Rbm45 Phylogenetics, Protein Domain Conservation, and Gene Architecture in Clade Metazoa

Virdjinija Vuchkovska[1,2], Teagan N. Marti[1,2], Anali Cisneros[1,2], Lauren M. Saiki[1], Jeffrey O. Henderson[1,2,*]

[1] Department of Science and Mathematics, Judson University, Elgin, IL 60123, USA

[2] Equal contribution

Abstract

Mammalian Rbm45 is predominately expressed in neuronal tissue and is integral in brain development and neuronal differentiation under physiological conditions. Dysregulation of Rbm45 has been strongly associated with neurodegenerative disorders in humans and can drive hepatocellular carcinoma through reprogramming lipid metabolism. Intriguingly, Rbm45 is an ancient protein, evolutionarily conserved throughout metazoans, including in sponges which lack a nervous system. Curiously, the evolution of Rbm45 gene structure and protein domain conservation across kingdom Animalia is largely unknown. We performed phylogenetic analysis of Rbm45 nucleotide and amino acid sequences from 36 species representing 9 phyla: Porifera, Cnidaria, Priapulida, Mollusca, Brachiopoda, Arthropoda, Echinodermata, Hemichordata, and Chordata. While the tree from Rbm45 nucleotide sequence data resulted in clades Protostomia and Deuterostomia showing paraphyly, the phylogeny derived from Rbm45 amino acid sequence largely recapitulated known monophyletic relationships among metazoans. Human RBM45 protein structure includes three RNA-binding domains (RBD), a homo-oligomerization association (HOA) domain, a nuclear localization sequence (NLS), and a nuclear export sequence (NES). Multiple sequence alignment across the same 36 taxa used for phylogenetic analysis revealed conservation of all three RBDs, the HOA, and NLS; in contrast the NES was only detected in clade Craniata and not in clades Ambulacraria and Protostomia. *Rbm45* gene structure analysis revealed increasing gene complexity concomitant with increasing evolutionary complexity. *Rbm45* from non-bilaterian taxa had from 2 to 4 large exons, while bilaterian taxa had between 6 to 17 small exons. These findings demonstrate that *Rbm45* is an ancient, highly conserved gene among metazoans suggesting a function in a breadth of neural/sensory systems.

Introduction

RNA-binding proteins (RBP) are an evolutionarily conserved [1] family of proteins that have been shown to participate in a constellation of cellular

functions. Specifically, RNA recognition motif-type (RRM) binding domain proteins (RBDPs) have been demonstrated to regulate post-transcriptional RNA metabolism by modulating the longevity [2-5] and translational efficacy [6-11] of target mRNAs and contributing to splicing reactions [12]. RBDPs have also been identified participating in RNA-mediated [13-18] and RNA-independent protein-protein complexes [19-21]. These biochemical properties of RBDPs have been linked to roles in apoptosis [22], tumorigenesis [23], and neuropathology [24,25].

Rbm45 is an ancient gene, conserved from sponges to humans [26]. The human *RBM45* locus on chromosome 2q31.2 [26] encodes a promiscuous RBDP found to be expressed under normal physiological conditions predominantly in neuronal tissue of rats [27], mice [26,28], and humans [29,30]. Rbm45 has three canonical RNA-binding domains (RBD I, II, III) [27,31,32] that preferentially bind GC-rich RNA sequence motifs [27,33,34] and have recently been shown to facilitate binding to single-stranded DNA [35]; additionally, N-terminal RBD I and RBD II have been demonstrated to participate in RNA-independent protein-protein interactions [36]. Furthermore, Rbm45 contains a homo-oligomer assembly (HOA) domain that mediates homo- and heteromerization of Rbm45 and binding partners [31,37]. Finally, Rbm45 is capable of shuttling between the nucleus and cytoplasm by means of a nuclear localization signal (NLS) [31,32] and a nuclear export signal (NES) [32] facilitating its presumed role in regulating nuclear spliceosome and mRNA splicing functions [37-40].

Rbm45 was initially identified as being expressed spatiotemporally in neuronal precursor cells during brain development in rats [27] and has recently been shown to be necessary for neuroblastoma differentiation in the SH-SY5Y cell line [41]. Early work associated upregulation of Rbm45 with suppression of hypoxia induced apoptosis of cardiomyocytes [42] in mice and neuronal repair during spinal cord injury in a neonatal opossum model [43]. Subsequently, biochemical [25,29,31,32,37,44] and molecular genetic [45] studies have linked human RBM45 protein dysfunction/aggregation to the neurodegenerative syndromes: frontotemporal dementia, amyotrophic lateral sclerosis, and Alzheimer's disease. These data implicate Rbm45 in neurogenesis, neural repair, and neuropathophysiology. We [26] and others [31,32] have shown that *Rbm45* is highly conserved across metazoan taxa; intriguingly, our previous work [26] identified an *Rbm45* orthologue in sponges (phylum Porifera) which lack a nervous system. However, sponges do have a neural toolkit of genes [46] as well as neuroid cells [47] with secretory vesicles capable of communicating with adjacent cell types and coordinating cellular activity; this is suggestive of a vital role for phylum Porifera in nervous system evolution. To explore the deep homology of *Rbm45* in neural development, we have analyzed the phylogenetic history, protein domain conservation, and gene structure across 36 Rbm45 orthologues from 9 metazoan phyla. In this study, we show conservation of *Rbm45* over 650 million years of evolutionary history with preservation of RNA-binding, HOA, and NLS regions, as well as increasing complexity of gene architecture during the radiation of clade Metazoa.

Experimental Procedures

Rbm45 Nucleic Acid and Amino Acid Sequences

All Rbm45 sequence information (genomic; cDNA; amino acid) in this study was retrieved from the National Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/). *Rbm45* orthologues [48] were identified using the BLAST algorithm by querying GenBank with the *Homo sapiens* (human; Gene ID:129831) *RBM45* cDNA, accession number (no.) NM_152945.3 [26]. The

following accession nos. were utilized in this study: phylum Porifera: *Amphimedon queenslandica* (sponge: NW_003546244.1; XM_003382583.3; XP_003382631.1); phylum Brachiopoda: *Lingula anatine* (lamp shell: NW_019775657.1; XM_013560776.1; XP_013416230.1); phylum Priapulida: *Priapulus caudatus* (priapulid worm: NW_014578398.1; XM_014820871.1; XP_014676357.1); phylum Hemichordata: *Saccoglossus kowalevskii* (acorn worm: NW_003156738.1; XM_006825506.1; XP_006825569.1); phylum Echinodermata: *Acanthaster planci* (crown-of-thorns starfish: NW_019091356.1; XM_022231661.1; XP_022087353.1), *Strongylocentrotus purpuratus* (purple sea urchin: NW_011995289.1; XM_780088.4; XP_785181.2); phylum Cnidaria: *Exaiptasia pallida* (pale anemone: NW_018384422.1; XM_021061761.1; XP_020917420.1), *Hydra vulgaris* (freshwater polyp: NW_004169798.1; XM_012704848.1; XP_012560302.1), *Stylophora pistillata* (smooth cauliflower coral: NW_019217956.1; XM_022936599.1; XP_022792334.1), *Orbicella faveolata* (mountainous star coral: NW_018149652.1; XM_020763593.1; XP_020619252.1), *Acropora digitfera* (staghorn coral: NW_015441794.1; XM_015923864.1; XP_015779350.1); phylum Mollusca: *Octopus bimaculoides* (California two-spot octopus: NW_014653530.1; XM_014915327.1; XP_014770813.1), *Aplysia californica* (California sea hare: NW_004797628.1; XM_013085237.1; XP_012940691.1), *Crassostrea gigas* (Pacific oyster: NW_011936719.1; XM_011449747.2; XP_011448049.1); phylum Arthropoda: *Limulus polyphemus* (Atlantic horseshoe crab: NW_013677256.1; XM_013937428.2; XP_013792882.2), *Bactrocera dorsalis* (oriental fruit fly: NW_011876306.1; XM_011202560.2; XP_011200862.1), *Drosophila melanogaster* (fruit fly: NT_037436.4; NM_139630.3; NP_647887.1), *Aedes aegypti* (yellow fever mosquito: NC_035108.1; XM_001660954.2; XP_001661004.2); phylum Chordata: *Branchiostoma floridae* (Florida lancelet: NW_003101314.1; XM_002586844.1; XP_002586890.1), *Latimeria chalumnae* (West Indian Ocean coelacanth: NW_005822196.1; XM_006012287.1; XP_006012349.1), *Callorhinchus milii* (Australian ghost shark: NW_006890067.1; XM_007890069.1; XP_007888260.1), *Rhincodon typus* (whale shark: NW_018067517.1; XM_020531702.1; XP_020387291.1), *Orcinus orca* (killer whale: NW_004438435.1; XM_004267367.2; XP_004267415.1), *Lipotes vexillifer* (Yangtze River dolphin: NW_006776904.1; XM_007450717.1; XP_007450779.1), *Chelonia mydas* (green sea turtle: NW_006631183.1; XM_027822939.1; XP_027678740.1), *Alligator mississippiensis* (American alligator: NW_017714267.1; XM_019480600.1; XP_019336145.1), *Xenopus laevis* (African clawed frog: NC_030741.1; NM_001086621.1; NP_001080090.1), *Danio rerio* (Zebrafish: NC_007120.7; NM_001127402.1; NP_001120874.1), *Ornithorhynchus anatinus* (platypus: NW_001794453.1; XM_016228192.1; XP_016083678.1), *Monodelphis domestica* (gray short-tailed opossum: NC_008804.1; XM_007494431.2; XP_007494493.1), *Canis lupus familiaris* (dog: NC_006618.3; XM_022414739.1; XP_022270447.1), *Rattus norvegicus* (Norway rat: NC_005102.4; NM_153306.1; NP_695218.1), *Mus musculus* (house mouse: NC_000068.7; NM_153405.2; NP_700454.1), *Gallus gallus* (chicken: NC_006094.5; NM_001031252.1; NP_001026423.1), *Loxodonta africana* (African savanna elephant: NW_003573423.1; XM_023543049.1; XP_023398817.1), *Gorilla gorilla* (western gorilla: NC_018426.2; XM_019021614.1; XP_018877159.1), *Pan troglodytes* (common chimpanzee: NC_036881.1; XM_515938.6; XP_515938.2), *Homo sapiens* (human: NC_000002.12; NM_001365578.1; NP_001352507.1).

Molecular Phylogenetic Trees

All molecular phylogenetic trees were created using Molecular Evolutionary Genetic Analysis v.7.0 software (MEGA7) [49]. cDNA sequence alignments were performed using ClustalW within the MEGA7 program using these default parameters: Pairwise Alignment-Gap Opening Penalty: 15, Gap

Extension Penalty: 6.66; Multiple Alignment-Gap Opening Penalty: 15, Gap Extension Penalty: 6.66; DNA Weight Matrix: IUB; Transition Weight: 0.5; Use Negative Matrix: OFF; Delay Divergent Cutoff (%): 30. Molecular phylogenetic tree creation with Rbm45 cDNA sequences used the following parameters: Statistical Method: Maximum Likelihood [50]; Test of Phylogeny: Bootstrap method [51], No. of Bootstrap Replications: 100 and 1000; Model: Tamura-Nei model; Rate among Sites: Uniform rates; Gaps/Missing Data Treatment: Use all sites; ML Heuristic Method: Nearest-Neighbor-Interchange; Initial Tree for ML: NJ/BioNJ; Codons Included: 1st+2nd+3rd+Non-Coding.

