

Research



**Cite this article:** Ramos-Vicente D, Bayés À. 2020 AMPA receptor auxiliary subunits emerged during early vertebrate evolution by neo/subfunctionalization of unrelated proteins. *Open Biol.* **10**: 200234. <http://dx.doi.org/10.1098/rsob.200234>

Received: 29 July 2020  
Accepted: 3 October 2020

**Subject Area:**  
neuroscience/molecular biology/bioinformatics

**Keywords:**  
AMPA receptor auxiliary subunit, TARP, cornichon, shisa, Dispanin C, phylogeny

**Author for correspondence:**  
Àlex Bayés  
e-mail: [abayesp@santpau.cat](mailto:abayesp@santpau.cat)

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.5179219>.

# AMPA receptor auxiliary subunits emerged during early vertebrate evolution by neo/subfunctionalization of unrelated proteins

David Ramos-Vicente<sup>1,2</sup> and Àlex Bayés<sup>1,2</sup>

<sup>1</sup>Molecular Physiology of the Synapse Laboratory, Biomedical Research Institute Sant Pau, Barcelona, Spain  
<sup>2</sup>Universitat Autònoma de Barcelona, Barcelona, Spain

DR, 0000-0002-2730-0850; AB, 0000-0002-5265-6306

In mammalian synapses, the function of ionotropic glutamate receptors is critically modulated by auxiliary subunits. Most of these specifically regulate the synaptic localization and electrophysiological properties of AMPA-type glutamate receptors (AMPA receptors). Here, we comprehensively investigated the animal evolution of the protein families that contain AMPAR auxiliary subunits (ARASs). We observed that, on average, vertebrates have four times more ARASs than other animal species. We also demonstrated that ARASs belong to four unrelated protein families: CACNG-GSG1, cornichon, shisa and Dispanin C. Our study demonstrates that, despite the ancient origin of these four protein families, the majority of ARASs emerged during vertebrate evolution by independent but convergent processes of neo/subfunctionalization that resulted in the multiple ARASs found in present vertebrate genomes. Importantly, although AMPARs appeared and diversified in the ancestor of bilateral animals, the ARAS expansion did not occur until much later, in early vertebrate evolution. We propose that the surge in ARASs and consequent increase in AMPAR functionalities, contributed to the increased complexity of vertebrate brains and cognitive functions.

## 1. Introduction

Ionotropic glutamate receptors are key to the physiology of the nervous system, as they mediate fast excitatory neurotransmission [1]. Previously, we reported that the evolution of the proteins that form these tetrameric receptors has been much more sophisticated than previously acknowledged [2]. We found that beyond the well-known AMPA, Kainate, NMDA and Delta classes, there are four other classes exclusive to invertebrate species. Because of this, species with simple nervous systems, such as the sea anemone *N. vectensis*, have a similar number of ionotropic glutamate receptor subunits to animals with complex brains, such as mammals [2]. We argued that the high diversity of neuronal types [3] or the wide array of glutamate receptor functionalities [4] found in vertebrates is unlikely to be the result of an increased repertoire of genes coding for them. Notably, the subcellular traffic and function of ionotropic glutamate receptors is controlled by their auxiliary subunits [5]. These proteins add a new layer of complexity to glutamatergic transmission and might have contributed to an expanded functionality of these receptors in animals with complex brains. In mammals, sixteen proteins have been identified as ionotropic glutamate receptor auxiliary subunits. Of these, fourteen modulate AMPA-type glutamate receptors (AMPA receptors) [6], and they are referred to as AMPA receptor auxiliary subunits (ARASs). Only one auxiliary subunit has been reported for NMDA receptors, Neto1 [7], which also regulates Kainate receptors, as does Neto2 [8,9]. Currently known mammalian ARASs belong to five

protein families: CACNG and GSG1 (both within the superfamily of claudins), cornichon, shisa and Dispanin C.

Genes belonging to the same family are defined as paralogues; these originate from gene or genome duplication events [10]. In vertebrates, many paralogues resulted from the two rounds of whole-genome duplication (2R) occurred at the base of this lineage, approximately 400 Ma [11]. After duplication, the two new genes can experience processes of neofunctionalization or subfunctionalization. In the first scenario, one of the new genes retains all functions performed by the ancestral gene, while the second acquires new ones [12,13]. Alternatively, genes subfunctionalize when the multiple functions carried out by the ancestral gene are realized separately by descendent paralogues [13,14]. Establishing if paralogues have undergone a process of neo or subfunctionalization requires a precise understanding of all functions carried out by the ancestral gene and the descending paralogues. When this knowledge is not available, it is not possible to determine which one of these two processes took place [15]; in these occasions, the term neo/subfunctionalization can be used [13].

Known invertebrate ARAS belong to the CACNG or cornichon families. These have been described in *C. elegans* [16,17] and the fruit fly [18,19]. The first ARAS to be reported was CACNG2, also called stargazin, as it is mutated in *stargazer* mice [20]. Although this protein is phylogenetically related to CACNG1, an auxiliary subunit of voltage-dependent calcium channels (VDCC), it specifically interacts with AMPAR [21,22]. Subsequently, six other CACNGs have been discovered in mammals (CACNG3 to CACNG8 [23,24]). Apart from CACNG6, which also acts as a VDCC auxiliary subunit [25,26], all other CACNGs function as AMPA receptor auxiliary subunits [27–29]; these are usually referred to as TARPs (transmembrane AMPA receptor regulatory proteins) [27]. Phylogenetic analysis of mammalian TARPs identifies that they are more related to each other than to auxiliary subunits of VDCCs [25,27], being divided in two types [27,30,31] that differentially modulate AMPAR [28,29]. The most recently identified ARAS is GSG1L [32,33]. In vertebrate genomes, GSG1L has two paralogues: GSG1, which interacts with the polymerase TPAP, and GSG1L2 [32,34], of unknown biological function. These proteins also belong to the superfamily of claudins, like CACNGs. Despite its structural similarity with TARPs, GSG1L modulates AMPAR in a different way, downregulating its traffic to the plasma membrane and accelerating deactivation and desensitization [35,36]. The remaining ARASs have been identified in three other protein families: cornichon, shisa and Dispanin C. In all cases, these families contain ARAS and proteins with different biological functions. Among cornichons, CNIH2 and CNIH3 modulate AMPAR function, but CNIH1 and CNIH4 regulate the traffic of TGF $\alpha$  and GPCRs, respectively [37–41]. Four members of the shisa family, Shisa6 to Shisa9 (originally referred to as CKAMPs [42–46]), interact with AMPAR, yet Shisa2 and Shisa3 are involved in the traffic of FGF [47] and WNT [48] receptors, respectively, and Shisa5 participates in the p53/TP53-dependent apoptosis pathway [49]. Finally, SynDIG1, from the Dispanin C family, has also been identified as an ARAS [50,51], this protein has two paralogues in vertebrates, SynDIG1L and TMEM91, with poorly understood functions.

The evolutionary origin of ARASs is well established for cornichons, which have been identified in the ancestor of eukaryotes and are present in a large range of species, including plants and yeasts [52,53]. The claudin superfamily, to which

CACNG and GSG1 proteins belong, present homologues in the basal metazoan phylum of porifera [54,55], placing its origin prior to the divergence of these organisms. While CACNG homologues have been reported in different bilaterian species, including vertebrates and the fruit fly or *C. elegans* [17,19], it is still unknown when GSG1s appeared during evolution. In addition, some shisa homologues have been described in porifera and Dispanin C homologues have been identified in brown algae [56,57]. Here, we present a comprehensive study of the animal evolution of all protein families that include AMPAR auxiliary subunits. Our work shows that cornichons, present in the ancestor of all metazoans, would be the most ancient of all ARAS. TARPs, which appeared together with AMPARs in the ancestor of bilaterians, also function as ARASs in invertebrate organisms. Nevertheless, most ARASs, including shisas, GSG1L and SynDIG1, would have been recruited to modulate AMPAR function early in vertebrate evolution, suggesting that during this period there was an evolutionary pressure that favoured an expansion of the functionality of AMPARs.

## 2. Materials and methods

### 2.1. Identification of genes coding for AMPA receptor auxiliary subunits in metazoan genomes

Phylogenetic analysis was performed as described previously [2]. We searched sequences in 31 species belonging to different metazoan phyla: Porifera, Ctenophora, Placozoa, Cnidaria, Lophotrochozoa, Ecdysozoa, Hemichordata, Chordata and Vertebrata. The same species were used to construct all phylogenies. We used slowly evolving species whenever possible. All sequences were retrieved from public databases. *A. digitifera* and *P. flava* sequences were obtained from the Marine Genomics Unit [58,59]. A complete list of species included in the analysis and the corresponding database where sequence search was done can be found in electronic supplementary material, table S1.

