pre-duplication jawless fishes, CHO = chondrichthyes, ACT = actinopterygii, SAR = sarcopterygii, AMP = amphibian, REP = reptile, MAM = mammal, AVE = aves). The exact number of sequences per taxonomic group are provided in Table A in [S1 Data](#). A detailed list of sequences and information on exclusion criteria is available in Table B in [S1 Data](#). (PDF)

**S2 Fig. Phylogeny of non-chordate calsequestrins with some select chordate sequences.** Multiple sequence alignment used to construct the tree is available in [S3 Data](#). (PDF)

**S3 Fig. Calsequestrin synteny and exon structure in vertebrates.** CASQ1 (A) and CASQ2 (B) exhibit varying degrees of conservation in vertebrates. CASQ1 is usually proximal to *PEA15*, *IRRE-L*, *DCAF8*, and/or *ATPIA2*. In all vertebrates, CASQ2 is near *VANG-L* and

*NHLH2-L*. Calsequestrin paralogs are depicted in black, and representative species are included above each gene block. Arrow depicts direction of transcription. C) *CASQ1* exon structure in vertebrates. In passerines (highlighted in blue) *CASQ1* displays a breakdown in exon structure with missing or modified exons. The length of each exon (boxes) are proportional to the exon length in human *CASQ1*. Present exons are black, missing exons are white, and exons that are present but not homologous are depicted in gray.

(PDF)

**S1 Data. Table A: Composition of 1,071 species included in phylogenetic reconstruction by paralog and taxonomic group.** Table B: CASQ phylogeny sequences. Included are sequences used to construct a phylogenetic tree of CASQ1, CASQ2, and pre-duplication CASQ. Sequences that were pruned (with justification provided) or modified (with a description of the modification) are also included. Table C: CASQ expression date. RNA expression data in total counts, counts per million, and transcripts per million for CASQ1/CASQ2/ CASQ. Source material is also provided. Table D: Key for columns in Table C. Table E: Expression of CASQ paralogs in muscle and non-muscle tissues of chicken. Table F: Expression of CASQ paralogs in non-muscle tissues of birds. Table G: Table of chordate protein and nucleotide sequences (when available) obtained through search efforts. De novo assembly sequences are not included. Table H: Table of non-chordate protein and nucleotide sequences obtained through search efforts. Table I: NCBI Reference Transcriptomes used for kallisto mapping (XLSX)

**S2 Data. De novo assemblies.** PDF containing all de novo assembled sequences used for expression analysis and phylogenetic reconstruction.

(PDF)

**S3 Data. Invertebrate multiple sequence alignment .** FASTA file of invertebrate CASQ sequences used to construct the phylogenetic tree in [S2 Fig](#).

(FASTA)

**S4 Data. Vertebrate multiple sequence alignment.** FASTA file of vertebrate CASQ sequences used to construct the phylogenetic tree in [Fig 1C](#).

(FASTA)

**S5 Data. Annotated multiple sequence alignment (FASTA file) used to estimate electro-negativity .**

(FASTA)

**S6 Data. Vertebrate multiple sequence alignment (FASTA file) of CASQ sequences used to construct the MDS in [Fig 1D](#).**

(FASTA)

**S7 Data. Nucleotide alignment (FASTA file) of sequences used to run aBSREL.**

(FASTA)

**S8 Data. Output results from aBSREL.**

(JSON)

## Acknowledgments

We thank Eric Schuppe for early conversations about this project. We thank undergraduate researchers Justin Absher, Rebecca Epling, and L. Morgyn Hill at Wake Forest University for contributing preliminary work to data collection from this paper. Thanks to Olivia Delgado,

Andria Harkey, Nicholas Kortessis, Stephen Smith, Nathanael Walker-Hale, and Ellen Weinheimer for feedback on this manuscript.