Amino acid sequence alignments were executed using ClustalW within the MEGA7 program and used the following default parameters: Pairwise Alignment-Gap Opening Penalty: 10, Gap Extension Penalty: 0.1; Multiple Alignment-Gap Opening Penalty: 10, Gap Extension Penalty: 0.2; Protein Weight Matrix: Gonnet; Residue-specific Penalties: ON; Hydrophilic Penalties: ON; Gap Separation Distance: 4; End Gap Separation: OFF; Genetic Code Table: Standard; Use Negative Matrix: OFF; Delay Divergent Cutoff (%): 30. Molecular phylogenetic tree creation with Rbm45 amino acid sequences used the following parameters: Statistical Method: Maximum Likelihood [52]; Test of Phylogeny: Bootstrap method [51], No. of Bootstrap Replications: 100 and 1000; Model: Jones- Taylor-Thornton (JTT) model; Rate among sites: Uniform rates; Gaps/Missing Data Treatment: Use all sites; ML Heuristic Method: Nearest-Neighbor-Interchange; Initial Tree for ML: NJ/BioNJ.

Rbm45 Orthologue Protein Domain Conservation

Multiple sequence alignments were performed using Clustal Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/) [53-55] with the following parameters: Output Format: ClustalW with character counts; Dealign input sequences: No; Mbed-like clustering guide: Yes; Mbed-like clustering: Yes; Number of combined: default (0); Max guide tree iterations: Default; Max HMM Iteration: Default; Order: input.

NCBI gene annotation (https://www.ncbi.nlm.nih.gov/homologene/?term=Rbm45 [accessed 2024 February 4]), our previous work [26], and data from Tamada *et al.* [27] reveal that Rbm45 contains four RBDs. However, ensuing work by two leading Rbm45 research teams show an alternative structure with three RBDs [31,32]; for consistency, we have adopted these research team's Rbm45 domain nomenclature for this study. RBDs I, II, and III, HOA domain, NES, and the monopartite NLS of Rbm45 were identified by visual inspection after performing multiple sequence alignment using human RBM45 RBDs, HOA, NES, and NLS sequences as reference [21,26,32]. Additionally, the Locating Nuclear Export Signals (LocNES) algorithm was used to extend the analysis of NES sequences [56]. Orthologous *Rbm45* exon and intron sequence lengths were retrieved from NCBI Genome (https://www.ncbi.nlm.nih.gov/genome/).

Rbm45 Orthologue Gene Architecture Evolution

The Mean Exon Size versus Exon Number and Age of Taxonomic Lineage versus Exon Number were calculated for the following representative animals across metazoan taxa (common name; exon number; mean exon size in base pairs; approximate lineage age in years): phylum Porifera: *Amphimedon queenslandica* (sponge; 2; 984; 650,000,000 [57]); phylum Brachiopoda: *Lingula anatine* (lamp shell; 11; 238; 66,000,000 [58]); phylum Hemichordata: *Saccoglossus kowalevskii* (acorn worm; 9; 234; 373,000,000 [59]); phylum Echinodermata: *Acanthaster planci* (crown-of-thorns starfish; 11; 404; 4,000,000 [60]), *Strongylocentrotus purpuratus* (purple sea urchin; 10; 183; 199,000,000 [61]); phylum Cnidaria: *Exaiptasia pallida* (pale anemone; 4; 544; 500,000,000 [62]), *Hydra vulgaris* (freshwater polyp; 3; 420; 540,000,000 [63]), *Orbicella faveolata* (mountainous star coral; 3; 937; 570,000,000 [64]); phylum Mollusca: *Octopus bimaculoides* (California two-spot octopus; 10; 219;

155,000,000 [65]), *Crassostrea gigas* (Pacific oyster; 13; 195; 15,000,000 [66]); phylum Arthropoda: *Limulus polyphemus* (Atlantic horseshoe crab; 6; 206; 250,000,000 [67]), *Drosophila melanogaster* (fruit fly; 9; 174; 5,400,000 [68]); phylum Chordata: *Branchiostoma floridae* (Florida lancelet; 7; 154; 100,000,000 [69]), *Latimeria chalumnae* (West Indian Ocean coelacanth; 10; 181; 70,000,000 [70]), *Orcinus orca* (killer whale; 10; 182; 11,000,000 [71]), *Lipotes vexillifer* (Yangtze River dolphin; 10; 170; 11,000,000 [71]), *Alligator mississippiensis* (American alligator; 10; 210; 53,000,000 [72]), *Xenopus laevis* (African clawed frog; 11; 173; 18,000,000 [73]), *Danio rerio* (Zebrafish; 10; 199; 150,000,000 [74]), *Monodelphis domestica* (gray short-tailed opossum; 10; 196; 3,000,000 [75]), *Rattus norvegicus* (Norway rat; 10; 178; 2,000,000 [76]), *Mus musculus* (house mouse; 10; 190; 6,000,000 [77]), *Gallus gallus* (chicken; 10; 180; 21,000,000 [78]), *Pan troglodytes* (common chimpanzee; 10; 181; 8,000,000 [79]), *Homo sapiens* (human; 10; 149; 300,000 [80]). Scatter plots and regression analysis (R^2: Coefficient of Determination) were generated in Microsoft Excel.

Results and Discussion

Molecular Phylogenetics of Rbm45 Orthologues

Our lab [26] has previously demonstrated that nucleotide and amino acid phylograms of 10 vertebrate *Rbm45* orthologues recapitulate accepted taxonomic relationships between classes Actinopterygii (ray-finned fishes), Amphibia, Reptilia, and Mammalia; additionally, through NCBI database interrogation, we reported *Rbm45* orthologues, both empirically confirmed and Gnomon algorithm predicted (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/process/), across metazoan taxa including animals from the non-Bilateria phyla Porifera (sponges) and Cnidaria (e.g., hydra), and within clade Bilateria phyla from the nephrozoan lineages Protostomia (e.g., phyla Mollusca and Arthropoda) and Deuterostomia (e.g., phyla Echinodermata and Chordata). Molecular phylogenetic analysis has been successfully used to reconstruct the evolutionary history of populations, genes, and proteins; furthermore, it has been utilized to understand genome organization and gene conservation [81,82]. Therefore, to gain a better understanding of the evolutionary history of *Rbm45*, a gene involved in neuronal development [27] and neuronal pathogenesis [29], we have expanded our phylogenetic analysis to 36 *Rbm45* orthologues from 9 phyla: Porifera, Cnidaria, Brachiopoda, Mollusca, Arthropoda, Echinodermata, Priapulid, Hemichordata, and Chordata. When more than one organism's *Rbm45* sequence was available within a phylum, we often chose those species that are on the IUCN Red List of Threatened Species (e.g., *Lipotes vexillifer* [Yangtze River dolphin] [83], *Loxodonta africana* [African savanna elephant] [84]), are molecular genetic and developmental biology model systems (e.g., *Danio rerio* [Zebrafish], *Xenopus laevis* [African clawed frog], *Gallus gallus* [chicken], and *Mus musculus* [house mouse]), or give us multiple subgroups in a taxonomic unit (e.g., class Mammalia: order Monotremata [*Ornithorhynchus anatinus*: platypus], infraclass Marsupialia [*Monodelphis domestica*: gray short-tailed opossum], and infraclass Placentalia [*Homo sapiens*: humans]). Additionally, we attempted to have representative members from a variety of crown clades [85]: Ecdysozoa (phyla Priapulida and Arthropoda), Spiralia (phyla Brachiopoda and Mollusca), Ambulacraria (phyla Echinodermata and Hemichordata), and Chordata (subphylum Craniata) [86,87].

We used MEGA7 software (Materials and Methods) to build rooted phylogenetic trees using 36 Rbm45 cDNA and amino acid orthologous sequences. The bootstrap consensus tree from 1000 iterations is taken to represent the evolutionary history of the gene [51]. In our unbiased tree analysis (data not shown), phylum Porifera resolved as the sister group to all other animals; therefore, phylum Porifera
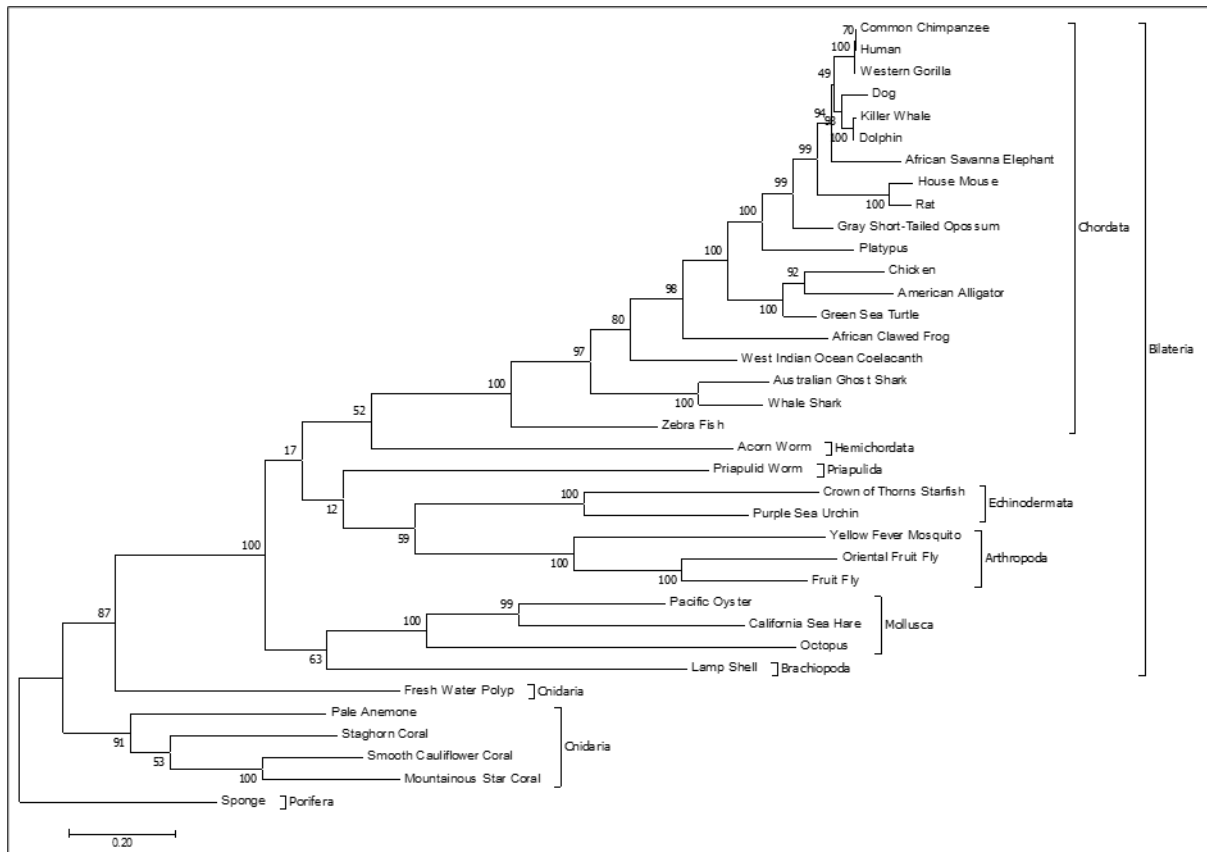
Figure 1. Rbm45 cDNA molecular phylogeny. Evolutionary analysis of 36 Rbm45 orthologous cDNA sequences across metazoan taxa conducted in MEGA7. The evolutionary history was inferred by using the Maximum Likelihood method based on the Tamura-Nei model [50]. The tree with the highest log likelihood (-82723.91) is shown. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed [51]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood approach, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. There were a total of 8510 positions in the final dataset. Phyla are indicated by the inner brackets. The monophyletic clade Bilateria is indicated by the outside bracket. The tree is rooted on phylum Porifera.