Sequences from protein families of interest were identified by homology-based searches with reciprocal identifications. Mouse proteins were used as search queries, when different isoforms were present we used the longest amino acid sequence. We searched for homologues using the BLASTP and TBLASTN tools [60] with default parameters. Subject sequences with an E-value below 0.05 were selected as candidate homologues. These sequences were re-blasted against mammal proteins in the NCBI database of ‘non-redundant protein sequences’ using the BLASTP and BLASTX tools, respectively. Identified homologues which have a protein length that is less than 50% of the query protein were discarded.

### 2.2. Phylogenetic analyses

The CACNG-GSG1 tree was constructed using a total of 114 sequences, from which fifteen are used as the outgroup (electronic supplementary material, file S1). The cornichon tree includes 70 sequences, of which five are cornichons from *Arabidopsis thaliana* used as the outgroup (electronic supplementary material, file S2). The shisa tree includes 98 sequences, eight of which are used as the outgroup (electronic supplementary material, file S3). Finally, the Dispanin C tree includes 41

sequences, four of which were used as the outgroup (electronic supplementary material, file S4).

Protein sequences were aligned with the MUSCLE algorithm [61], included in the software package MEGAX [62] with default parameters. ProtTest v3.4.2 was used to establish the best evolutionary model [63]. Trees were constructed using MrBayes v3.2.7 [64] for Bayesian inference and IQ-TREE [65] for maximum-likelihood analysis. For Bayesian inference phylogenies were run for 10 000 000 generations. Markov chain Monte Carlo (MCMC) was used to approximate the posterior probability of the Bayesian trees. Bayesian analyses included two independent MCMC runs, each using four parallel chains composed of three heated and one cold chain. Twenty-five per cent of initial trees were discarded as burn-in. Convergence was assessed when the potential scale reduction factor (PSRF) value was between 1.002 and 1.000. In maximum-likelihood analysis the starting tree was estimated using a neighbour-joining method and branch support was obtained after 1000 iterations of ultrafast bootstrapping [66]. Gene/protein names were given based on their position in the tree. Phylogenetic trees were rendered using FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>). Phylogenetic calculations were performed at the CIPRES science gateway [67].

### 2.3. Protein nomenclature

Proteins from non-vertebrate species were systematically named following this nomenclature: (i) the name of the subfamily which they belong, or family if the sequence is not assigned to a subfamily; (ii) a Greek letter to identify non-vertebrate paralogues, if any; and (iii) a three-letter species code.

### 2.4. Prediction of PDZ motifs classes

The software Eukaryotic Linear Motif (ELM) [68] was used to identify if the C-terminus of proteins presented a PDZ binding motif and, if so, to what class it belonged.

## 3. Results

### 3.1. TARPs are ancient AMPAR auxiliary subunits widespread in bilaterians

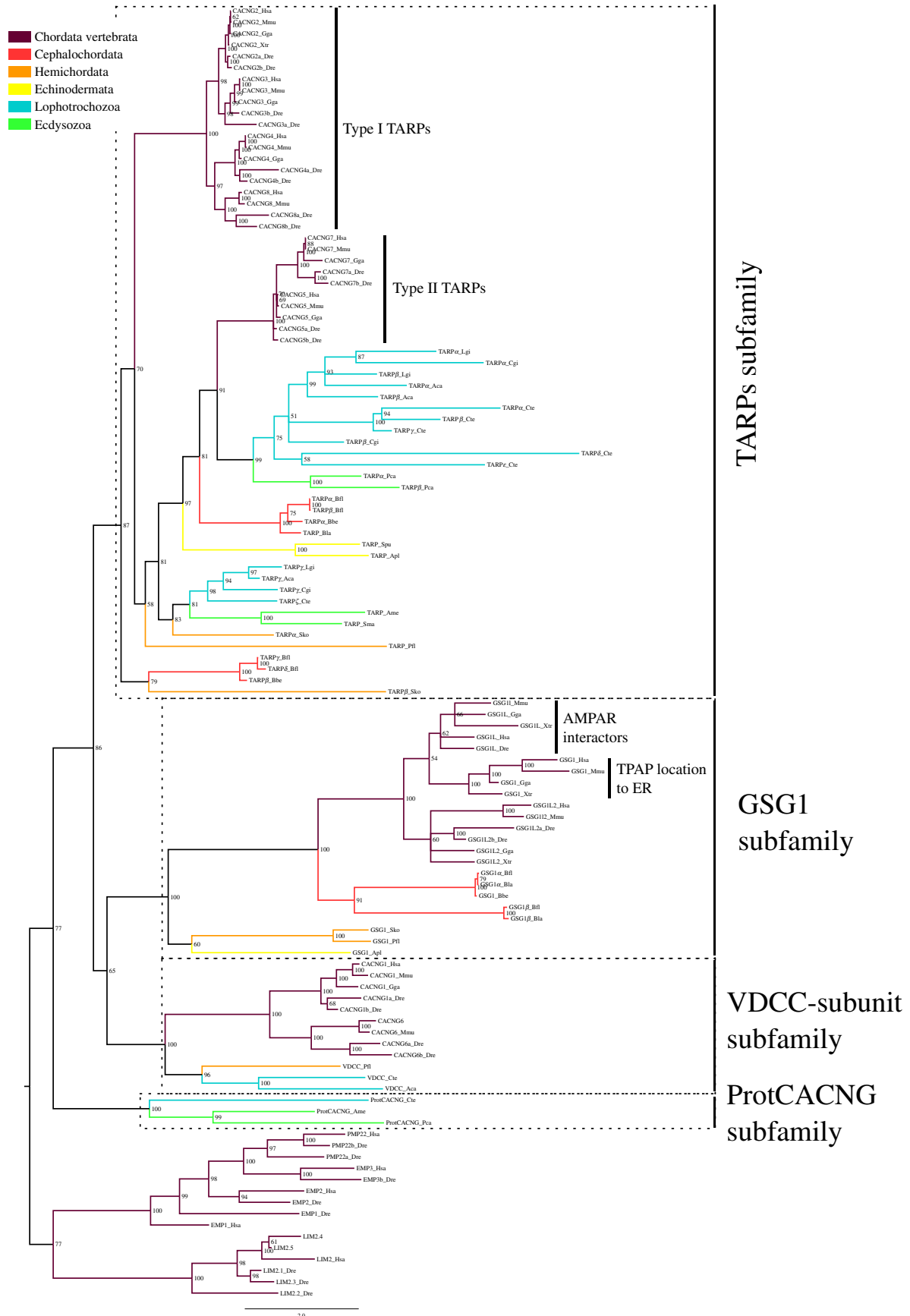
Among protein families including AMPA receptor auxiliary subunits, CACNGs and GSG1s belong to the superfamily of claudins [32,69,70]. We thus incorporated GSG1s into the phylogeny of CACNGs to investigate if they are part of the same family within the claudin superfamily. We found that GSG1 proteins confidently fell within the ingroup (figure 1; electronic supplementary material, figure S1). Thus CACNGs and GSG1s are more evolutionarily related to each other than they are to the rest of the claudin superfamily, belonging to the same family, which we have named CACNG-GSG1. We only found CACNG-GSG1 sequences in bilaterian species (table 1), indicating that this family appeared in a common ancestor of bilaterians. Phylogenetic analysis identified four subfamilies among CACNG-GSG1s: TARPs, GSG1s, VDCCs and a protostome-specific subfamily. As these are widely represented among bilaterian phyla, we propose that the ancestral CACNG-GSG1 experienced three duplication events before the split of bilaterians (electronic supplementary material, figure S2A). Nevertheless, only the TARP subfamily is conserved in

all bilaterians investigated. The other families were lost in certain lineages. For instance: (i) the GSG1 subfamily is only present in deuterostomes; (ii) molluscs would have lost the protostome-specific subfamily; and (iii) the VDCC-subunit subfamily was independently lost by cephalochordates, echinodermites and ecdysozoans.

Among the phyla studied, vertebrates present the highest number of CACNG-GSG1s (11 proteins; table 1), of which 7 are ARAS. Although *Capitella teleta*, an annelid, has eight members in this family, of which six would be TARPs according to the phylogenies, most invertebrates investigated generally have fewer CACNG-GSG1s. All mammalian proteins within the branch of TARPs are known AMPAR regulatory proteins [27,28,71]. Due to their position in the trees we propose that invertebrate orthologues to vertebrate TARPs and their bilaterian ancestor would interact and regulate AMPARs. However, with the exception of cephalochordate sequences, invertebrate TARPs are more divergent than their vertebrate counterparts, as indicated by their longer branches. It is thus plausible that the ability to interact with AMPARs has been altered in this species. Functionally, vertebrate TARPs are classified into Type I and Type II [27,28,71]. This classification is mirrored by the phylogenies, as vertebrate sequences form two monophyletic groups, one for each type. While invertebrate proteins cannot be unambiguously related to Type I or Type II TARPs, as the statistics metrics generated by the Bayesian (figure 1) and maximum-likelihood (electronic supplementary material, figure S1) phylogenies are not high enough, the trees suggest that the majority of them belong to Type II and that Type I would have been lost in non-vertebrates. Among GSG1s, vertebrate GSG1s are the only ones known to act as ARASs [32,33]. Yet, our tree indicates that all GSG1 paralogues arise from recent duplications at the base of the vertebrate phylum, suggesting that the emergence of ARAS in this family occurred by neo/subfunctionalization early in vertebrate evolution.