## Author contributions

**Conceptualization:** Christina M Harvey, Matther J Fuxjager, James B Pease.

**Data curation:** Christina M Harvey, Matther J Fuxjager, James B Pease.

**Formal analysis:** Christina M Harvey, James B Pease.

**Funding acquisition:** Matther J Fuxjager, James B Pease.

**Investigation:** Christina M Harvey, Matther J Fuxjager, James B Pease.

**Methodology:** Christina M Harvey, James B Pease.

**Project administration:** James B Pease.

**Resources:** Christina M Harvey, James B Pease.

**Software:** Christina M Harvey, James B Pease.

**Supervision:** Matther J Fuxjager, James B Pease.

**Visualization:** Christina M Harvey.

**Writing – original draft:** Christina M Harvey, James B Pease.

**Writing – review & editing:** Christina M Harvey, Matther J Fuxjager, James B Pease.

## References

1. Ohno S. Evolution by gene duplication. Berlin: Springer-Verlag; 1970.
2. Demuth JP, De Bie T, Stajich JE, Cristianini N, Hahn MW. The evolution of mammalian gene families. *PLoS One*. 2006;1(1):e85. <https://doi.org/10.1371/journal.pone.0000085> PMID: 17183716
3. Gout J-F, Lynch M. Maintenance and loss of duplicated genes by dosage subfunctionalization. *Mol Biol Evol*. 2015;32(8):2141–8. <https://doi.org/10.1093/molbev/msv095> PMID: 25908670
4. Kuzmin E, VanderSluis B, Nguyen Ba AN, Wang W, Koch EN, Usaj M, et al. Exploring whole-genome duplicate gene retention with complex genetic interaction analysis. *Science*. 2020;368(6498):eaaz5667. <https://doi.org/10.1126/science.aaz5667> PMID: 32586993
5. Innan H, Kondrashov F. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet*. 2010;11(2):97–108. <https://doi.org/10.1038/nrg2689> PMID: 20051986
6. Qian W, Liao B-Y, Chang AY-F, Zhang J. Maintenance of duplicate genes and their functional redundancy by reduced expression. *Trends Genet*. 2010;26(10):425–30. <https://doi.org/10.1016/j.tig.2010.07.002> PMID: 20708291
7. Birchler JA, Yang H. The multiple fates of gene duplications: Deletion, hypofunctionalization, subfunctionalization, neofunctionalization, dosage balance constraints, and neutral variation. *Plant Cell*. 2022;34(7):2466–74. <https://doi.org/10.1093/plcell/koac076> PMID: 35253876
8. MacLennan DH, Wong PT. Isolation of a calcium-sequestering protein from sarcoplasmic reticulum. *Proc Natl Acad Sci U S A*. 1971;68(6):1231–5. <https://doi.org/10.1073/pnas.68.6.1231> PMID: 4256614
9. Park H, Wu S, Dunker AK, Kang C. Polymerization of calsequestrin. Implications for Ca<sup>2+</sup> regulation. *J Biol Chem*. 2003;278(18):16176–82. <https://doi.org/10.1074/jbc.M300120200> PMID: 12594204
10. Park H, Park IY, Kim E, Youn B, Fields K, Dunker AK, et al. Comparing skeletal and cardiac calsequestrin structures and their calcium binding: a proposed mechanism for coupled calcium binding and protein polymerization. *J Biol Chem*. 2004;279(17):18026–33. <https://doi.org/10.1074/jbc.M311553200> PMID: 14871888
11. Beard NA, Dulhunty AF. C-terminal residues of skeletal muscle calsequestrin are essential for calcium binding and for skeletal ryanodine receptor inhibition. *Skelet Muscle*. 2015;5:6. <https://doi.org/10.1186/s13395-015-0029-7> PMID: 25861445
12. Manno C, Figueroa LC, Gillespie D, Fitts R, Kang C, Franzini-Armstrong C, et al. Calsequestrin depolymerizes when calcium is depleted in the sarcoplasmic reticulum of working muscle. *Proc Natl Acad Sci U S A*. 2017;114(4):E638–47. <https://doi.org/10.1073/pnas.1620265114> PMID: 28069951