served as the outgroup [88-90] in the cDNA and amino acid molecular phylogenies (Figs. 1 and 2). As we progress "up" the tree from most ancient to most recent lineages, phylum Cnidaria exhibits paraphyly, with the freshwater polyp (*Hydra vulgaris*) diverging after corals and anemones as sister group to Bilateria, having 87% and 97% bootstrap support for the evolutionary node in the cDNA and amino acid molecular phylogenies, respectively (Figs. 1 and 2). These data are an example of incomplete lineage sorting, where a gene tree does not match the history of the taxa [82,91], within clade Cnidaria [92]. Where we have sequence from more than one organism in a phylum, the phyla are monophyletic in both the cDNA and amino acid molecular phylogeny (e.g., Mollusca, Arthropoda, Echinodermata, and Chordata). We also observed incomplete lineage sorting in the cDNA molecular
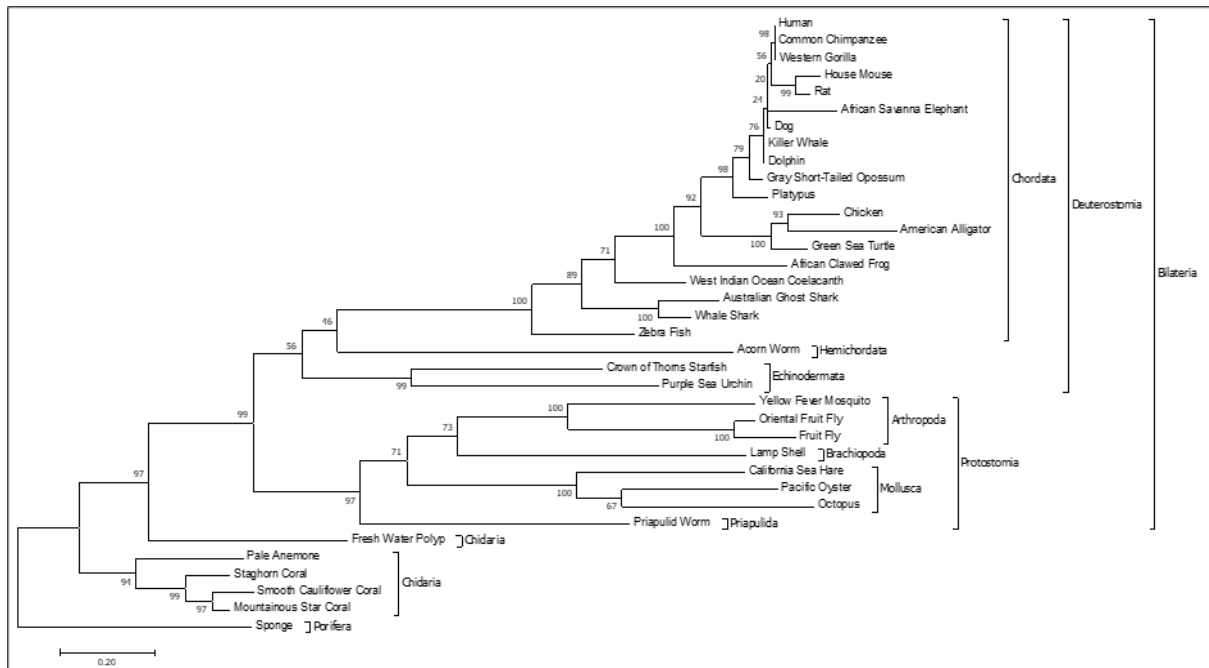
Figure 2. Rbm45 protein molecular phylogeny. Evolutionary analysis of 36 Rbm45 orthologous amino acid sequences across metazoan taxa conducted in MEGA7. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model [52]. The tree with the highest log likelihood (-22250.34) is shown. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed [51]. Branches corresponding to partitions reproduced in less than 50% bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are shown next to the branches. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. There were a total of 930 positions in the final dataset. Phyla are indicated by the inner brackets. The monophyletic clades Protostomia, Deuterostomia, and Bilateria are indicated by the outside brackets. The tree is rooted on phylum Porifera.

phylogeny with clades Protostomia and Deuterostomia exhibiting paraphyly (Fig. 1). However, since any bootstrap value less than 70 is considered unreliable [93,94], this node, at 17% bootstrap support, splitting the cluster taxa Chordata and Hemichordata away from Priapulida, Echinodermata, Arthropoda, Mollusca, and Brachiopoda, is not well supported. However, the node between non-bilaterians and bilaterians has 87% bootstrap support. In contrast, the amino acid molecular phylogeny (Fig. 2) shows Protostomia and Deuterostomia as monophyletic clades with 99% bootstrap support of the node at bilaterian diversification. Like the cDNA molecular phylogeny, the node at the diversification of non-bilaterians and bilaterians has 97% bootstrap support in the amino acid molecular phylogeny (Figs. 1 and 2).

Both the cDNA (Fig. 1) and the amino acid (Fig. 2) molecular phylogenies show phylum Hemichordata as the sister group to phylum Chordata. However, in contrast to recent work [59], neither phylogeny places Hemichordata with Echinodermata (i.e., they are paraphyletic) into clade Ambulacraria, the sister group to clade Chordata, albeit these nodes have weak bootstrap support of 52% and 46%,

respectively. Conversely, in the amino acid molecular phylogeny (Fig. 2), Echinodermata and Hemichordata, though paraphyletic, do form basal groups to Chordata as expected [59,86,87]. Interestingly, both the cDNA and amino acid molecular phylogenies place Zebrafish (*Danio rerio*; group Osteichthyes; class Actinopterygii), with 100% bootstrap support, as ancestral to all other chordates (subphylum Craniata) including cartilaginous fish (class Chondrichthyes). These data are in opposition to accepted cladograms based on morphological and molecular traits [87,95], but support a hypothesis that sharks evolved from a common ancestor having a bony skeleton, with generalized bone loss as a synapomorphy for class Chondrichthyes [96]. The coelacanth (*Latimeria chalumnae*; class Sarcopterygii) has 80% (Fig. 1) and 71% (Fig. 2) bootstrap support as a sister group to clade Tetrapoda in agreement with the current hypothesis on the evolution of the tetrapod lineage [97]. Similarly, our phylogenies have 100% (Fig. 1) and 92% (Fig. 2) bootstrap support for clade Amniota.

Protostomes are separated into clades Spiralia and Ecdysozoa [86]. Within the Rbm45 amino acid molecular phylogeny (Fig. 2), which delineates clade Protostomia as monophyletic, all nodes, except for the split between octopus and Pacific oyster within phylum Mollusca, have greater than 70% bootstrap support. However, we observed incomplete lineage sorting of Rbm45 within Protostomia as clades Ecdysozoa and Spiralia do not exhibit monophyly. Instead, the taxa that make up these two clades are interspersed within the protostomes. The phylogenetic incongruences (i.e., conflicting branching orders) and lack of coalescence between our Rbm45 molecular phylogenies and currently accepted species phylogeny is not surprising [82,98] and is most likely a product of coalescent stochasticity [82,99,100] and variations in lineage-specific evolutionary rates over time (heterotachy) [101,102]. In contrast to other published data [103], we found that our amino acid molecular phylogeny recapitulated accepted species tree topology more closely than the cDNA molecular phylogeny [104]. Additionally, we have observed that where there is certainty in taxa resolution (e.g., clades Tetrapoda and Amniota), our data also appropriately resolves these nodes; whereas, where there is uncertainty in the placement of taxa (e.g., polytomy of clade Spiralia), our molecular phylogenies also reflect similar difficulties in node placement consistent with previous work by others [105].