### 3.2. Sequence analysis reveals common motifs in CACNG-GSG1s

We next constructed multiple sequence alignments to investigate primary sequence features among CACNG-GSG1s (figure 2a; electronic supplementary material, figure S3). We found a highly conserved motif of 7 residues in all CACNG-GSG1s (consensus sequence: Y(174)SYGWSF, residue numbering corresponds to human CACNG2). The last five residues of this motif present the highest conservation, S179 being the most conserved position. We also investigated which proteins from this family would present PDZ binding motifs, as mammalian TARPs [5] present Class 1 PDZ motifs. These short C-terminal sequences contribute to the anchoring of TARPs at the postsynaptic membrane [72,73]. All TARP subfamily sequences except two, TARP\_Pfl and TARP $\alpha$ \_Pca, are predicted to have a PDZ binding motif; most of these being Class 1 motifs. TARP\_Spu, TARP\_Apl, TARP $\beta$ \_Lgi and TARP $\beta$ \_Aca are predicted to have changed this into Class 2 motifs. Similarly, most GSG1 subfamily proteins also have a PDZ binding motif, except GSG1s from hemichordates (figure 2b). These motifs are present in GSG1 proteins that act as ARAS but also in those with other functions. Interestingly, this subfamily presents three types of PDZ motifs. Vertebrate GSG1s are predicted to have Class 3 motifs,



**Figure 1.** Bayesian inference phylogenetic tree of the CACNG-GSG1 family. The subfamilies in which the CACNG-GSG1 protein family is divided are highlighted by dashed line boxes. The name of each subfamily is presented at the right of the corresponding box. Posterior probabilities are shown at tree nodes and protein names at the end of branches. Tree branches are coloured based on phylum, as indicated in the legend. The closest relatives to vertebrate CACNG in the claudin superfamily were used as outgroup. The Type I and Type II TARPs are indicated. If known, the function of vertebrate GSG1 subfamily sequences was also included. Scale bar denotes number of amino acid substitutions per site. The amino acid substitution model used was Vt + G + F, the analysis ran for 10 000 000 generations, the final standard deviation was 0.012330 and the final potential scale reduction factor (PSRF) was 1.000.

**Table 1.** Number of proteins found in the four families known to include AMPA receptor auxiliary subunits (ARASs). The number of ARASs is given in brackets. Proteins from *Homo sapiens* are experimentally confirmed as ARAS. Invertebrate proteins are defined as candidate ARASs based exclusively in the phylogenies.

		phylum	species	CACNG-GSG1	CNIH	Shisa	Dispanin C	
Metazoa	Bilateria	Vertebrata	<i>H. sapiens</i>	11 (7)	4 (2)	10 (4)	3 (1)	
			Chordata	<i>C. intestinalis</i>	0	2 (1)	0	0
				<i>B. lanceolatum</i>	3 (1)	3 (1)	6 (0)	1 (1)
		<i>B. belcheri</i>		3 (2)	2 (1)	6 (0)	7 (1)	
		<i>B. floridae</i>		6 (4)	3 (2)	8 (0)	3 (0)	
		Hemichordata		<i>S. kowalevskii</i>	3 (2)	2 (1)	16 (0)	1 (1)
			<i>P. flava</i>	3 (1)	1 (0)	4 (0)	0	
		Echinodermata	<i>A. planci</i>	2 (1)	1 (1)	2 (0)	0	
			<i>S. purpuratus</i>	1 (1)	1 (1)	1 (0)	2 (0)	
		Mollusca	<i>L. gigantea</i>	3 (3)	2 (1)	0	2 (0)	
			<i>C. gigas</i>	3 (3)	3 (1)	0	1 (1)	
			<i>A. californica</i>	4 (3)	3 (2)	1 (0)	0	
		Annellida	<i>C. teleta</i>	8 (6)	2 (1)	2 (0)	1 (0)	
		Arthropoda	<i>A. mellifera</i>	2 (1)	1 (1)	0	0	
			<i>S. maritima</i>	1 (1)	1 (0)	1 (0)	0	
		Priapulida	<i>P. caudatus</i>	3 (2)	2 (1)	0	0	
		non-Bilateria	Cnidaria	<i>N. vectensis</i>	0	2 (1)	0	0
				<i>A. digitifera</i>	0	2 (1)	0	0
				<i>O. faveolata</i>	0	2 (1)	0	0
	<i>H. magnipapillata</i>			0	1 (1)	0	1 (0)	
	Placozoa		<i>T. adhaerens</i>	0	2 (1)	0	0	
	Porifera		<i>O. carmela</i>	0	2 (1)	0	2 (1)	
			<i>S. ciliatum</i>	0	1 (0)	0	0	
<i>L. complicata</i>			0	2 (1)	0	1 (0)		
<i>A. queenslandica</i>			0	2 (1)	0	0		
Ctenophora	<i>M. leidy</i>		0	1 (1)	0	1 (0)		

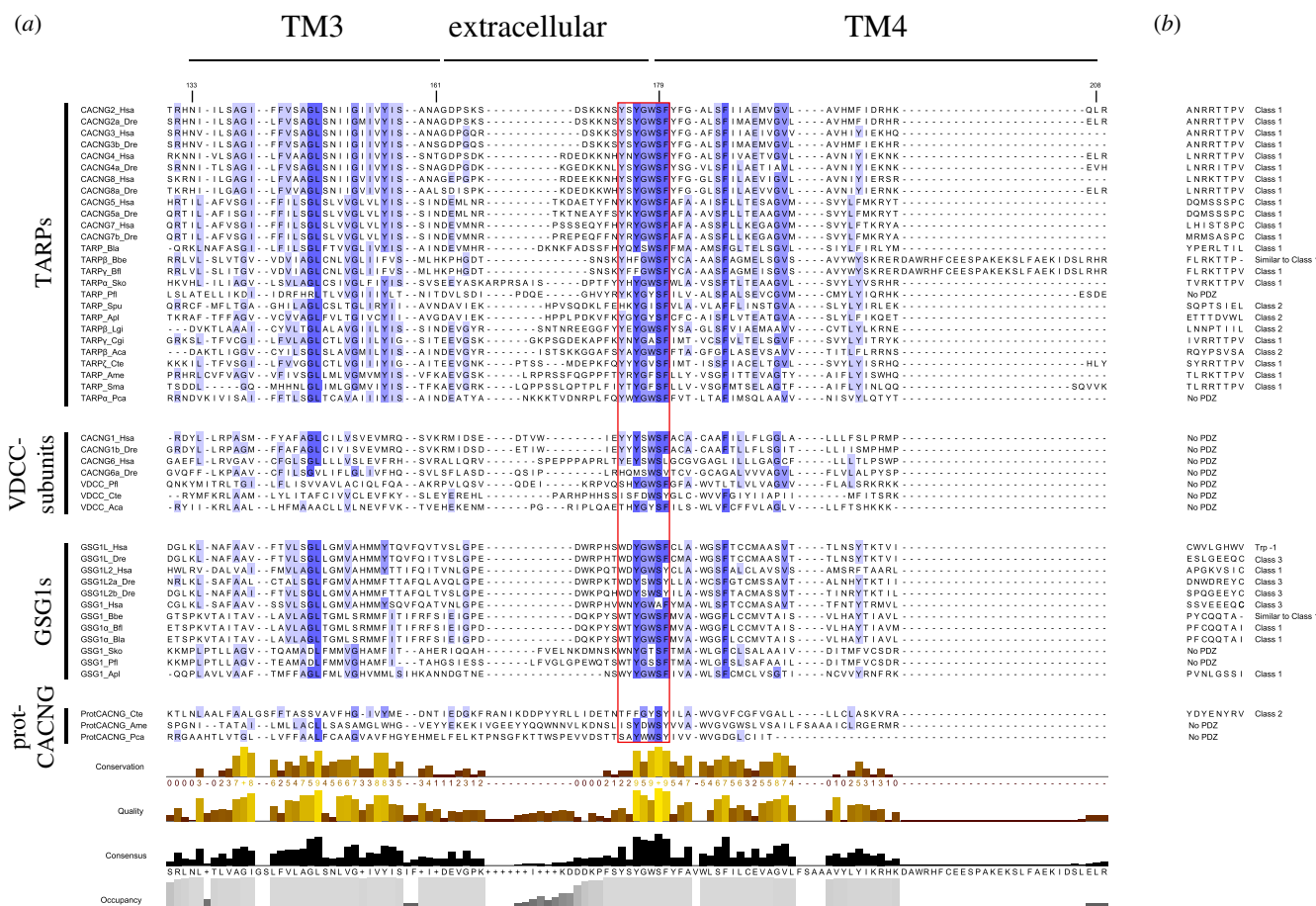
although their cephalochordate orthologues would have Class 1 motifs. Mammalian GSG1L2 would also present Class 1 motifs although *Danio rerio* orthologues present a Class 3 motif. Finally, mammalian GSG1Ls, which are the only members of the family known to interact with AMPARs, present a PDZ motif class termed Trp-1 [74]. This motif results from an insertion of 7 amino acids exclusive to the mammalian lineage. Other vertebrate species present a Class 3 PDZ motif (electronic supplementary material, figure S4).