13. Wang Q, Paskevicius T, Filbert A, Qin W, Kim HJ, Chen X-Z, et al. Phylogenetic and biochemical analysis of calsequestrin structure and association of its variants with cardiac disorders. *Sci Rep*. 2020;10(1):18115. <https://doi.org/10.1038/s41598-020-75097-3> PMID: [33093545](https://pubmed.ncbi.nlm.nih.gov/33093545/)
14. Ng K, Titus EW, Lieve KV, Roston TM, Mazzanti A, Deiter FH, et al. An international multicenter evaluation of inheritance patterns, arrhythmic risks, and underlying mechanisms of *CASQ2*-catecholaminergic polymorphic ventricular tachycardia. *Circulation*. 2020;142(10):932–47. <https://doi.org/10.1161/CIRCULATIONAHA.120.045723> PMID: [32693635](https://pubmed.ncbi.nlm.nih.gov/32693635/)
15. Bal NC, Sharon A, Gupta SC, Jena N, Shaikh S, Györke S, et al. The catecholaminergic polymorphic ventricular tachycardia mutation R33Q disrupts the N-terminal structural motif that regulates reversible calsequestrin polymerization. *J Biol Chem*. 2010;285(22):17188–96. <https://doi.org/10.1074/jbc.M109.096354> PMID: [20353949](https://pubmed.ncbi.nlm.nih.gov/20353949/)
16. Knollmann BC, Chopra N, Hlaing T, Akin B, Yang T, Etensohn K, et al. *Casq2* deletion causes sarcoplasmic reticulum volume increase, premature Ca<sup>2+</sup> release, and catecholaminergic polymorphic ventricular tachycardia. *J Clin Invest*. 2006;116(9):2510–20. <https://doi.org/10.1172/JCI29128> PMID: [16932808](https://pubmed.ncbi.nlm.nih.gov/16932808/)
17. Basaki M, Asasi K, Tabandeh MR, Aminlari M. Polymorphism identification and cardiac gene expression analysis of the calsequestrin 2 gene in broiler chickens with sudden death syndrome. *Br Poult Sci*. 2016;57(2):151–60. <https://doi.org/10.1080/00071668.2015.1099615> PMID: [26953612](https://pubmed.ncbi.nlm.nih.gov/26953612/)
18. Xing T, Zhao X, Wang P, Chen H, Xu X, Zhou G. Different oxidative status and expression of calcium channel components in stress-induced dysfunctional chicken muscle. *J Anim Sci*. 2017;95(4):1565–73. <https://doi.org/10.2527/jas.2016.0868> PMID: [28464077](https://pubmed.ncbi.nlm.nih.gov/28464077/)
19. Wei L, Hanna AD, Beard NA, Dulhunty AF. Unique isoform-specific properties of calsequestrin in the heart and skeletal muscle. *Cell Calcium*. 2009;45(5):474–84. <https://doi.org/10.1016/j.ceca.2009.03.006> PMID: [19376574](https://pubmed.ncbi.nlm.nih.gov/19376574/)
20. Murphy RM, Larkins NT, Mollica JP, Beard NA, Lamb GD. Calsequestrin content and SERCA determine normal and maximal Ca<sup>2+</sup> storage levels in sarcoplasmic reticulum of fast- and slow-twitch fibres of rat. *J Physiol*. 2009;587(2):443–60. <https://doi.org/10.1113/jphysiol.2008.163162> PMID: [19029185](https://pubmed.ncbi.nlm.nih.gov/19029185/)
21. Damiani E, Volpe P, Margreth A. Coexpression of two isoforms of calsequestrin in rabbit slow-twitch muscle. *J Muscle Res Cell Motil*. 1990;11(6):522–30. <https://doi.org/10.1007/BF01745219> PMID: [2084148](https://pubmed.ncbi.nlm.nih.gov/2084148/)