Unfortunately, we were unable to include in our analysis two basal groups, phyla Placozoa and Ctenophora, whose positions in the metazoan phylogeny are disputed [106,107]. The predicted sequence [108] for the placozoan *Trichoplax adhaerans* Rbm45 orthologue (Gene ID: 6752484; NW_002060945.1, XM_002110678.1, XP_002110714.1; accessed 2024 February 4) is a partial sequence predicted to have at least three exons encoding 274 amino acids roughly corresponding to amino acids 21-317 (including gaps; data not shown) of human RBM45. Therefore, the hypothetical *T. adhaerans* Rbm45 orthologue sequence lacks a start codon, RNA-binding domain III, homo-oligomer assembly domain, nuclear localization signal, and a stop codon. However, the presence of at least 3 exons supports the assertion that placozoans are more derived than their simple body plan would imply [106,108] (Fig. 3). Inclusion of the *T. adhaerans* partial Rbm45 amino acid sequence into our molecular phylogenetic analysis placed it as a sister group to phyla Porifera and Cnidaria as expected [86,87] (data not shown). The nuclear genomes of six species from phylum Ctenophora have been sequenced to date (https://www.ncbi.nlm.nih.gov/search/all/?term=Ctenophora; accessed 2024 February 4); however, an Rbm45 orthologue has not yet been identified. This is unfortunate, as we wished to include this gene product in our analysis considering the hypothesis [107,109] that ctenophores are the sister group to all animals, and not poriferans, challenging the predominant paradigm that metazoan nervous systems have evolved complexity in a stepwise manner over deep time [110].
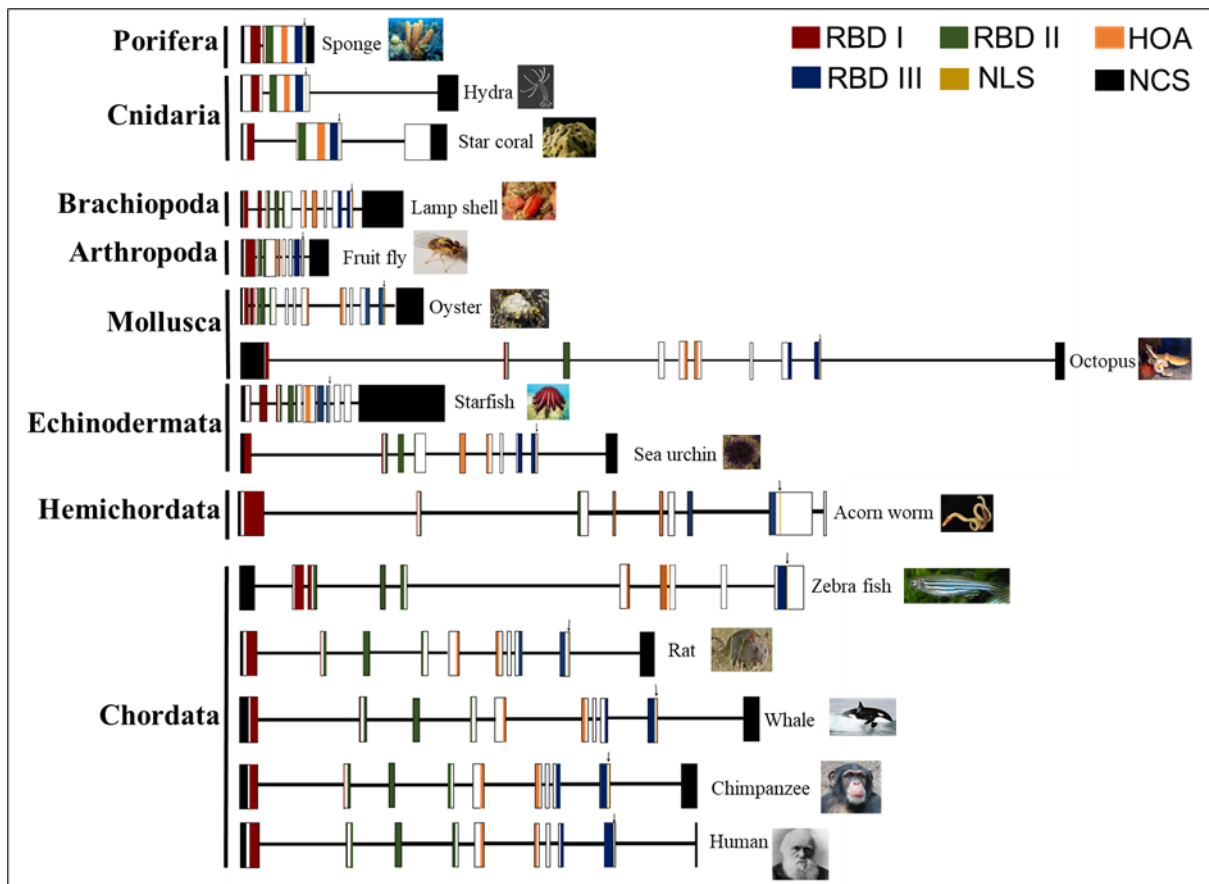
Figure 3. Gene architecture of Rbm45 orthologues. Rbm45 protein domains are linearly conserved from sponges to humans. Schematic diagram of the exon-intron structure from 15 representative Rbm45 orthologues across metazoan taxa. Vertical boxes represent exons, and the solid horizontal lines represent introns. The diagram shows the relative sizes of the exons and introns. The introns and exons are scaled by type; the intron width scale is half that of the exon width scale. The DNA sequences encoding RNA-binding domains (RBD) I, II, and III, and the homo-oligomer assembly (HOA) domain, are indicated by red, green, blue, and orange fill, respectively. The DNA sequences encoding the monopartite nuclear localization signal (NLS) are indicated by a vertical yellow line and a downward pointing arrow. Non-coding exon sequences (NCS) are shaded black. All animal images were retrieved from Wikimedia (commons.wikimedia.org) and are public domain (Creative Commons license CC0: https://creativecommons.org/).

Rbm45 Orthologue Protein Domain Conservation

Considering the strong recapitulation of the metazoan tree of life using Rbm45 amino acid sequences, we also analyzed the conservation of protein domains from sponges to humans across the same 36 taxa used in the phylogenetic analysis (Figs. 1 and 2). Multiple sequence alignment using the Clustal Omega algorithm demonstrates that Rbm45 protein domains are linearly conserved from sponges to humans in the order: RBD I, RBD II, HOA, RBD III, and NLS (Fig. 3). RBDs I, II, and III show 72%, 80%, and 65 % similarity across taxa, respectively, with the HOA domain displaying 68% similarity across taxa analyzed (data not shown). Furthermore, the monopartite NLS [32,111], located just downstream of RBD III (Fig. 3), from all 36 orthologues conforms to the canonical consensus sequence K(R/K)(X)(R/K). Example NLSs are sponge KRQK, sea anemone KRPR, mosquito KRMR, and human KRQR (data not shown) exemplifying core basic amino acids.

In contrast to the NLS, the human RBM45 NES does not conform to a classical leucine-rich domain but instead was empirically determined [32] to be made up of a clique of two hydrophobic amino acids: leucine-leucine. Additionally, this clique of two hydrophobic residues L(L/I) are conserved among Rbm45 mammalian orthologues [27,32]. Using multiple sequence alignment (Materials and Methods), we identified the L(L/I) clique only within clade Tetrapoda of clade Craniata. To identify an NES in non-tetrapods, we interrogated the online NES prediction program LocNES [56] using human RBM45 as the input sequence. The algorithm identified a majority-rule canonical sequence from all 19 Craniata taxa (i.e., clade Tetrapoda, clade Chondrichthyes, and group Osteichthyes) between the HOA and RBD III domains [31,32] having the general form: $R^{15}K^{16}MA(T^{14}/S^2)Q(M^{12}/L^7)VA^{16}A^{18}Q^{18}(L^{11}/M^5/V^3)(A^{11}/M^4)(S^{18}/T^1)(M^{15}/V^3)$ where the superscript indicates the number of taxa with that amino acid, if not at identity, with conserved amino acids (e.g., hydrophobic) grouped in parentheses. In tetrapods, this conserved sequence was downstream and immediately adjacent to the L(L/I) clique. In contrast, using sponge or sea anemone Rbm45 as the query sequence, we were unable to deduce a consensus NES in invertebrates (clades Ambulacraria and Protostomia). These data agree with a pairwise alignment of sponge and human Rbm45 amino acid sequences where the Clustal Omega algorithm inserted a gap in the sponge sequence across from the human NES (data not shown). The inability to identify a putative NES in the invertebrate taxa analyzed is not necessarily surprising given the complexity and variability of NES sequences [56,112]. Furthermore, the LocNES algorithm searched for CRM1 dependent NESs [56]. It is possible that invertebrate Rbm45 proteins use a CRM1-independent pathway (e.g., passive diffusion) like vertebrate TDP-43 and FUS [113]. If invertebrate Rbm45 exits the nucleus by passive diffusion, then the evolution of an NES would be a novel trait in the Craniata lineage. Future work on this question would necessitate the empirical determination of whether Rbm45 is able to be trafficked from the nucleus to the cytoplasm in these invertebrate organisms.

Rbm45 Orthologue Gene Architecture Evolution

Concurrent with protein domain conservation analysis, we also analyzed the gene structure of 25 *Rbm45* orthologues (Materials and Methods). *Rbm45* from the non-bilaterian phylum Porifera has 2 large exons, and from phylum Cnidaria hydra and star coral have 3 exons (Fig. 3) and the pale anemone has 4 exons, including a cryptic exon (data not shown). Interestingly, all protein domains are conserved in the first two exons of cnidarian *Rbm45* with a similar spacing to sponge (Fig. 3 and data not shown). The bilaterian phyla have between 6 and 13 exons, with the Rbm45 protein domains fragmented across multiple exons (Fig. 3 and data not shown). Unsurprisingly, though the localization of protein domains occurs in different exons depending on the organism, within each phylum the domains are localized in very similar exon positions. This phenomenon is especially evident in class Mammalia of phylum Chordata where the domain distribution among exons is almost identical (Fig. 3 and data not shown) between taxa. Furthermore, regression analysis of the 25 *Rbm45* orthologues revealed a statistically significant strong negative correlation between mean exon length and total number of exons ($R^2 = 0.6169$, $P < 0.0001$; Fig. 4) in accord with the work of others [114]. Our data are in agreement with earlier work demonstrating that more evolutionarily advanced organisms, as measured by genomic and metabolomic complexity [114-117], have more short exons and longer introns while less evolutionary advanced organisms have fewer large exons and short introns (Fig. 3 and data not shown) consistent with whole genome analysis across the five kingdoms [118] Protozoa, Chromista, Plantae, Fungi, and Animalia [114].

We extended our analysis of the evolution of *Rbm45* gene complexity by examining the correlation of a
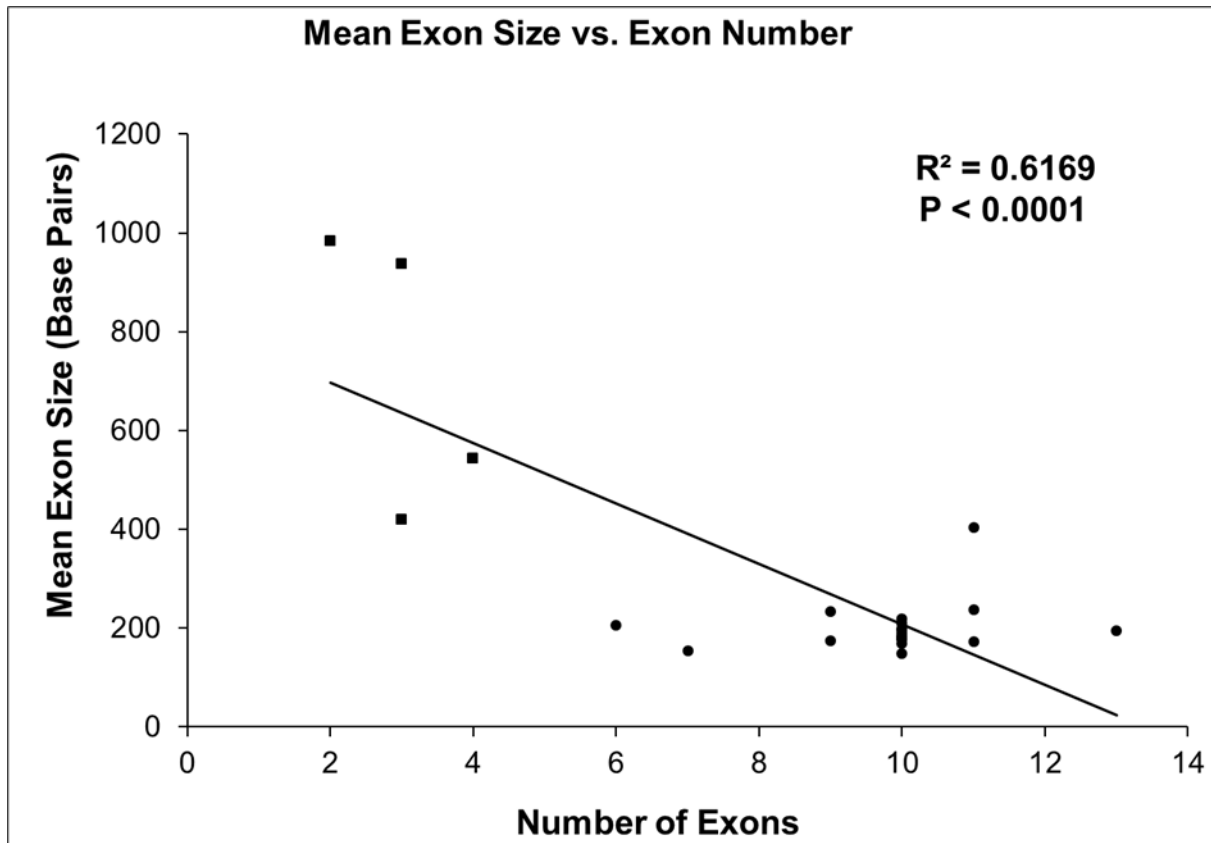
Figure 4. Regression analysis of mean exon length versus number of exons in Rbm45 orthologues. Rbm45 mean exon length decreases as exon number increases. A scatter plot, with linear regression, was produced from 25 representative Rbm45 orthologues. There is a statistically significant correlation (Coefficient of Determination: R^2) between mean exon length and the number of exons in the organism's gene structure. Closed squares (■) are non-bilaterian (i.e., diploblastic) animals of phyla Porifera and Cnidaria; closed circles (•) are members of clade Bilateria (i.e., triploblastic) from phyla Arthropoda, Brachiopoda, Chordata, Echinodermata, Hemichordata, and Mollusca [86].