### 3.3. A process of subfunctionalization would have led to vertebrate cornichon ARAS

We next investigated the phylogeny of cornichon proteins. Noticeably, we found cornichon sequences in all species investigated, covering all metazoan phyla (table 1). Our phylogenetic analysis indicates that this family is divided into two subfamilies: CNIH1/2/3 and CNIH4 (figure 3; electronic supplementary material, figure S5). Furthermore, it also shows that the ancestor of all metazoans already presented the two genes that later gave rise to these two subfamilies (electronic supplementary material, figure S2B). Interestingly, both cornichon

subfamilies are highly conserved, having experienced few duplication or deletion events throughout metazoan evolution. Nevertheless, while the CNIH1/2/3 subfamily is found in all metazoan phyla studied, ctenophores and echinoderms apparently lost the CNIH4 subfamily. From all species studied, vertebrates have the highest number of cornichon proteins, presenting four. Among invertebrates, *B. lanceolatum*, *B. floridae*, *C. gigas* and *A. californica* are the species with more cornichons, with three coding genes in their genomes.

The vertebrate CNIH4 protein interacts with GPCRs, promoting their traffic to the cell surface [41]. The high conservation of this subfamily allows us to hypothesize that all its members will act in GPCR trafficking to the cell surface, like their vertebrate counterparts. Instead, vertebrate members of the CNIH1/2/3 subfamily have different functions. While CNIH1 is involved in the maturation of TGF $\alpha$  [18], both CNIH2 and CNIH3 function as ARASs [33]. These three genes appeared by duplication in the vertebrate ancestor; thus, the phylogeny alone does not allow us to hypothesize about the function of the ancestral gene. Yet both the fruit fly and *C. elegans* present a single cornichon orthologue each [16,18]. Interestingly, the fly protein has been found to participate in the maturation of TGF $\alpha$  [18], as mammalian CNIH1s,



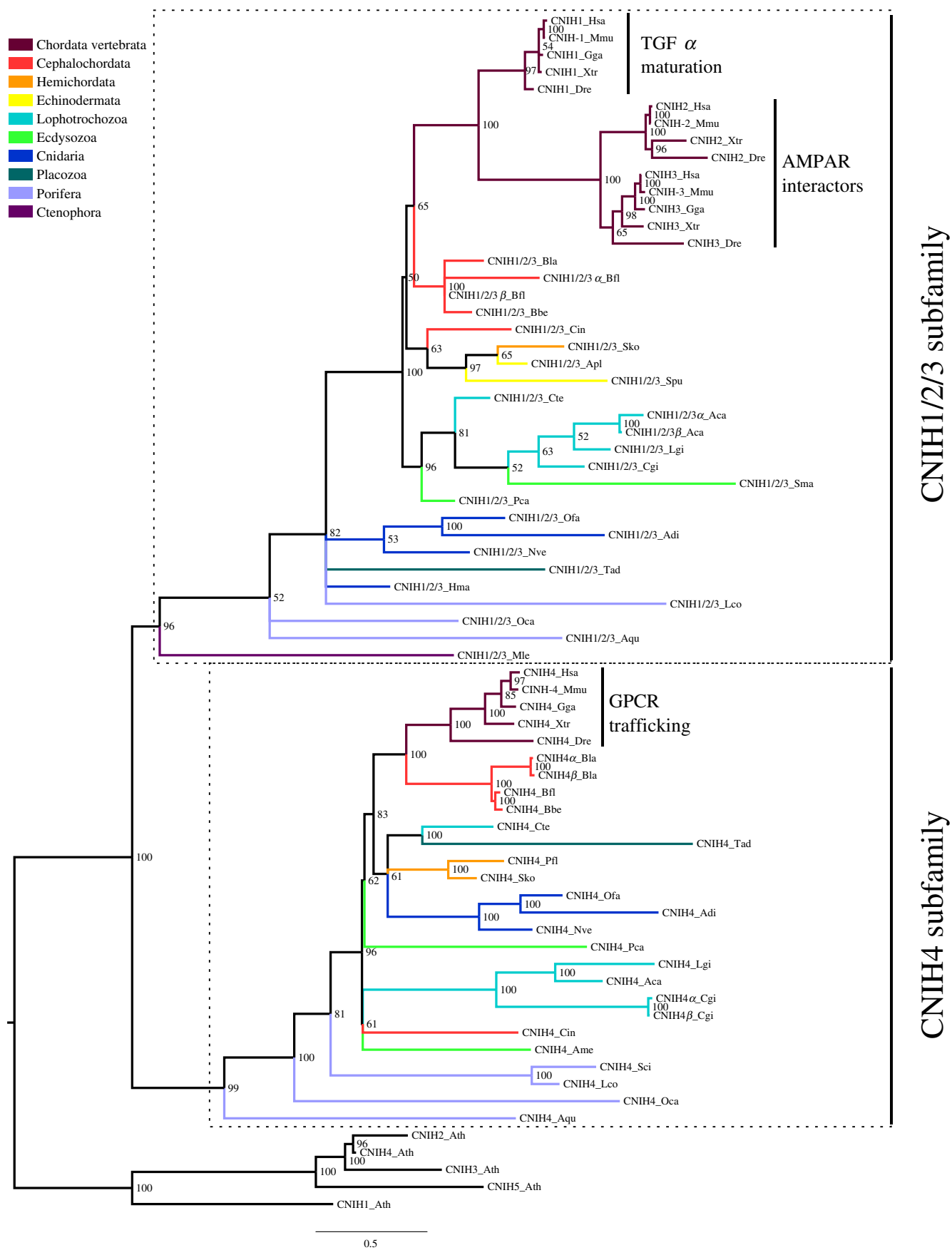
**Figure 2.** Multiple sequence alignment of the TM3, TM4 and PDZ binding motif of CACNG-GSG1 proteins. The sequence alignment includes sequences from the four subfamilies in the CACNG-GSG1 protein family. From vertebrate species, only *Homo sapiens* and *Danio rerio* sequences were included. In the TARPs subfamily, only representative sequences from invertebrate species are shown. A complete alignment of TARPs sequences can be seen in electronic supplementary material, figure S3. The name of the subfamilies is indicated at the left of sequence names. (a) Alignment of the TM3 and TM4 of CACNG-GSG1s proteins, which are implicated in the interaction with AMPAR. The extension of the transmembrane segments and the extracellular loop is marked on the top of the alignment. Residues involved in the highly conserved YSYGWSF motif are highlighted by a red box. The conservation of each position of the alignment is represented by an intensity gradient of the background, higher conservation corresponding to more intense blue, and by a bar chart at the bottom. Also the quality chart (a measure of the probability of seeing mutation in an alignment position), the consensus sequence and the occupancy chart are shown in the bottom. Residue numbering corresponds to human CACNG2 sequence. (b) The last eight residues of proteins containing a PDZ binding motif are shown. If no PDZ binding motif is found it is labelled as No PDZ. The class of each PDZ binding motif is also displayed. The figure was prepared with Jalview v2.11.0.

while the *C. elegans* orthologue, *cni-1*, acts as an ARAS. To the best of our knowledge the role of the fly protein as ARAS has not been investigated, nor has the role of *cni-1* been studied in the context of TFG $\alpha$  maturation. Nevertheless, we propose that the invertebrate orthologue of CNIH1, 2 and 3 performs both functions and that in the vertebrate lineage this gene was duplicated and subfunctionalized so that CNIH1 retained the role as a factor for TFG $\alpha$  maturation, and the ancestor of CNIH2 and 3 retained the ARAS function.