22. Wang Q, Michalak M. Calsequestrin. Structure, function, and evolution. *Cell Calcium*. 2020;90:102242. <https://doi.org/10.1016/j.ceca.2020.102242> PMID: [32574906](https://pubmed.ncbi.nlm.nih.gov/32574906/)
23. Sun Z, Wang L, Han L, Wang Y, Zhou Y, Li Q, et al. Functional calsequestrin-1 is expressed in the heart and its deficiency is causally related to malignant hyperthermia-like arrhythmia. *Circulation*. 2021;144(10):788–804. <https://doi.org/10.1161/CIRCULATIONAHA.121.053255> PMID: [34162222](https://pubmed.ncbi.nlm.nih.gov/34162222/)
24. Choi ES, Clegg DO. Identification and developmental expression of a chicken calsequestrin homolog. *Dev Biol*. 1990;142(1):169–77. [https://doi.org/10.1016/0012-1606\(90\)90160-k](https://doi.org/10.1016/0012-1606(90)90160-k) PMID: [2227093](https://pubmed.ncbi.nlm.nih.gov/2227093/)
25. Furlan S, Paradiso B, Greotti E, Volpe P, Nori A. Calsequestrin in Purkinje cells of mammalian cerebellum. *Acta Histochem*. 2023;125(2):152001. <https://doi.org/10.1016/j.acthis.2023.152001> PMID: [36669254](https://pubmed.ncbi.nlm.nih.gov/36669254/)
26. Qin J, Valle G, Nani A, Chen H, Ramos-Franco J, Nori A, et al. Ryanodine receptor luminal Ca<sup>2+</sup> regulation: swapping calsequestrin and channel isoforms. *Biophys J*. 2009;97(7):1961–70. <https://doi.org/10.1016/j.bpj.2009.07.030> PMID: [19804727](https://pubmed.ncbi.nlm.nih.gov/19804727/)
27. Sanchez EJ, Lewis KM, Danna BR, Kang C. High-capacity Ca<sup>2+</sup> binding of human skeletal calsequestrin. *J Biol Chem*. 2012;287(14):11592–601. <https://doi.org/10.1074/jbc.M111.335075> PMID: [22337878](https://pubmed.ncbi.nlm.nih.gov/22337878/)
28. Bal NC, Jena N, Chakravarty H, Kumar A, Chi M, Balaraju T, et al. The C-terminal calcium-sensitive disordered motifs regulate isoform-specific polymerization characteristics of calsequestrin. *Biopolymers*. 2015;103(1):15–22. <https://doi.org/10.1002/bip.22534> PMID: [25091206](https://pubmed.ncbi.nlm.nih.gov/25091206/)
29. Fuxjager MJ, Eaton J, Lindsay WR, Salwiczek LH, Rensel MA, Barske J, et al. Evolutionary patterns of adaptive acrobatics and physical performance predict expression profiles of androgen receptor - but not oestrogen receptor - in the forelimb musculature. *Funct Ecol*. 2015;29(9):1197–208. <https://doi.org/10.1111/1365-2435.12438> PMID: [26538789](https://pubmed.ncbi.nlm.nih.gov/26538789/)
30. Pease JB, Driver RJ, de la Cerda DA, Day LB, Lindsay WR, Schlinger BA, et al. Layered evolution of gene expression in “superfast” muscles for courtship. *Proc Natl Acad Sci U S A*. 2022;119(14):e2119671119. <https://doi.org/10.1073/pnas.2119671119> PMID: [35363565](https://pubmed.ncbi.nlm.nih.gov/35363565/)
31. Paolini C, Quarta M, Nori A, Boncompagni S, Canato M, Volpe P, et al. Reorganized stores and impaired calcium handling in skeletal muscle of mice lacking calsequestrin-1. *J Physiol*. 2007;583(Pt 2):767–84. <https://doi.org/10.1113/jphysiol.2007.138024> PMID: [17627988](https://pubmed.ncbi.nlm.nih.gov/17627988/)