taxon's approximate lineage age, as determined by a robust analysis of current literature (Materials and Methods), to the number of exons in the *Rbm45* orthologue. We demonstrate a statistically significant very strong negative correlation ($R^2 = 0.8057$, *P* < 0.0001; Fig. 5) where the most ancient lineages (e.g., Porifera at 650,000,000 years; Cnidaria at 500-570,000,000 years; see Materials and Methods) have the fewest number of exons, between 2 and 4, while generally more recent taxa lineages have more exons (> 9 exons; e.g., Zebrafish at 150,000,000 years; crown-of-thorns starfish at 4,000,000 years; humans at 300,000 years; see Materials and Methods). The majority of taxa followed the trend of fewer exons correlating to an ancient lineage age and more exons to a more recent lineage age. A notable exception is the acorn worm, *Saccoglossus kowalevskii*, from phylum Hemichordata (clade Ambulacraria) which has an approximate lineage age of 370,000,000 years (Upper Devonian) [59]. The reference genomic sequence (NW_003156738.1) of *S. kowalevskii* predicts 9 exons which is closer in number to what is observed for other members of clade Bilateria we analyzed. However, all other bilaterian taxa that we used in Fig. 5 had lineage ages of less than 250,000,000 years (e.g., horseshoe crab with 6 exons). A rigorous analysis of the gene structure of all 36 *Rbm45* orthologues used in this study revealed that the *Priapulis caudatus* (phylum Priapulida; clade Scalidophora) *Rbm45* orthologue reference genomic
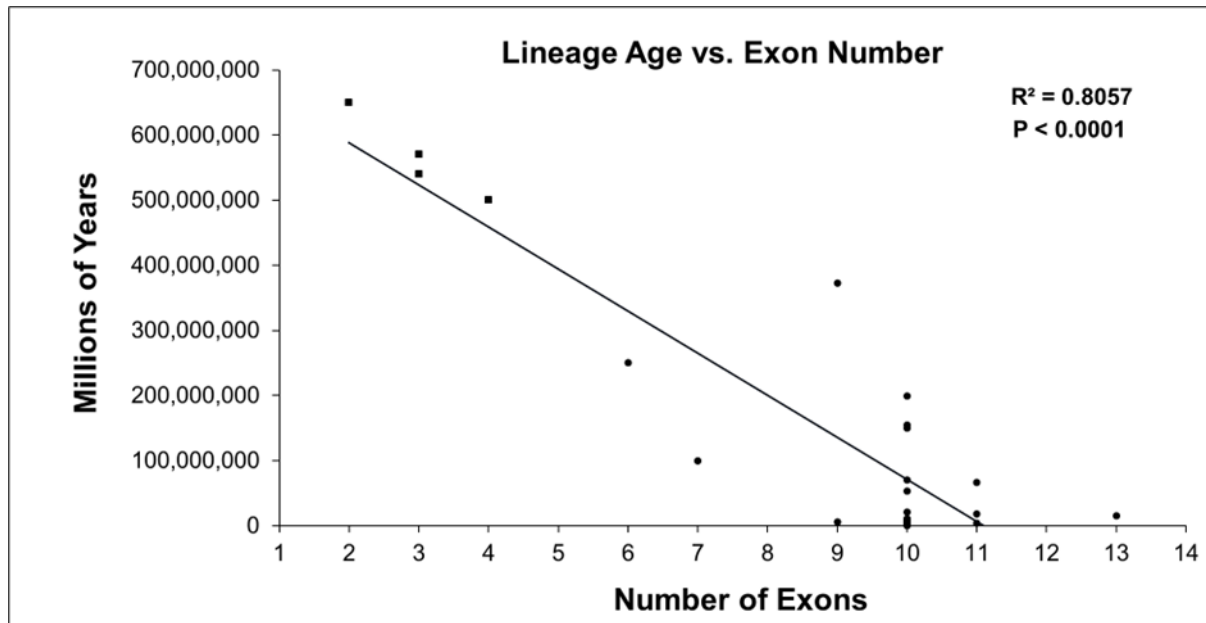
Figure 5. Regression analysis of species lineage age versus number of exons in Rbm45 orthologues. Rbm45 exon number increases as lineage age decreases. A scatter plot, with linear regression, was produced from 25 representative Rbm45 orthologues. There is a statistically significant high correlation (Coefficient of Determination: $R^2 \geq 0.7$) between lineage age and the number of exons in the organism's gene structure. Closed squares (■) are non-bilaterian (i.e., diploblastic) animals of phyla Porifera and Cnidaria; closed circles (•) are members of clade Bilateria (i.e., triploblastic) from phyla Arthropoda, Brachiopoda, Chordata, Echinodermata, Hemichordata, and Mollusca [86].

sequence (NW_014578398.1) is predicted to have 17 exons (data not shown), the most of any organism examined by us. Extant priapulins date from the late Carboniferous (~350,000,000 years ago) [119], while extinct stem- and crown-group priapulins are found in the middle Cambrian (~500,000,000 years ago) [120,121]. These data suggest that division of the *Rbm45* gene into many exons (i.e., > 4) occurred relatively early in the adaptive radiation of the evolutionary complex bilateral body plan during and after the Cambrian period [122,123]. Taken together, these data indicate an ancient origin for *Rbm45* in the metazoan lineage. In accordance with this observation, we were able to identify an *Rbm45* orthologue in *Monosiga brevicollis* [(Monbr1│Name: e_gw1.8.135.1; Protein ID: 16822; Location:scaffold_8:72832-74415); https://mycocosm.jgi.doe.gov/cgi-bin/dispGeneModel? db=Monbr1&id=16822] from phylum Choanoflagellata, clade Holozoa. These flagellated protists are hypothesized to be the ancestors of phyla Porifera, and thus all metazoans [124-126], further demonstrating the ancient roots of *Rbm45*. Consistent with this well-accepted hypothesis [86], an unrooted phylogeny using amino acid sequence data from the 36 Rbm45 orthologues from this study plus the Rbm45 orthologue from Choanoflagellata places choanoflagellates at the root of the tree as the sister to all animals (data not shown). This high level of conservation among crown clades reveals that *Rbm45* may play a role in neurogenesis across metazoans.

Conclusions

We have demonstrated that *Rbm45* is an ancient gene conserved from clade Holozoa to clade Metazoa with phylogenetic analysis of Rbm45 orthologue amino acid sequence mirroring known monophyletic relationships among metazoans. Additionally, we have verified and extended the observation of deep

OPEN ACCESS

homology of the RBD, HOA, and NLS regions in the Rbm45 protein from sponges to chordates, whereas the NES is a possible synapomorphy unique to clade Craniata. Furthermore, we have shown a statistically significant increase in complexity of *Rbm45* gene architecture contemporaneous with increasing evolutionary complexity moving from non-bilaterian to bilaterian animals over evolutionary time. Elucidation of Rbm45 function in neural development and homeostasis in a broad range of molecular genetic model systems/taxa will merit detailed attention in the future to holistically understand its function in a breadth of neural/sensory networks.

Author contributions

Conceptualization: J.O.H.; Formal Analysis: V.V., A.C., and J.O.H.; Investigation: V.V., T.N.M., A.C., L.M.S., and J.O.H.; Supervision: J.O.H.; Visualization: V.V., T.N.M., A.C., and J.O.H.; Writing – original draft: J.O.H.; Writing – reviewing and editing: V.V., T.N.M., A.C., and J.O.H.

Competing interests

The authors declare no competing interests.

Abbreviations

Rbm45 - RNA-binding motif protein 45; RBP - RNA-binding protein; RRM - RNA recognition motif; RBDP - RNA recognition motif-type binding domain protein; RBD - RNA-binding domain; HOA - homo-oligomer assembly; NLS - nuclear localization signal; NES - nuclear export signal; NCBI - National Center for Biotechnology Information; no - number; MEGA7 - Molecular Evolutionary Genetic Analysis v.7.0 software; JTT - Jones-Taylor-Thornton model; LocNES - Locating Nuclear Export Signals algorithm; IUCN - International Union for Conservation of Nature; NCS - non-coding exon sequence

References

1. Gerstberger, S., Hafner, M., Ascano, M., & Tuschl, T. (2014) Evolutionary conservation and expression of human RNA-binding proteins and their role in human genetic disease. DOI: 10.1007/978-1-4939-1221-6_1. *Adv. Exp. Med. Biol.* 825,1–55.

2. Abe, R., Sakashita, E., Yamamoto, K., & Sakamoto, H. (1996) Two different RNA binding activities for the AU-rich element and the poly(A) sequence of the mouse neuronal protein mHuC. DOI: 10.1093/nar/24.24.4895. *Nucleic Acids Res.* 24(24),4895-901.

3. Anant, S., Henderson, J.O., Mukhopadhyay, D., Min, J., Kennedy, S.M., *et al.* (2001) Novel role for RNA binding protein CUGBP2 in mammalian mRNA editing: CUGBP2 modulates C to U editing of apolipoprotein B mRNA by interacting with apobec-1 and ACF, the apobec-1

complementation factor. DOI: 10.1074/jbc.M104911200. *J. Biol. Chem.* 276(50),47338-51.

4. Joseph, B., Orlian, M., & Furneaux, H. (1998) p21 (waf1) mRNA contains a conserved element in its 3′-untranslated region that is bound by the Elav-like mRNA-stabilizing proteins. DOI: 10.1074/jbc.273.32.20511. *J. Biol. Chem.* 273(32),20511-6.

5. Mukhopadhyay, D., Houchen, C.W., Kennedy, S., Dieckgraefe, B.K., & Anant, S. (2003) Coupled mRNA stabilization and translational silencing of cyclooxygenase-2 by a novel RNA binding protein, CUGBP2. DOI: 10.1016/s1097-2765(03)00012-1. *Mol. Cell.* 11(1),113-26.