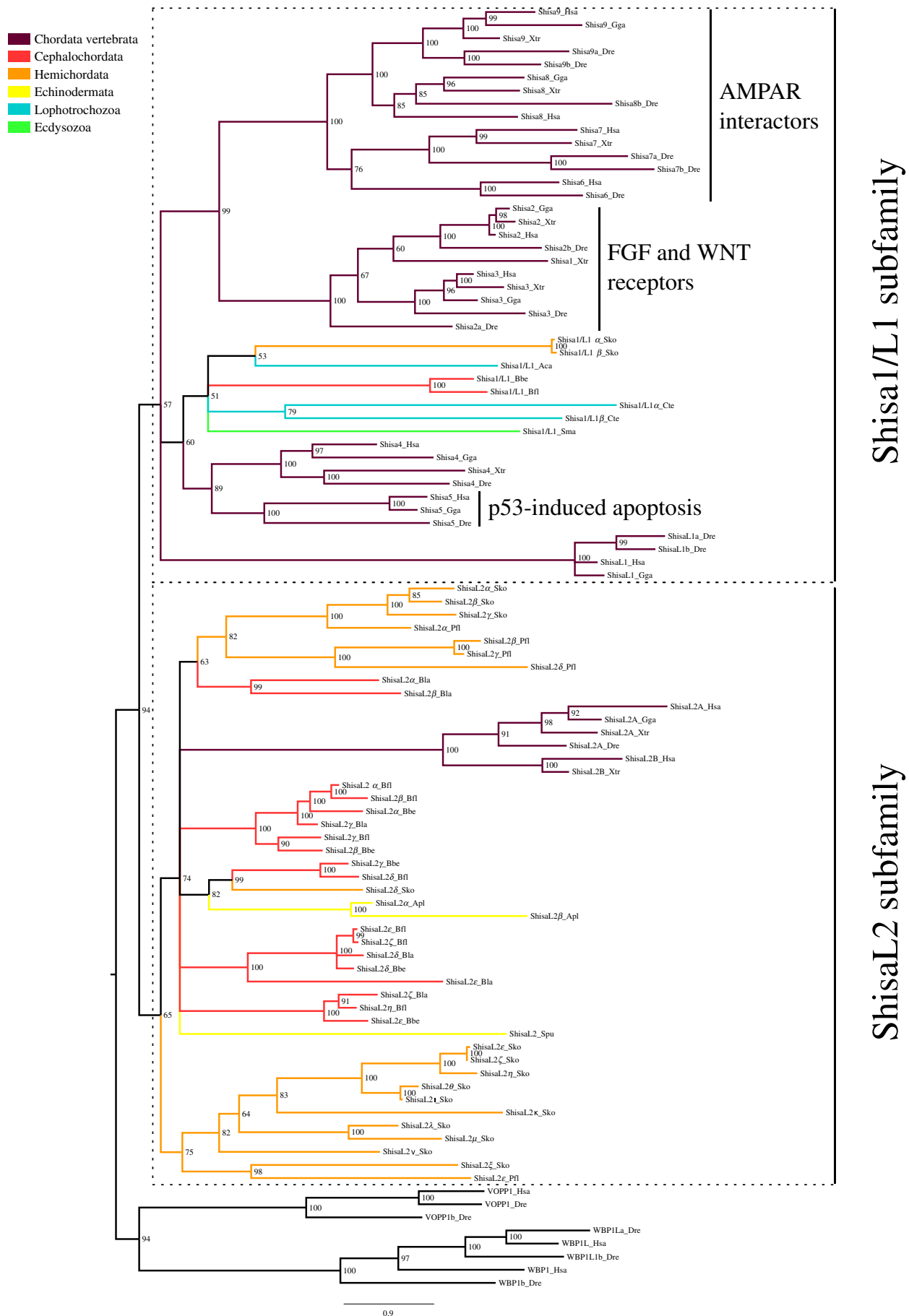
### 3.4. The phylogeny of the shisa family shows independent expansions in deuterostome species

Shisa proteins were only found among bilaterian species, although not in all of them. Species such as *Ciona intestinalis*, an urochordate, or *L. gigantea*, a mollusc, have lost this family (table 1). The phylogenies revealed that this family is organized in two subfamilies. The ancestral shisa gene would have appeared in bilaterians and duplicated before their diversification to generate these two subfamilies (electronic

supplementary material, figure S2C). Based on the proteins they contain, we have termed them Shisa1/L1 and ShisaL2 (figure 4; electronic supplementary material, figure S6). The Shisa1/L1 subfamily is better conserved, as it has only been lost in echinoderms; instead, the ShisaL2 is only found in deuterostome species, having been lost in the ancestor of protostomes. While the Shisa1/L1 greatly expanded in vertebrates, the ShisaL2 expanded in cephalochordates and hemichordates. Protostome species with shisas show a low number of sequences compared to chordates and *S. kowalevskii*, having 1 or 2 genes per species. The ML phylogeny shows how the Shisa1/L1 subfamily might be further divided in two classes: one comprising vertebrate Shisa4 and Shisa5 with invertebrate sequences and a second including vertebrate shisas: L1, 2, 3 and 6 to 9 (electronic supplementary material, figure S6). Nevertheless, the tree constructed with the BI method presents a different topology for the Shisa1/L1 subfamily (figure 4); this does not allow us to fully conclude that there would be two classes within the Shisa1/L1 subfamily, one being specific to vertebrates, as we found for Type I TARPs.



**Figure 3.** Bayesian inference phylogenetic tree of the cornichon family. The two subfamilies in which the cornichon family is divided are highlighted by dashed line boxes. The name of each subfamily is presented at the right of the box. Posterior probabilities are shown at tree nodes and protein names at the end of each branch. Tree branches are coloured based on phylum, as indicated in the legend. Cornichon proteins from *Arabidopsis thaliana* were used as outgroup. When known the function of vertebrate sequences is indicated. Scale bar denotes number of amino acid substitutions per site. The amino acid substitution model used was Lg + G, the analysis ran for 10 000 000 generations, the final standard deviation was 0.004018 and the final potential scale reduction factor (PSRF) was 1.000.



**Figure 4.** Bayesian inference phylogenetic tree of the shisa family. The two subfamilies in which the shisa family is divided are highlighted by dashed line boxes. The name of each subfamily is presented at the right of the box. Posterior probabilities are shown at tree nodes and protein names at the end of each branch. Tree branches are coloured based on phylum, as indicated in the legend. The closely related vertebrate proteins VOPP and WBP1 were used as outgroup. The function of vertebrate sequences, if known, is indicated. Scale bar denotes number of amino acid substitutions per site. The amino acid substitution model used was Vt + I + G + F, the analysis ran for 10 000 000 generations, the final standard deviation was 0.005589 and the final potential scale reduction factor (PSRF) was 1.000.



Vertebrate proteins from the Shisa1/L1 perform very different biological functions. Shisa5 is involved in p53-induced apoptosis, Shisa2 in FGF receptor traffic, Shisa3 in WNT receptor function and Shisas 6 to 9 are ARASs. Due to its position in both phylogenies and its grouping with invertebrate shisas, we propose that Shisa5 would be phylogenetically closer to the ancestral shisa than the other paralogues and that the ancestral gene would also be involved in apoptosis. In this scenario the emergence of ARAS in this subfamily would be the result of neofunctionalization events occurring during vertebrate evolution in the branch of Shisa6 to Shisa9.

### 3.5. The Dispanin C family is poorly conserved among metazoans

SynDIG1, a member of the Dispanin C family, has recently been reported as an AMPAR auxiliary subunit [75]. Our phylogenies indicate that members of this family can be traced to the early-diverging phylum of Ctenophores (figure 5; electronic supplementary material, figure S7). Yet, this family has been lost in multiple lineages and species during metazoan evolution (table 1). Only half of the 26 species investigated present at least one member of the Dispanin C family. Our trees indicate that in metazoans this family is also divided into two subfamilies, which we have named Dispanin C1 and Dispanin C2. The *L. gigantea* DispaninC1/2\_Lgi does not belong to any subfamily, since we didn't find other proteins in the same tree branch we did not define a new subfamily for it. The ctenophore sequence Dispanin\_C\_Mle is the first to diverge, not belonging to any subfamily; thus we propose that the ancestor of all metazoans had a single gene coding for Dispanin C and after the split of ctenophores it duplicated, giving rise to the two subfamilies (electronic supplementary material, figure S2D). Virtually nothing is known about the two vertebrate paralogues of SynDIG1, SynDIG1L and TMEM91, although SynDIG1L has been found downregulated in mouse models of Huntington's disease [57]. Due to the lack of functional information in other vertebrate or invertebrate Dispanin Cs it not possible to establish when SynDIG1 function as ARAS arose during evolution.