32. Canato M, Scorzeto M, Giacomello M, Protasi F, Reggiani C, Stienen GJM. Massive alterations of sarcoplasmic reticulum free calcium in skeletal muscle fibers lacking calsequestrin revealed by a genetically encoded probe. *Proc Natl Acad Sci U S A*. 2010;107(51):22326–31. <https://doi.org/10.1073/pnas.1009168108> PMID: [21135222](https://pubmed.ncbi.nlm.nih.gov/21135222/)
33. Gitelman I. Evolution of the vertebrate twist family and synfunctionalization: a mechanism for differential gene loss through merging of expression domains. *Mol Biol Evol*. 2007;24(9):1912–25. <https://doi.org/10.1093/molbev/msm120> PMID: [17567594](https://pubmed.ncbi.nlm.nih.gov/17567594/)
34. Liberles DA. Gene duplication: red queens, linkage, redundancy and synfunctionalization. *Heredity (Edinb)*. 2008;101(4):299–300. <https://doi.org/10.1038/hdy.2008.81> PMID: [18698335](https://pubmed.ncbi.nlm.nih.gov/18698335/)
35. Peona V, Weissensteiner MH, Suh A. How complete are “complete” genome assemblies?—An avian perspective. *Mol Ecol Resour*. 2018;18(6):1188–95. <https://doi.org/10.1111/1755-0998.12933> PMID: [30035372](https://pubmed.ncbi.nlm.nih.gov/30035372/)
36. Fernández R, Gabaldón T. Gene gain and loss across the metazoan tree of life. *Nat Ecol Evol*. 2020;4(4):524–33. <https://doi.org/10.1038/s41559-019-1069-x> PMID: [31988444](https://pubmed.ncbi.nlm.nih.gov/31988444/)
37. Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, et al. The evolution of gene expression levels in mammalian organs. *Nature*. 2011;478(7369):343–8. <https://doi.org/10.1038/nature10532> PMID: [22012392](https://pubmed.ncbi.nlm.nih.gov/22012392/)
38. Sebé-Pedrós A, Saudemont B, Chomsky E, Plessier F, Mailhé M-P, Renno J, et al. Cnidarian cell type diversity and regulation revealed by whole-organism single-cell RNA-Seq. *Cell*. 2018;173(6):1520–1534.e20. <https://doi.org/10.1016/j.cell.2018.05.019> PMID: [29856957](https://pubmed.ncbi.nlm.nih.gov/29856957/)
39. Mah JL, Dunn CW. Cell type evolution reconstruction across species through cell phylogenies of single-cell RNA sequencing data. *Nat Ecol Evol*. 2024;8(2):325–38. <https://doi.org/10.1038/s41559-023-02281-9> PMID: [38182680](https://pubmed.ncbi.nlm.nih.gov/38182680/)
40. Infante C, Ponce M, Manchado M. Duplication of calsequestrin genes in teleosts: molecular characterization in the Senegalese sole (*Solea senegalensis*). *Comp Biochem Physiol B Biochem Mol Biol*. 2011;158(4):304–14. <https://doi.org/10.1016/j.cbpb.2011.01.002> PMID: [21256971](https://pubmed.ncbi.nlm.nih.gov/21256971/)
41. Glasauer SMK, Neuhauss SCF. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol Genet Genomics*. 2014;289(6):1045–60. <https://doi.org/10.1007/s00438-014-0889-2> PMID: [25092473](https://pubmed.ncbi.nlm.nih.gov/25092473/)