6. Anant, S., Murmu, N., Houchen, C.W., Mukhopadhyay, D., Riehl, T.E., *et al.* (2004) Apobec-1 protects intestine from radiation injury through posttranscriptional regulation of cyclooxygenase-2 expression. DOI: 10.1053/j.gastro.2004.06.022. *Gastroenterology,* 124(4),1139-49.

7. Antic, D., Lu, N., & Keene, J.K. (1999) ELAV tumor antigen, Hel-N1, increases translation of neurofilament M mRNA and induces formation of neuritis in human teratocarcinoma cells. DOI: 10.1101/gad.13.4.449. *Genes Dev.* 13(4),449-61.

8. Imai, T., Tokunga, A., Yoshida, T., Hashimoto, M., Mikoshiba, K., *et al.* (2001) The neural RNA-binding protein Musashi1 translationally regulates mammalian numb gene expression by interacting with its mRNA. DOI: 10.1128/MCB.21.12.3888-3900.2001. *Mol. Cell. Biol.* 21(12),3888-900.

9. Okabe, M., Imai, T., Kurusu, M., Hiromi, Y., & Okano, H. (2001) Translational repression determines a neuronal potential in *Drosophila* asymmetric cell division. DOI: 10.1038/35075094. *Nature.* 411(6833),94-8.

10. Sureban, S.M., Murmu, N., Rodriguez, P., May, R., Maheshwari, R., *et al.* (2007) Functional antagonism between RNA binding proteins HuR and CUGBP2 determines the fate of COX-2 mRNA translation. DOI: 10.1053/j.gastro.2006.12.031. *Gastroenterology*. 132(3),1055-65.

11. Sureban, S.M., May, R., George, R.J., Dieckgraefe, B.K., McLeod, H.L., *et al.* (2008) Knockdown of RNA binding protein musashi-1 leads to tumor regression *in vivo*. DOI: 10.1053/j.gastro.2008.02.057. *Gastroenterology*. 134(5),1448-58.

12. Mitchell, S.F., & Parker, R. (2014) Principles and properties of eukaryotic mRNPs. DOI: 10.1016/j.molcel.2014.04.033. *Mol. Cell.* 54(4),547-58.

13. Blanc, V., Henderson, J.O., Kennedy, S., & Davidson, N.O. (2001) Mutagenesis of apobec-1 complementation factor reveals distinct domains that modulate RNA binding, protein-protein interaction with apobec-1, and complementation of C to U RNA-editing activity. DOI: 10.1074/jbc.M107654200. *J. Biol. Chem.* 276(49),46386-93.

14. Blanc, V., Navaratnam, N., Henderson, J.O., Anant, S., Kennedy, S., *et al.* (2001) Identification of GRY-RBP as an apolipoprotein B RNA binding protein that interacts with both apobec-1 and with apobec-1 complementation factor (ACF) to modulate C to U editing. DOI: 10.1074/jbc.M006435200. *J. Biol. Chem.* 276(13),10272-83.

15. Blanc, V., Henderson, J.O., Newberry, E.P., Kennedy, S., Luo, J., *et al.* (2005) Targeted deletion of the murine apobec-1 complementation factor (*Acf*) gene results in embryonic lethality. DOI: 10.1128/MCB.25.16.7260-7269.2005. *Mol. Cell. Biol.* 25(16),7260-9.

16. Blanc, V., Henderson, J.O., Newberry, R.D., Xie, Y., Cho, S-J., *et al.* (2007) Deletion of the AU-rich RNA binding protein apobec-1 reduces intestinal tumor burden in Apc^{min} mice.

DOI: 10.1158/0008-5472.CAN-07-1593. *Cancer Res.* 67(18),8565-73.

17. Henderson, J.O., Blanc, V., & Davidson, N.O. (2001) Isolation, characterization and developmental regulation of the human apobec-1 complementation factor (ACF) gene. DOI: 10.1016/s0167-4781(01)00295-0. *Biochim. Biophys. Acta.* 1522(1),22-30.

18. Lau, P.P., Chang, B. H-J., & Chan, L. (2002) Two-hybrid cloning identifies an RNA-binding protein, GRY-RBP, as a component of apobec-1 editosome. DOI: 10.1006/bbrc.2001.4679. *Biochem. Biophys. Res. Commun.* 282(4),977-83.

19. Lunde, B.M., Moore, C., & Varani, G. (2007) RNA-binding proteins: modular design for efficient function. DOI: 10.1038/nrm2178. *Nat. Rev. Mol. Cell Biol.* 8(6),479-490.

20. Eck, A.G., Lopez, K.J., & Henderson, J.O. (2018) RNA-binding motif protein 45 (Rbm45)/ developmentally regulated RNA-binding protein-1 (Drbp1): Association with neurodegenerative disorders. DOI: https://doi.org/10.47611/jsr.v7i2.417. *J. Stud. Res.* 7(2),33-37.

21. Ugras, S.E., & Shorter, J. (2012) RNA-Binding proteins in amyotrophic lateral sclerosis and neurodegeneration. DOI: 10.1155/2012/432780. *Neur. Res. Int.* 2012,432780.

22. Mukhopadhyay, D., Jung, J., Murmu, N., Houchen, C.W., Dieckgraefe, B.K., *et al.* (2003) CUGBP2 plays a critical role in apoptosis of breast cancer cells in response to genotoxic injury. DOI: 10.1196/annals.1299.093. *Ann. N. Y. Acad. Sci.* 1010,504-9.

23. Qin, H., Ni, H., Liu, Y., Yuan, Y., Xi, T., *et al.* (2020) RNA-binding proteins in tumor progression. DOI: 10.1186/s13045-020-00927-w. *J. Hematol. Oncol.* 13(1),90.

24. Agrawal, S., Kuo, P.-H., Chu, L.-Y., Golzarroshan, B., Jain, M., *et al.* (2019) RNA recognition motifs of disease-linked RNA-binding proteins contribute to amyloid formation. DOI: 10.1038/ s41598-019-42367-8. *Sci. Rep.* 9(1),6171.

25. Collins, M., Li, Y., & Bowser, R. (2020) RBM45 associates with nuclear stress bodies and forms nuclear inclusions during chronic cellular stress and in neurodegenerative diseases. DOI: 10.1186/ s40478-020-00965-y. *Acta Neuropathol. Commun.* 8(1),91.

26. Price, L.E., Faul, A.B.L, Vuchkovska, A., Lopez, K.J., Fast, K.M., *et al.* (2018) Molecular genetic analysis of Rbm45/Drbp1: Genomic structure, expression, and evolution. DOI: https:// doi.org/10.47611/jsr.v7i2.426. *J. Stud. Res.* 7(2),49-61.

27. Tamada, H., Sakashita, E., Shimazaki, K., Ueno, E., Hamamoto, T., *et al.* (2002) cDNA cloning and characterization of Drb1, a new member of RRM-type neural RNA-binding protein. DOI: 10.1016/s0006-291x(02)02132-0. *Biochem. Biophys. Res. Commun.* 297(1),96-104.

28. McKee, A.E., Minet, E., Stern, C., Riahi, S., Stiles, C.D., *et al.* (2005) A genome-wide *in situ* hybridization map of RNA-binding proteins reveals anatomically restricted expression in the developing mouse brain. DOI: 10.1186/1471-213X-5-14. *BMC Dev. Biol.* 5,14.

29. Collins, M., Riascos, D., Kovalik, T., An, J., Krupa, K., *et al.* (2012) The RNA-binding motif 45 (RBM45) protein accumulates in inclusion bodies in amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration with TDP-43 inclusions (FTLD-TDP) patients. DOI: 10.1007/ s00401-012-1045-x. *Acta Neuropathol.* 124(5),717–732.

30. Strausberg, R.L., Feingold E.A., Grouse, L.H., Derge, J.G., Klausner, R.D., *et al.* (2002) Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. DOI: 10.1073/pnas.242603899. *Proc. Natl. Acad. Sci. U S A.* 99(26),16899-16903.

31. Li, Y., Collins, M. Geiser, R., Bakkar, N., Riascos, D., *et al.* (2015) RBM45 homo-oligomerization mediates association with ALS-linked proteins and stress granules. DOI: 10.1038/srep14262. *Sci. Rep*. 5,14262.

32. Mashiko, T., Sakashita, E., Kasashima, K., Tominaga, K., Kuroiwa, K., *et al.* (2016) Developmentally regulated RNA-binding motif protein 45 (RBM45), a nuclear-cytoplasmic trafficking protein, forms TAR DNA-binding protein 43 (TDP-43)-mediated cytoplasmic aggregates. DOI: 10.1074/jbc.M115.712232. *J. Biol. Chem.* 291(29),14996-15007.

33. Chen, X., Yang, Z., Wang, W., Qian, K., Liu, M., *et al.* (2021) Structural basis for RNA recognition by the N-terminal tandem RRM domains of human RBM45. DOI: 10.1093/nar/gkab075. *Nucleic Acids Res.* 49(5),2946-58.

34. Ray, D., Kazan, H., Cook, K.B., Weirauch, M.T., Najafabadi, H.S., *et al.* (2013) A compendium of RNA-binding motifs for decoding gene regulation. DOI: 10.1038/nature12311. *Nature*. 499 (7457),172-7.

35. Wang, R., Sun, H., Chen, W., Zhao, B., & Chen, L. (2023) Molecular basis of ssDNA recognition by RBM45 protein of neurodegenerative disease from multiple molecular dynamics simulations and energy predictions. DOI: 10.1016/j.jmgm.2022.108377. *J. Mol. Graph. Model.* 118,108377.

36. Bakkar, N., Kousari, A., Kovalik, T., & Bowser, R. (2015) RBM45 modulates the antioxidant response in amyotrophic lateral sclerosis through interactions with KEAP1. DOI: 10.1128/MCB.00087-15. *Mol. Cell. Bio.* 35,2385-99.

37. Li, Y., Collins, M., An, J., Geiser, R., Tegeler, T., *et al.* (2016) Immunoprecipitation and mass spectrometry defines an extensive RBM45 protein-protein interaction network. DOI: 10.1016/j.brainres.2016.02.047. *Brain Res*. 1647,79-93.

38. Del Río-Moreno, M., Alors-Pérez, E., González-Rubio, S., Ferrín, G., Reyes, *et al.* (2019) Dysregulation of the splicing machinery is associated to the development of nonalcoholic fatty liver disease. DOI: 10.1210/jc.2019-00021. *J. Clin. Endocrinol. Metab.* 104(8),3389-3402.

39. Herrero-Aguayo, V., Jiménez-Vacas, J., Sáez-Martinez, P., Gómez-Gómez, E., López-Cánovas, J.L., *et al.* (2021) Influence of obesity in the miRNome: miR-4454, a key regulator of insulin response via splicing modulation in prostate. DOI: 10.1210/clinem/dgaa580. *J. Clin. Endocrinol. Metab.* 106(2),e469-e484.