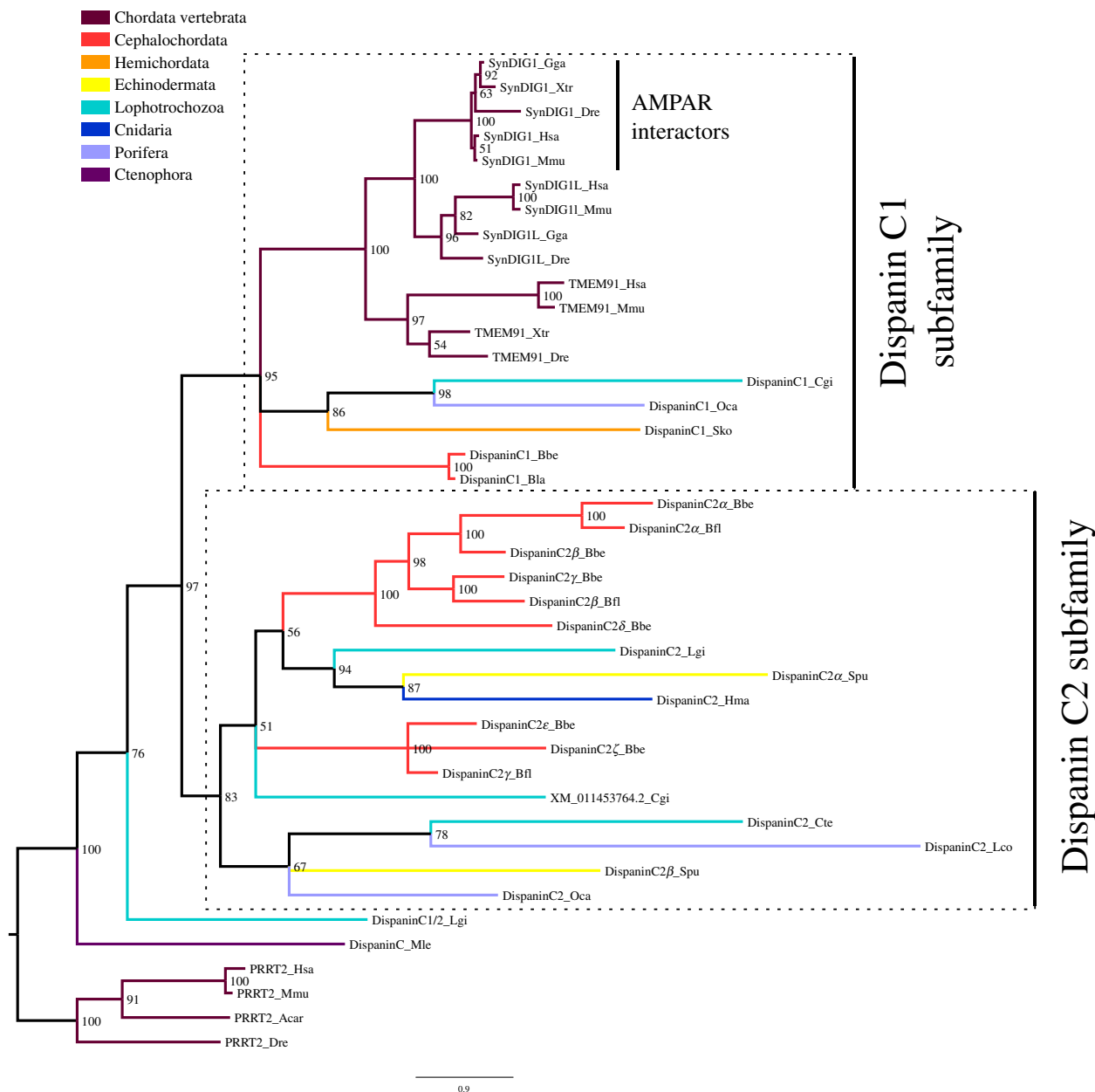
## 4. Discussion

Evolutionary studies have demonstrated that the synaptic proteome has importantly increased in vertebrates [76], expanding the molecular tools available to synaptic physiology [77–79] and multiplying synaptic molecular types [80,81]. The two rounds of whole-genome duplication (2R) that occurred at the base of the vertebrate lineage [82] were the major driving force behind this expansion, as genes expressed at the synapse were retained at high frequencies after these duplication events [83]. The result of this increased complexity in the vertebrate synaptic proteome has been associated with the higher cognitive functions found in mammals [84]. Nevertheless, there are exceptions to this general model of synaptic proteome evolution. This is the case of glutamate receptors, key nervous system proteins driving the excitatory synaptic transmission. These proteins have undergone a highly sophisticated evolutionary pattern in animals, with many lineage-specific gains, losses and expansions of entire classes of these receptors. This process has resulted in a similar number of glutamate receptors being present in most animal species, regardless of

the complexity of their nervous system [2]. Here, we present how the neo/subfunctionalization of unrelated proteins into ARASs has resulted in a different evolutionary strategy to increase the synaptic proteome in vertebrates.

Mammalian ARASs were thought to belong to five protein families: CACNG, GSG1, cornichon, shisa and Dispanin C. CACNGs and GSG1s are known to belong to the superfamily of claudins [85], a large group of proteins that presents over 40 members in mammals and that includes three families, one of them being that of CACNGs [86]. Nevertheless, the exact position of GSG1s within claudins was unknown. Our phylogenetic analysis indicates that GSG1s actually belong to the CACNG family of claudins. We thus refer to this family as CACNG-GSG1. Therefore, ARASs are organized into four evolutionarily unrelated protein families. In contrast with what we reported for the family of ionotropic glutamate receptors, which diversified into 12 phylogenetic groups, including four subfamilies and ten classes [2], the animal evolution of the families containing ARAS has not been particularly complex. cornichon, shisa and Dispanin C families can be divided into just two phylogenetic groups (subfamilies) and the CACNG-GSG1 family in four. Nevertheless, all these families have increased their members along animal evolution, and vertebrates have especially increased their number of ARAS. This is particularly true for the CACNG-GSG1 and shisa families, which include 11 and 10 proteins, respectively, in the vertebrate species investigated, of which 7 and 4 are ARASs. Altogether vertebrates generally present 14 ARASs, while, based on our study, invertebrate bilaterals would have fewer, between 1 and 7, and non-bilaterals even fewer, 1 or 2. Sequences from invertebrate bilaterals only fall confidently in two subfamilies with experimentally identified ARAS: that of TARPs within the CACNG-GSG1 family, and that of CNIH1/2/3 among cornichons. Furthermore, non-bilaterals would only present proteins phylogenetically related to the ARAS subfamily of CNIH1/2/3 cornichons. Additionally, a number of invertebrates, including the basal sponge *O. carmela*, could have retained one Dispanin C1 phylogenetically close to the mammalian ARAS SynDIG1, although the phylogenies are not fully conclusive in this regard. The loss of the shisa and Dispanin C families in multiple invertebrate species, and even in entire phyla, suggests a less relevant role of these proteins in invertebrates as compared with vertebrates, which present high conservation levels in both families. Furthermore, our analysis suggests that ARAS from the shisa family are only present in vertebrates, postulating them as an innovation of this lineage.

As ARASs belong to four unrelated protein families that include proteins with other functions, we aimed at using our phylogenetic study to propose when in evolution these proteins acquired their function as AMPAR modulators, although functional information would be required to completely establish their role. Our data indicate that all proteins identified in the TARP subfamily might function as ARASs, which would mean that the TARP subfamily would be the only one in which their proteins are solely dedicated to modulating AMPAR function. On the other hand, we found a possible example of vertebrate neofunctionalization of ARASs in the subfamily of GSG1Ls, as its vertebrate paralogue GSG1 is involved in trafficking of TPAP. Identifying the function of non-vertebrate GSG1s would be required to establish more conclusively this event of neofunctionalization. Vertebrates have two cornichons acting as ARASs (CNIH2



**Figure 5.** Bayesian inference phylogenetic tree of the Dispanin C family. The two subfamilies in which the Dispanin C family is divided are highlighted by dashed line boxes. The name of each subfamily is presented at the right of the box. Posterior probabilities are shown at tree nodes and protein names at the end of each branch. Tree branches are coloured based on phylum, as indicated in the legend. The vertebrate proteins PRRT2 from the Dispanin B family were used as outgroup. The function of vertebrate sequences, if known, is also indicated. Scale bar denotes number of amino acid substitutions per site. The amino acid substitution model used was Vt + I + G + F, the analysis ran for 10 000 000 generations, the final standard deviation was 0.003312 and the final potential scale reduction factor (PSRF) was 1.000.

and 3), emerging from a vertebrate-specific duplication. Their closest paralogue is CNIH1, which is involved in TGF $\alpha$  maturation. Interestingly, the fly sole orthologue to CNIH1, 2 and 3 is known to participate in the maturation of TGF $\alpha$  [18], while the unique orthologue of these proteins in *C. elegans* has a well-proven role as an ARAS [16]. We thus propose that invertebrate orthologues of CNIH1/2/3 perform both functions and that vertebrate paralogues underwent a process of subfunctionalization by which CNIH1 retained the TGF $\alpha$  maturation function and the ancestor of CNIH2 and 3 retained the ARAS role. Within the shisa family all known ARASs (Shisa6 to 9) are in the Shisa1/L1 subfamily, yet this phylogenetic group also contains proteins performing other functions, as Shisa5 participates in p53-induced apoptosis [49] and Shisa2 and 3, the closest paralogues to Shisas6–9 with known function, are involved in FGF [47] and WNT [48] receptor trafficking, respectively. In opposition to what

we found in the CACNG-GSG1 and cornichon subfamilies containing ARAS, the Shisa1/L1 subfamily includes very few sequences from non-vertebrates. Thereby, TARPs and CNIH1/2/3 subfamilies present 31 and 23 sequences from invertebrate species, respectively, while the Shisa1/L1 subfamily includes only eight. Importantly, all these eight sequences fall in the branch of Shisa5, which is not an ARAS. Thus, due to the topology of the phylogenetic trees we propose that Shisa6 to 9 would have experienced a process of neo/subfunctionalization early in the vertebrate lineage to become ARASs. Finally, the phylogenetic evolution of the Dispanin C family would also suggest a neo/subfunctionalization process in the vertebrate lineage resulting in members of the C1 subfamily becoming ARASs, although, as before, research on vertebrate and invertebrate orthologues of SynDIG1 will be required to fully validate this hypothesis. The fact that large-scale proteomics experiments of synaptic

preparations [77–79] and repositories of synaptic proteins [87] do not identify SynDIG1L or TMEM91 suggests that they are unlikely to function as ARASs, thus supporting the hypothesis of a neo/subfunctionalization process of ARASs in this family.