42. Holland LZ, Ocampo Daza D. A new look at an old question: when did the second whole genome duplication occur in vertebrate evolution?. *Genome Biol*. 2018;19(1):209. <https://doi.org/10.1186/s13059-018-1592-0> PMID: [30486862](https://pubmed.ncbi.nlm.nih.gov/30486862/)
43. Nakatani Y, Shingate P, Ravi V, Pillai NE, Prasad A, McLysaght A, et al. Reconstruction of proto-vertebrate, proto-cyclostome and proto-gnathostome genomes provides new insights into early vertebrate evolution. *Nat Commun*. 2021;12(1). <https://doi.org/10.1038/s41467-021-24573-z>
44. Marletaz F, Timoshevskaya N, Timoshevskiy A, Parey E, Simakov O, Gavriouchkina D, et al. The hagfish genome and the evolution of vertebrates. *Nature* 2024;627. doi: [10.1038/s41467-021-24573-z](https://doi.org/10.1038/s41467-021-24573-z).
45. Yu D, Ren Y, Uesaka M, Beavan AJS, Muffato M, Shen J, et al. Hagfish genome elucidates vertebrate whole-genome duplication events and their evolutionary consequences. *Nat Ecol Evol*. 2024;8(3):519–35. <https://doi.org/10.1038/s41559-023-02299-z> PMID: [38216617](https://pubmed.ncbi.nlm.nih.gov/38216617/)
46. Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*. 2014;346(6215):1311–20. <https://doi.org/10.1126/science.1251385>
47. Huang Z, De O Furo I, Liu J, Peona V, Gomes AJB, Cen W, et al. Recurrent chromosome reshuffling and the evolution of neo-sex chromosomes in parrots. *Nat Commun*. 2022;13(1):944. <https://doi.org/10.1038/s41467-022-28585-1> PMID: [35177601](https://pubmed.ncbi.nlm.nih.gov/35177601/)
48. Kapusta A, Suh A, Feschotte C. Dynamics of genome size evolution in birds and mammals. *Proc Natl Acad Sci U S A*. 2017;114(8):E1460–9. <https://doi.org/10.1073/pnas.1616702114> PMID: [28179571](https://pubmed.ncbi.nlm.nih.gov/28179571/)
49. Waters PD, Patel HR, Ruiz-Herrera A, Álvarez-González L, Lister NC, Simakov O, et al. Microchromosomes are building blocks of bird, reptile, and mammal chromosomes. *Proc Natl Acad Sci U S A*. 2021;118(45):e2112494118. <https://doi.org/10.1073/pnas.2112494118> PMID: [34725164](https://pubmed.ncbi.nlm.nih.gov/34725164/)
50. Kumar A, Chakravarty H, Bal NC, Balaraju T, Jena N, Misra G, et al. Identification of calcium binding sites on calsequestrin 1 and their implications for polymerization. *Mol Biosyst*. 2013;9(7):1949–57. <https://doi.org/10.1039/c3mb25588c> PMID: [23629537](https://pubmed.ncbi.nlm.nih.gov/23629537/)
51. Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol*. 2015;32(5):1342–53. <https://doi.org/10.1093/molbev/msv022> PMID: [25697341](https://pubmed.ncbi.nlm.nih.gov/25697341/)