40. Wang, J., Ganaie, S.S., Cheng, G., Xu, P., Ning, K., *et al.* (2020) RNA binding motif protein RBM45 regulates expression of the 11-kilodalton protein of parvovirus B19 through binding to novel intron splicing enhancers. DOI: 10.1128/mBio.00192-20. *mBio*. 11(2),e00192-20.

41. Choi, S.H., Flamand, M.N., Liu, B., Zhu, H., Hu, M., *et al.* (2022) RBM45 is an m^6A-binding protein that affects neuronal differentiation and the splicing of a subset of mRNAs. DOI: 10.1016/j.celrep.2022.111293. *Cell Rep.* 40(9),111293.

42. Pyo, J.O., Nah, J., Kim, H.J., Chang, J.W., Song, Y.W. *et al.* (2008) Protection of cardiomyocytes from ischemic/hypoxic cell death via Drbp1 and pMe2GlyDH in cardio-specific ARC transgenic mice. DOI: 10.1074/jbc.M804209200. *J. Biol. Chem.* 283(45),30707-14.

43. Mladinic, M., Lefèvre, C., Del Bel, E., Nicholls, J., & Digby, M. (2010) Developmental changes of gene expression after spinal cord injury in neonatal opossums. DOI: 10.1016/j.brainres.2010.09.024. *Brain Res.* 1363,20-39.

44. Gong, J., Huang, M., Wang, F., Ma, X., Liu, H., *et al.* (2017) RBM45 competes with HDAC1 for binding to FUS in response to DNA damage. DOI: 10.1093/nar/gkx1102. *Nucleic Acids Res.* 45 (22),12862-12876.

45. van der Zee, J., Dillen, L., Baradaran-Heravi, Y., Gossye, H., Koçoğlu, C., *et al.* (2021) Family-based exome sequencing identifies RBM45 as a possible candidate gene for frontotemporal dementia and amyotrophic lateral sclerosis. DOI: 10.1016/j.nbd.2021.105421. *Neurobiol. Dis.* 156,105421.

46. Leys, S.P. (2015) Elements of a 'nervous system' in sponges. DOI: 10.1242/jeb.110817. *J. Exp. Biol.* 218(Pt 4),581-91.

47. Musser, J.M., Schippers, K.J., Nickel, M., Mizzon, G., Kohn, A.B., *et al.* (2021) Profiling cellular diversity in sponges informs animal cell type and nervous system evolution. DOI: 10.1126/science.abj2949. *Science.* 374(6568),717-723.

48. Fitch, W.M. (1970) Distinguishing homologous from analogous proteins. PMID: 5449325. *Syst. Zool.* 19(2),99-113.

49. Kumar, S., Stecher, G., & Tamura, K. (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. DOI: 10.1093/molbev/msw054. *Mol. Biol. Evol.* 33(7),1870-4.

50. Tamura, K. & Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. DOI: 10.1093/oxfordjournals.molbev.a040023. *Mol. Biol. Evol.* 10(3),512-26.

51. Felsenstein, J. (1985) Confidence limits on phylogenies: An approach using the bootstrap. DOI: 10.1111/j.1558-5646.1985.tb00420.x. *Evolution.* 39(4),783-791.

52. Jones, D.T., Taylor, W.R., & Thornton, J.M. (1992) The rapid generation of mutation data matrices from protein sequences. DOI: 10.1093/bioinformatics/8.3.275. *Comput. Appl. Biosci.* 8 (3),275-82.

53. Madeira, F., Pearce, M., Tivey, A.R.N., Basutkar, P., Lee, J., *et al.* (2022) Search and sequence analysis tools services from EMBL-EBI in 2022. DOI: 10.1093/nar/gkac240. *Nucleic Acids Res.* 5 (50;W1),W276-W279.

54. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. DOI: 10.1038/msb.2011.75. *Mol. Syst. Biol.* 11(7),539.

55. Sievers, F., & Higgins, D.G. (2021) The clustal omega multiple alignment package. DOI: 10.1007/978-1-0716-1036-7_1. *Methods Mol. Biol.* 2231,3-16.

56. Xu, D., Marquis, K., Pei, J., Fu, S.C., Cağatay, T., *et al.* (2015) LocNES: a computational tool for locating classical NESs in CRM1 cargo proteins. DOI: 10.1093/bioinformatics/btu826. *Bioinformatics.* 31(9),1357-65.

57. Love, G.D., Grosjean, E., Stalvies, C., Fike, D.A., Grotzinger, J.P., *et al.* (2009) Fossil steroids record the appearance of Demospongiae during the Cryogenian Period. DOI: 10.1038/nature07673. *Nature.* 457(7230),718-21.

58. Biernat, G. & Emig, C.C. (1993) Anatomical distinctions of the Mesozoic lingulide brachiopods. https://www.app.pan.pl/article/item/app38-001.html. *Acta Palaeontol. Pol.* 38(1-2),1-20.

59. Simakov, O., Kawashima, T., Marlétaz, F., Jenkins, J., Koyanagi, R., *et al.* (2015) Hemichordate genomes and deuterostome origins. DOI: 10.1038/nature16150. *Nature*. 527(7579),459–465.

60. Yuasa, H., Kajitani, R., Nakamura, Y., Takahashi, K., Okuno, M., *et al.* (2021) Elucidation of the speciation history of three sister species of crown-of-thorns starfish (*Acanthaster* spp.) based on genomic analysis. DOI: 10.1093/dnares/dsab012. *DNA Res.* 28(4),dsab012.

61. Schultz, H.A.G. (2015) Echinoidea, Vol.1. In *Handbook of Zoology: Echinodermata* (Schmidt-Rhaesa, A., ed.). Walter de Gruyter GmbH & Co KG. pp. 36 ff, section 2.4. ISBN 978-3-11-038601-1.

62. Morris, S.C. (1993) Ediacaran-like fossils in Cambrian Burgess Shale–type faunas of North America. https://www.palass.org/publications/palaeontology-journal/archive/36/3/article_pp593-635. *Palaeontology*. 36(3),593–635.

63. Chapman, J., Kirkness, E., Simakov, O., Hampson, S.E., Mitros, T. *et al.* (2020) The dynamic genome of *Hydra*. DOI: 10.1038/nature08830. *Nature.* 464(7288),592–6.

64. Sorauf, J.E., & Savarese, M. (1995) A Lower Cambrian coral from south Australia. https://www.palass.org/publications/palaeontology-journal/archive/38/4/article_pp757-770. *Palaeontology.* 38(4),757-770.

65. Fuchs, D., Iba, Y., Heyng, A., Iijima, M., Klug, C., *et al.* (2020) The Muensterelloidea: phylogeny and character evolution of Mesozoic stem octopods. DOI: 10.1002/spp2.1254 *Pap. Palaeontol.* 6(1),31-92.

66. Li, C., Kou, Q., Zhang, Z., Hu, L., Huang, W., *et al.* (2021) Reconstruction of the evolutionary biogeography reveal the origins and diversification of oysters (Bivalvia: Ostreidae). DOI: 10.1016/j.ympev.2021.107268. *Mol. Phylogenet. Evol.* 164,107268.

67. Lamsdell, J.C., & McKenzie, S.C. (2015) *Tachypleus syriacus*(Woodward)—a sexually dimorphic Cretaceous crown limulid reveals underestimated horseshoe crab divergence times. DOI: 10.1007/s13127-015-0229-3. *Org. Divers Evol.* 15,681–693.

68. Tamura, K., Subramanian, S., & Kumar, S. (2004) Temporal patterns of fruit fly (Drosophila) evolution revealed by mutation clocks. DOI: 10.1093/molbev/msg236. *Mol. Biol. Evol*. 21(1),36–44.

69. Donoghue, P.C.J., & Keating, J.N. (2014) Early vertebrate evolution. DOI: 10.1111/pala.12125. *Palaeontology.* 57(5),879-893.

70. Cavin, L., & Guinot, G. (2014) Coelacanths as "almost living fossils". DOI: 10.3389/fevo.2014.00049. *Front. Ecol. Evol*. 2,49.

71. Steeman, M.E., Hebsgaard, M.B., Fordyce, R.E., Ho, S.Y., Rabosky, D.L., *et al*. (2009) Radiation of extant cetaceans driven by restructuring of the oceans. DOI: 10.1093/sysbio/syp060. *Syst. Biol*. 58(6),573-85.

72. Pan, T., Miao, J.-S., Zhang, H.-B., Yan, P., Lee, P.-S., *et al.* (2020) Near-complete phylogeny of extant Crocodylia (Reptilia) using mitogenome-based data. DOI: 10.1093/zoolinnean/zlaa074. *Zool. J. Linn. Soc.* 91(4),1075–1089.

73. Session, A.M, Uno, Y., Kwon, T., Chapman, J.A., Toyoda, A., *et al.* (2016) Genome evolution in the allotetraploid frog *Xenopus laevis*. DOI: 10.1038/nature19840. *Nature.* 538(7625),336–343.

74. Friedman, M. (2015) The early evolution of ray-finned fishes. DOI: 10.1111/pala.12150. *Palaeontology*. 58(2),213–228.

75. Woodburne, M.O. (2010) The great American biotic interchange: dispersals, tectonics, climate, sea level and holding pens. DOI: 10.1007/s10914-010-9144-8. *J. Mammal. Evol.* 17(4),245–264.

76. Verneau, O., Catzeflis, F., & Furano, A.V. (1998) Determining and dating recent rodent speciation events by using L1 (LINE-1) retrotransposons. DOI: 10.1073/pnas.95.19.11284. *Proc. Natl. Acad. Sci. U S A*. 95(19),11284-9.

77. Jaeger, J.-J., Tong, H., & Denys, C. (1986) The age of *Mus-Rattus* divergence: paleontological data compared with the molecular clock. https://gallica.bnf.fr/ark:/12148/bpt6k5675350v/f923.item. *C.R. Acad. Sc. Paris.* 302(ser. II)(14),917-922.

78. Lawal, R.A., Martin, S.H., Vanmechelen, K., Vereijken, A., Silva, P., *et al.* (2020) The wild species genome ancestry of domestic chickens. DOI: 10.1186/s12915-020-0738-1. *BMC Biol.* 18 (1),13.

79. Reis, M.D., Gunnell, G.F., Barba-Montoya, J., Wilkins, A., Yang, Z., *et al.* (2018) Using phylogenomic data to explore the effects of relaxed clocks and calibration strategies on divergence time estimation: Primates as a test case. DOI: 10.1093/sysbio/syy001. *Syst. Biol*. 67(4),594-615.

80. Hublin, J-J., Ben-Neer, A., Bailey, S.E., Freidline, S.E, Neubauer, S. *et al.* (2017) New fossils from Jebel Irhoud, Morocco and the pan-African origin of *Homo sapiens*. DOI: 10.1038/nature22336. *Nature*. 546(7657),289-92.

81. Penny, D., Foulds, L. & Hendy, M. (1982) Testing the theory of evolution by comparing phylogenetic trees constructed from five different protein sequences. DOI: 10.1038/297197a0. *Nature.* 297(5863),197–200.

82. Som, A. (2015) Causes, consequences and solutions of phylogenetic incongruence. DOI: 10.1093/bib/bbu015. *Brief Bioinform*. 16(3),536-48.