In summary, this study reveals that the large set of ARASs found in vertebrates is absent from other species. Interestingly, the surge in ARASs happened much later than the emergence of AMPARs [2]. The class of AMPA receptors appears and diversifies in the ancestor of bilateral species, around 800 Ma, while the increase in ARAS occurs early in vertebrate evolution, approximately 400 Ma [11]. Our analysis suggests that this increase is due to an expansion of proteins belonging to the TARP subfamily, the neofunctionalization of a group of Shisa1/L1 subfamily proteins and parallel processes of neo/subfunctionalization occurring in the cornichon and Dispanin C families. All these resulted in the recruitment of a large number of ARASs early in

vertebrate evolution, which suggests that these proteins might have importantly contributed to the development of the complex nervous systems found in these animals. The parallel recruitment of unrelated proteins to perform synaptic functions represents another strategy by which evolution has favoured an increased complexity in the synaptic proteome.

**Data accessibility.** This article has no additional data.

**Authors' contributions.** D.R.-V. performed the analysis. D.R.-V. and À.B. designed the study analysed the data and wrote the manuscript. À.B. allocated funding to the project.

**Competing interests.** The authors declare no conflict of interest.

**Funding.** D.R.-V. and À.B.'s financial support was provided by: BFU2012-34398, BFU2015-69717-P and RTI2018-097037-B-I00 (MINECO), Career Integration Grant (ref. 304111), Ramón y Cajal Fellowship (RYC-2011-08391p), IEDI-2017-00822 and AGAUR (SGR14-297 and 2017 SGR 1776).

**Acknowledgements.** Our gratitude to Dr David Soto (Universitat de Barcelona) who provided useful advice on this research.

## References

- Sobolevsky AI, Rosconi MP, Gouaux E. 2009 X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor. *Nature* **462**, 745–756. (doi:10.1038/nature08624)
- Ramos-Vicente D *et al.* 2018 Metazoan evolution of glutamate receptors reveals unreported phylogenetic groups and divergent lineage-specific events. *Elife* **7**, e35774. (doi:10.7554/eLife.35774)
- Hodge RD *et al.* 2019 Conserved cell types with divergent features in human versus mouse cortex. *Nature* **573**, 61–68. (doi:10.1038/s41586-019-1506-7)
- Traynelis SF *et al.* 2010 Glutamate receptor ion channels: structure, regulation, and function. *Pharmacol. Rev.* **62**, 405–496. (doi:10.1124/pr.109.002451)
- Jackson AC, Nicoll RA. 2011 The expanding social network of ionotropic glutamate receptors: TARPs and other transmembrane auxiliary subunits. *Neuron* **70**, 178–199. (doi:10.1016/j.neuron.2011.04.007)
- Haering SC, Tapken D, Pahl S, Hollmann M. 2014 Auxiliary subunits: shepherding AMPA receptors to the plasma membrane. *Membranes (Basel)* **4**, 469–490. (doi:10.3390/membranes4030469)
- Ng D *et al.* 2009 Neto1 is a novel CUB-domain NMDA receptor-interacting protein required for synaptic plasticity and learning. *PLoS Biol.* **7**, e41. (doi:10.1371/journal.pbio.1000041)
- Zhang W *et al.* 2009 A transmembrane accessory subunit that modulates kainate-type glutamate receptors. *Neuron* **61**, 385–396. (doi:10.1016/j.neuron.2008.12.014)
- Tang M, Pelkey KA, Ng D, Ivakine E, McBain CJ, Salter MW, McInnes RR. 2011 Neto1 is an auxiliary subunit of native synaptic kainate receptors. *J. Neurosci.* **31**, 10 009–10 018. (doi:10.1523/JNEUROSCI.6617-10.2011)
- Gogarten JP, Olendzki L. 1999 Orthologs, paralogs and genome comparisons. *Curr. Opin. Genet. Dev.* **9**, 630–636. (doi:10.1016/s0959-437x(99)00029-5)
- Kumar S, Stecher G, Suleski M, Hedges SB. 2017 TimeTree: a resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819. (doi:10.1093/molbev/msx116)
- Hopkins DH, Rane RV, Younus F, Coppin CW, Pandey G, Jackson CJ, Oakeshott JG. 2019 The molecular basis for the neofunctionalization of the juvenile hormone esterase duplication in *Drosophila*. *Insect Biochem. Mol. Biol.* **106**, 10–18. (doi:10.1016/j.ibmb.2019.01.001)
- He X, Zhang J. 2005 Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* **169**, 1157–1164. (doi:10.1534/genetics.104.037051)
- Lambert MJ, Cochran WO, Wilde BM, Olsen KG, Cooper CD. 2015 Evidence for widespread subfunctionalization of splice forms in vertebrate genomes. *Genome Res.* **25**, 624–632. (doi:10.1101/gr.184473.114)
- Innan H, Kondrashov F. 2010 The evolution of gene duplications: classifying and distinguishing between models. *Nat. Rev. Genet.* **11**, 97–108. (doi:10.1038/nrg2689)
- Brockie PJ *et al.* 2013 Cornichons control ER export of AMPA receptors to regulate synaptic excitability. *Neuron* **80**, 129–142. (doi:10.1016/j.neuron.2013.07.028)
- Wang R, Walker CS, Brockie PJ, Francis MM, Mellem JE, Madsen DM, Maricq AV. 2008 Evolutionary conserved role for TARPs in the gating of glutamate receptors and tuning of synaptic function. *Neuron* **59**, 997–1008. (doi:10.1016/j.neuron.2008.07.023)
- Bökel C, Dass S, Wilsch-Bräuninger M, Roth S. 2006 *Drosophila* Cornichon acts as cargo receptor for ER export of the TGF $\alpha$ -like growth factor Gurken. *Development* **133**, 459–470. (doi:10.1242/dev.02219)
- Walker CS, Brockie PJ, Madsen DM, Francis MM, Zheng Y, Koduri S, Mellem JE, Strutz-Seebohm N, Maricq AV. 2006 Reconstitution of invertebrate glutamate receptor function depends on stargazin-like proteins. *Proc. Natl Acad. Sci. USA* **103**, 10 781–10 786. (doi:10.1073/pnas.0604482103)
- Letts VA *et al.* 1998 The mouse stargazer gene encodes a neuronal Ca<sup>2+</sup>-channel gamma subunit. *Nat. Genet.* **19**, 340–347. (doi:10.1038/1228)
- Chen L, Chetkovich DM, Petralia RS, Sweeney NT, Kawasaki Y, Wenthold RJ, Brecht DS, Nicoll RA. 2000 Stargazin regulates synaptic targeting of AMPA receptors by two distinct mechanisms. *Nature* **408**, 936–943. (doi:10.1038/35050030)
- Hashimoto K, Fukaya M, Qiao X, Sakimura K, Watanabe M, Kano M. 1999 Impairment of AMPA receptor function in cerebellar granule cells of ataxic mutant mouse stargazer. *J. Neurosci.* **19**, 6027–6036. (doi:10.1523/JNEUROSCI.19-14-06027.1999)
- Klugbauer N, Dai S, Specht V, Lacinová L, Marais E, Bohn G, Hofmann F. 2000 A family of gamma-like calcium channel subunits. *FEBS Lett.* **470**, 189–197. (doi:10.1016/s0014-5793(00)01306-5)
- Burgess DL, Gefrides LA, Foreman PJ, Noebels JL. 2001 A cluster of three novel Ca<sup>2+</sup> channel gamma subunit genes on chromosome 19q13.4: evolution and expression profile of the gamma subunit gene family. *Genomics* **71**, 339–350. (doi:10.1006/geno.2000.6440)
- Chu PJ, Robertson HM, Best PM. 2001 Calcium channel gamma subunits provide insights into the evolution of this gene family. *Gene* **280**, 37–48. (doi:10.1016/s0378-1119(01)00738-7)
- Hansen JP, Chen R-S, Larsen JK, Chu P-J, Janes DM, Weis KE, Best PM. 2004 Calcium channel gamma6 subunits are unique modulators of low voltage-activated (Cav3.1) calcium current. *J. Mol. Cell. Cardiol.* **37**, 1147–1158. (doi:10.1016/j.yjmcc.2004.08.005)