52. Creazzo TL, Burch J, Godt RE. Calcium buffering and excitation-contraction coupling in developing avian myocardium. *Biophys J*. 2004;86(2):966–77. [https://doi.org/10.1016/S0006-3495\(04\)74172-7](https://doi.org/10.1016/S0006-3495(04)74172-7) PMID: [14747332](https://pubmed.ncbi.nlm.nih.gov/14747332/)
53. Filatova TS, Abramochkin DV. Characteristics of fast sodium current in isolated quail cardiomyocytes. *Moscow Univ BiolSci Bull*. 2022;77(3):159–64. <https://doi.org/10.3103/s0096392522030051>
54. Hillman SS, Hedrick MS. A meta-analysis of in vivo vertebrate cardiac performance: implications for cardiovascular support in the evolution of endothermy. *J Exp Biol* 2015;218(8):1143–50. <https://doi.org/10.1242/jeb.118372>
55. Shiels HA, Galli GLJ. The sarcoplasmic reticulum and the evolution of the vertebrate heart. *Physiology (Bethesda)*. 2014;29(6):456–69. <https://doi.org/10.1152/physiol.00015.2014> PMID: [25362639](https://pubmed.ncbi.nlm.nih.gov/25362639/)
56. Shiels HA. Avian cardiomyocyte architecture and what it reveals about the evolution of the vertebrate heart. *Phil Trans R Soc B*. 2022;377(1864). <https://doi.org/10.1098/rstb.2021.0332>
57. Filatova TS, Abramochkin DV, Shiels HA. Warmer, faster, stronger: Ca<sup>2+</sup> cycling in avian myocardium. *J Exp Biol*. 2020;223(Pt 19):jeb228205. <https://doi.org/10.1242/jeb.228205> PMID: [32843363](https://pubmed.ncbi.nlm.nih.gov/32843363/)
58. Pani P, Bal NC. Avian adjustments to cold and non-shivering thermogenesis: what, where and how. *Biol Rev Camb Philos Soc*. 2022;97(6):2106–26. <https://doi.org/10.1111/brv.12885> PMID: [35899483](https://pubmed.ncbi.nlm.nih.gov/35899483/)
59. Wang Y, Xu L, Duan H, Pasek DA, Eu JP, Meissner G. Knocking down type 2 but not type 1 caldesmon reduces calcium sequestration and release in C2C12 skeletal muscle myotubes. *J Biol Chem*. 2006;281(22):15572–81. <https://doi.org/10.1074/jbc.M600090200> PMID: [16595676](https://pubmed.ncbi.nlm.nih.gov/16595676/)
60. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403–10. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2) PMID: [2231712](https://pubmed.ncbi.nlm.nih.gov/2231712/)
61. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30(4):772–80. <https://doi.org/10.1093/molbev/mst010> PMID: [23329690](https://pubmed.ncbi.nlm.nih.gov/23329690/)
62. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*. 2019;35(21):4453–5. <https://doi.org/10.1093/bioinformatics/btz305> PMID: [31070718](https://pubmed.ncbi.nlm.nih.gov/31070718/)
63. Whelan S, Goldman N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol*. 2001;18(5):691–9. <https://doi.org/10.1093/oxfordjournals.molbev.a003851> PMID: [11319253](https://pubmed.ncbi.nlm.nih.gov/11319253/)
64. Lemoine F, Domelevo Entfellner J-B, Wilkinson E, Correia D, Dávila Felipe M, De Oliveira T, et al. Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature*. 2018;556(7702):452–6. <https://doi.org/10.1038/s41586-018-0043-0> PMID: [29670290](https://pubmed.ncbi.nlm.nih.gov/29670290/)
65. Ciccarelli FD, Doerks T, von Mering C, Creevey CJ, Snel B, Bork P. Toward automatic reconstruction of a highly resolved tree of life. *Science*. 2006;311(5765):1283–7. <https://doi.org/10.1126/science.1123061> PMID: [16513982](https://pubmed.ncbi.nlm.nih.gov/16513982/)
66. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 2011;29(7):644–52. <https://doi.org/10.1038/nbt.1883> PMID: [21572440](https://pubmed.ncbi.nlm.nih.gov/21572440/)
67. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol*. 2016;34(5):525–7. <https://doi.org/10.1038/nbt.3519>
68. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;14(4):417–9. <https://doi.org/10.1038/nmeth.4197> PMID: [28263959](https://pubmed.ncbi.nlm.nih.gov/28263959/)
69. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf*. 2011;12:323. <https://doi.org/10.1186/1471-2105-12-323> PMID: [21816040](https://pubmed.ncbi.nlm.nih.gov/21816040/)
70. Kiriakidou M, McAllister JM, Sugawara T, Strauss JF 3rd. Expression of steroidogenic acute regulatory protein (StAR) in the human ovary. *J Clin Endocrinol Metab*. 1996;81(11):4122–8. <https://doi.org/10.1210/jcem.81.11.8923870> PMID: [8923870](https://pubmed.ncbi.nlm.nih.gov/8923870/)
71. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j*. 2011;17(1):10. <https://doi.org/10.14806/ej.17.1.200>
72. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. 2009;25(9):1189–91. <https://doi.org/10.1093/bioinformatics/btp033> PMID: [19151095](https://pubmed.ncbi.nlm.nih.gov/19151095/)
73. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky SL. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol Biol Evol*. 2018;35(3):773–7. <https://doi.org/10.1093/molbev/msx335> PMID: [29301006](https://pubmed.ncbi.nlm.nih.gov/29301006/)