83. Smith, B.D., Wang, D., Braulik, G.T., Reeves, R., Zhou, K., *et al.* (2017) *Lipotes vexillifer*. *The IUCN Red List of Threatened Species.* e.T12119A50362206. [cited 2024 February 5]. Available from https://dx.doi.org/10.2305/IUCN.UK.2017-3.RLTS.T12119A50362206.en.

84. Gobush, K.S., Edwards, C.T.T, Balfour, D., Wittemyer, G., Maisels, F., *et al.* (2022) *Loxodonta africana* (amended version of 2021 assessment). *The IUCN Red List of Threatened Species.* e.T181008073A223031019. [cited 2024 February 5]. Available from https://dx.doi.org/10.2305/IUCN.UK.2022-2.RLTS.T181008073A223031019.en.

85. Gauthier, J.A. (1986) Saurischian monophyly and the origin of birds. https://biostor.org/reference/110202. BHL: 15651737. *Mem. Calif. Acad. Sci.* 8,1-55.

86. Brusca, R.C., Moore, W., & Shuster, S.M. (2016) *Invertebrates*, 3rd ed., Sinauer Associates Inc., Sunderland, MA, 1104 pp. ISBN: 978-1605353753.

87. Hickman, C.P., Roberts, L.S., Keen, S.L., Eisenhour, D.J., Larson, A., *et al.* (2011) *Integrated Principles of Zoology* 15th ed., McGraw-Hill Inc., New York, NY, 842 pp. ISBN: 978-0073040509.

88. Feuda, R., Dohrmann, M., Pett, W., Philippe, H., Rota-Stabelli, O., *et al.* (2017) Improved modeling of compositional heterogeneity supports sponges as sister to all other animals. DOI: 10.1016/j.cub.2017.11.008. *Curr. Biol.* 27(24),3864-3870.

89. Nixon, K.C., & Carpenter, J.M. (1993) On outgroups. DOI: 10.1111/j.1096-0031.1993.tb00234.x. *Cladistics* 9(4),413-426.

90. Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M.E.A., *et al.* (2010) The *Amphimedon queenslandica* genome and the evolution of animal complexity. DOI: 10.1038/nature09201. *Nature*. 466(7307),720-726.

91. Emlen, D.J. & Zimmer, C. (2020) *Evolution: Making sense of life*, 3rd ed., W.H. Freeman and Company, New York, NY, 782 pp. ISBN:978-1319079864.

92. Kayal, E., Bentlage, B., Sabrina Pankey, M., Ohedra, A.H., Medina, M., *et al.* (2018) Phylogenomics provides a robust topology of the major cnidarian lineages and insights on the origins of key organismal traits. DOI: 10.1186/s12862-018-1142-0. *BMC Evol. Biol.* 18,68.

93. Hillis, D.M., & Bull, J.J. (1993) An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. DOI: 10.1093/sysbio/42.2.182. *Syst. Biol.* 42(2),182-192.

94. Hall, B.G. (2013) Building phylogenetic trees from molecular data with MEGA. DOI: 10.1093/molbev/mst012. *Mol. Biol. Evol.* 30(5),1229–35.

95. Betancur-R, R., Wiley, E.O., Arratia, G., Acero, A., Bailly, N., *et al.* (2017) Phylogenetic classification of bony fishes. DOI: 10.1186/s12862-017-0958-3. *BMC Evol. Biol*. 17,162.

96. Brazeau, M.D., Giles, S., Dearden, R.P. Jerve, A., Ariunchimeg, Y., *et al.* (2020) Endochondral bone in an Early Devonian 'placoderm' from Mongolia. DOI: 10.1038/s41559-020-01290-2. *Nat. Ecol. Evol.* 4(11),1477–1484.

97. Amemiya, C.T., Alföldi, J., Lee, P.E., Fan, S., Philippe, H. *et al.* (2013) The African coelacanth genome provides insights into tetrapod evolution. DOI: 10.1038/nature12027. *Nature* 496 (7445),311-6.

98. Jeffroy, O., Brinkmann, H., Delsuc, F., & Philippe, H. (2006) Phylogenomics: the beginning of incongruence? DOI: 10.1016/j.tig.2006.02.003. *Trends Genet*. 22(4),225–31.

99. Degnan, J.H., & Rosenberg, N.A. (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. DOI: 10.1016/j.tree.2009.01.009. *Trends Ecol. Evol.* 24(6),332-40.

100. Pamilo, P., & Nei, M. (1988) Relationships between gene trees and species trees. 10.1093/oxfordjournals.molbev.a040517. *Mol. Biol. Evol.* 5(5),568-83.

101. Lopez, P., Casane, D., & Philippe, H. (2002) Heterotachy an important process of protein evolution. DOI: 10.1093/oxfordjournals.molbev.a003973. *Mol. Biol. Evol.* 19(1),1-7.

102. Kolaczkowski, B., & Thornton, J.W. (2004) Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. DOI: 10.1038/nature02917. *Nature*. 431 (7011),980-984.

103. Simmons, M.P. (2000) A fundamental problem with amino-acid-sequence characters for phylogenetic analyses. DOI: 10.1111/j.1096-0031.2000.tb00283.x. *Cladistics*. 16(3),274-282.

104. Wenzell, J.W. (2002) Phylogenetic analysis: The basic method. In *Techniques in Molecular Systematics and Evolution* (DeSalle, R., Giribet, G., Wheeler, W., eds.). Springer Basel AG, pp. 4-30. DOI: 10.1007/978-3-0348-8125-8_1

105. Castresana, J. (2007) Topological variation in single-gene phylogenetic trees. DOI: 10.1186/gb-2007-8-6-216. *Genome Biol.* 8(6),216.

106. Da Silva, F.B., Muschner, V.C., & Bonatto, L. (2007) Phylogenetic position of *Placozoa* based on large subunit (LSU) and small subunit (SSU) rRNA genes. https://pesquisa.bvsalud.org/portal/resource/pt/lil-445694?lang=en. *Genet. Mol. Biol.* 30(1),127-132.

107. Schultz, D.T., Haddock, S.H.D., Bredeson, J.V., Green, R.E., Simakov, O., *et al.* (2023) Ancient gene linkages support ctenophores as sister to other animals. DOI: 10.1038/s41586-023-05936-6. *Nature*. 618(7963),110-117.

108. Srivastava, M., Begovic, E., Chapman, J., Putnam, N.H., Hellsten, U., *et al.* (2008) The Trichoplax genome and the nature of placozoans. DOI: 10.1038/nature07191. *Nature*. 454(7207),955-60.

109. Burkhardt, P., Colgren, J., Medhus, A., Digel, L., Nauman, B., *et al.* (2023) Syncytial nerve net in a ctenophore adds insights on the evolution of nervous systems. DOI: 10.1126/science.ade5645. *Science*. 380(6642),293-297.

110. Nielsen, C. (2019) Early animal evolution: A morphologist's view. DOI: 10.1098/rsos.190638. *R. Soc. Open Sci.* 6(7),190638.

111. Chelsky, D., Ralph, R., & Jonak, G. (1989) Sequence requirements for synthetic peptide-mediated translocation to the nucleus. DOI: 10.1128/mcb.9.6.2487-2492.1989. *Mol. Cell Biol*. 9(6),2487-92.

112. Kosugi, S., Hasebe, M., Tomita, M., & Yanagawa, H. (2008) Nuclear export signal consensus sequences defined using a localization-based yeast selection system. DOI: 10.1111/j.1600-0854.2008.00825.x. *Traffic*. 9(12),2053-62.

113. Ederle, H., Funk, C., Abou-Ajram, C., Hutten, S., Funk, E.B.E., *et al.* (2018) Nuclear egress of TDP-43 and FUS occurs independently of Exportin-1/CRM1. DOI: 10.1038/s41598-018-25007-5. *Sci. Rep.* 8(1),7084.

114. Koralewski, T.E., & Krutovsky, K.V. (2011) Evolution of exon-intron structure and alternative splicing. DOI: 10.1371/journal.pone.0018055. *PLoS One.* 6(3),e18055.

115. Adami, C., Ofria, C., & Collier, T.C. (2000) Evolution of biological complexity. DOI: 10.1073/pnas.97.9.4463. *Proc. Natl. Acad. Sci. U S A*. 97(9),4463-8.

116. Lynch, M., & Conery, J.S. (2003) The origins of genome complexity. DOI: 10.1126/science.1089370. *Science*. 302(5649),1401-4.

117. Valentine, J.W. (2000) Two genomic paths to the evolution of complexity in body plans. https://www.jstor.org/stable/2666123. *Paleobiology*. 26(3),513-519.

118. Ruggiero, M.A., Gordon, D.P., Orrell, T.M., Bailly, N., Bourgoin, T. *et al.* (2015) A higher level classification of all living organisms. DOI: 10.1371/journal.pone.0119248. *PLoS One*. 10(4),e0119248.

119. Kolbasova, G., Schmidt-Rhaesa, A., Syomin, V., Bredikhin, D., Morozov, T., *et al.* (2023) Cryptic species complex or an incomplete speciation? Phylogeographic analysis reveals an intricate Pleistocene history of *Priapulus caudatus* Lamarck, 1816. DOI: 10.1016/j.jcz.2022.11.013. *Zoologischer Anzeiger*. 302,113-130.

120. Huang, D.Y., Vannier, J., & Chen, J.Y. (2004) Recent Priapulidae and their Early Cambrian ancestors: comparisons and evolutionary significance. DOI: 10.1016/j.geobios.2003.04.004. *GeoBios*. 37(2),217-228.

121. Kesidis, G., Slater, B. J., Sören, J., & Budd, G.E. (2019) Caught in the act: priapulid burrowers in

early Cambrian substrates. DOI: 10.1098/rspb.2018.2505. *Proc. R. Soc. B.* 286(1894),20182505.

122. Budd, G.E. (2013) At the origin of animals: the revolutionary Cambrian fossil record. DOI: 10.2174/13892029113149990011. *Curr. Genomics*. 14(6),344-54.

123. Peterson, K. J., Lyons, J. B., Nowak, K. S., Takacs, C. M., Wargo, M. J., *et al.* (2004) Estimating metazoan divergence times with a molecular clock. DOI: 10.1073/pnas.0401670101. *Proc. Natl. Acad. Sci. U S A*. 101(17),6536–6541.

124. Grau-Bové, X., Torruella, G., Donachie, S., Suga, H., Leonard, G., *et al.* (2017) Dynamics of genomic innovation in the unicellular ancestry of animals. DOI: 10.7554/eLife.26036. *eLife*. 6,e26036.

125. Haeckel, E. (1874) The gastraea-theory, the phylogenetic classification of the animal kingdom and the homology of the germ-lemellæ. DOI: 10.1242/jcs.S2-14.54.142. *Q. J. Microsc. Sci.*14(54),142-163.

126. Lang, B.F., O'Kelly, C.J., Nerad, T.A., Gray, M.W., & Burger, G. (2002) The closest unicellular relatives of animals. DOI: 10.1016/s0960-9822(02)01187-9. *Curr. Biol.* 12(20),1773-8.