27. Tomita S, Chen L, Kawasaki Y, Petralia RS, Wenthold RJ, Nicoll RA, Brecht DS. 2003 Functional studies and distribution define a family of transmembrane AMPA receptor regulatory proteins. *J. Cell Biol.* **161**, 805–816. (doi:10.1083/jcb.200212116)
28. Kato AS *et al.* 2007 New transmembrane AMPA receptor regulatory protein isoform, gamma-7, differentially regulates AMPA receptors. *J. Neurosci.* **27**, 4969–4977. (doi:10.1523/JNEUROSCI.5561-06.2007)
29. Soto D, Coombs ID, Renzi M, Zonouzi M, Farrant M, Cull-Candy SG. 2009 Selective regulation of long-form calcium-permeable AMPA receptors by an atypical TARP, gamma-5. *Nat. Neurosci.* **12**, 277–285. (doi:10.1038/nn.2266)
30. Deng F, Price MG, Davis CF, Mori M, Burgess DL. 2006 Stargazin and other transmembrane AMPA receptor regulating proteins interact with synaptic scaffolding protein MAGI-2 in brain. *J. Neurosci.* **26**, 7875–7884. (doi:10.1523/JNEUROSCI.1851-06.2006)
31. Tomita S. 2010 Regulation of ionotropic glutamate receptors by their auxiliary subunits. *Physiology (Bethesda)* **25**, 41–49. (doi:10.1152/physiol.00033.2009)
32. Shanks NF *et al.* 2012 Differences in AMPA and kainate receptor interactomes facilitate identification of AMPA receptor auxiliary subunit GSG1L. *Cell Rep* **1**, 590–598. (doi:10.1016/j.celrep.2012.05.004)
33. Schwenk J *et al.* 2012 High-resolution proteomics unravel architecture and molecular diversity of native AMPA receptor complexes. *Neuron* **74**, 621–633. (doi:10.1016/j.neuron.2012.03.034)
34. Choi H-S, Lee S-H, Kim H, Lee Y. 2008 Germ cell-specific gene 1 targets testis-specific poly(A) polymerase to the endoplasmic reticulum through protein-protein interactions. *FEBS Lett.* **582**, 1203–1209. (doi:10.1016/j.febslet.2008.01.065)
35. Gu X, Mao X, Lussier MP, Hutchison MA, Zhou L, Hamra FK, Roche KW, Lu W. 2016 GSG1L suppresses AMPA receptor-mediated synaptic transmission and uniquely modulates AMPA receptor kinetics in hippocampal neurons. *Nat Commun* **7**, 10873. (doi:10.1038/ncomms10873)
36. Mao X, Gu X, Lu W. 2017 GSG1L regulates the strength of AMPA receptor-mediated synaptic transmission but not AMPA receptor kinetics in hippocampal dentate granule neurons. *J. Neurophysiol.* **117**, 28–35. (doi:10.1152/jn.00307.2016)
37. Shi Y, Suh YH, Milstein AD, Iozaki K, Schmid SM, Roche KW, Nicoll RA. 2010 Functional comparison of the effects of TARPs and cornichons on AMPA receptor trafficking and gating. *Proc. Natl Acad. Sci. USA* **107**, 16 315–16 319. (doi:10.1073/pnas.1011706107)
38. Kato AS *et al.* 2010 Hippocampal AMPA receptor gating controlled by both TARP and cornichon proteins. *Neuron* **68**, 1082–1096. (doi:10.1016/j.neuron.2010.11.026)
39. Coombs ID, Soto D, Zonouzi M, Renzi M, Shelley C, Farrant M, Cull-Candy SG. 2012 Cornichons modify channel properties of recombinant and glial AMPA receptors. *J. Neurosci.* **32**, 9796–9804. (doi:10.1523/JNEUROSCI.0345-12.2012)
40. Castro CP, Piscopo D, Nakagawa T, Derynck R. 2007 Cornichon regulates transport and secretion of TGFalpha-related proteins in metazoan cells. *J. Cell. Sci.* **120**, 2454–2466. (doi:10.1242/jcs.004200)
41. Sauvageau E, Rochdi MD, Oueslati M, Hamdan FF, Percherancier Y, Simpson JC, Pepperkok R, Bouvier M. 2014 CNH4 interacts with newly synthesized GPCR and controls their export from the endoplasmic reticulum. *Traffic* **15**, 383–400. (doi:10.1111/tra.12148)
42. von Engelhardt J, Mack V, Sprengel R, Kavenstock N, Li KW, Stern-Bach Y, Smit AB, Seeburg PH, Monyer H. 2010 CKAMP44: a brain-specific protein attenuating short-term synaptic plasticity in the dentate gyrus. *Science* **327**, 1518–1522. (doi:10.1126/science.1184178)
43. Farrow P, Khodosevich K, Sapir Y, Schulmann A, Aslam M, Stern-Bach Y, Monyer H, von Engelhardt J. 2015 Auxiliary subunits of the CKAMP family differentially modulate AMPA receptor properties. *Elife* **4**, e09693. (doi:10.7554/eLife.09693)
44. Khodosevich K, Jacobi E, Farrow P, Schulmann A, Rusu A, Zhang L, Sprengel R, Monyer H, von Engelhardt J. 2014 Coexpressed auxiliary subunits exhibit distinct modulatory profiles on AMPA receptor function. *Neuron* **83**, 601–615. (doi:10.1016/j.neuron.2014.07.004)
45. Klaassen RV *et al.* 2016 Shisa6 traps AMPA receptors at postsynaptic sites and prevents their desensitization during synaptic activity. *Nat. Commun.* **7**, 10682. (doi:10.1038/ncomms10682)
46. von Engelhardt J. 2019 AMPA receptor auxiliary proteins of the CKAMP family. *Int. J. Mol. Sci.* **20**, 1460. (doi:10.3390/ijms20061460)
47. Hedge TA, Mason I. 2008 Expression of Shisa2, a modulator of both Wnt and Fgf signaling, in the chick embryo. *Int. J. Dev. Biol.* **52**, 81–85. (doi:10.1387/ijdb.072355th)
48. Chen C-C *et al.* 2014 Shisa3 is associated with prolonged survival through promoting  $\beta$ -catenin degradation in lung cancer. *Am. J. Respir. Crit. Care Med.* **190**, 433–444. (doi:10.1164/rccm.201312-2256OC)
49. Bourdon J-C, Renzing J, Robertson PL, Fernandes KN, Lane DP. 2002 Scotin, a novel p53-inducible proapoptotic protein located in the ER and the nuclear membrane. *J. Cell Biol.* **158**, 235–246. (doi:10.1083/jcb.200203006)
50. Kalashnikova E, Lorca RA, Kaur I, Barisono GA, Li B, Ishimaru T, Trimmer JS, Mohapatra DP, Diaz E. 2010 SynDIG1: an activity-regulated, AMPA-receptor-interacting transmembrane protein that regulates excitatory synapse development. *Neuron* **65**, 80–93. (doi:10.1016/j.neuron.2009.12.021)
51. Diaz E. 2010 SynDIG1 regulation of synaptic AMPA receptor targeting. *Commun. Integr. Biol.* **3**, 347–349. (doi:10.4161/cib.3.4.11765)
52. Wudick MM *et al.* 2018 CORNICHON sorting and regulation of GLR channels underlie pollen tube  $\text{Ca}^{2+}$  homeostasis. *Science* **360**, 533–536. (doi:10.1126/science.aar6464)
53. Powers J, Barlowe C. 1998 Transport of axl2p depends on erv14p, an ER-vesicle protein related to the *Drosophila* cornichon gene product. *J. Cell Biol.* **142**, 1209–1222. (doi:10.1083/jcb.142.5.1209)
54. Fahey B, Degnan BM. 2010 Origin of animal epithelia: insights from the sponge genome. *Evol. Dev.* **12**, 601–617. (doi:10.1111/j.1525-142X.2010.00445.x)
55. Leys SP, Riesgo A. 2012 Epithelia, an evolutionary novelty of metazoans. *J. Exp. Zool. B Mol. Dev. Evol.* **318**, 438–447. (doi:10.1002/jez.b.21442)
56. Pei J, Grishin NV. 2012 Unexpected diversity in Shisa-like proteins suggests the importance of their roles as transmembrane adaptors. *Cell. Signal.* **24**, 758–769. (doi:10.1016/j.cellsig.2011.11.011)
57. Sällman AM, Bringeland N, Fredriksson R, Schiöth HB. 2012 The dispanins: a novel gene family of ancient origin that contains 14 human members. *PLoS ONE* **7**, e31961. (doi:10.1371/journal.pone.0031961)
58. Simakov O *et al.* 2015 Hemichordate genomes and deuterostome origins. *Nature* **527**, 459–465. (doi:10.1038/nature16150)
59. Mao Y, Economo EP, Satoh N. 2018 The roles of introgression and climate change in the rise to dominance of *Acropora* corals. *Curr. Biol.* **28**, 3373–3382. (doi:10.1016/j.cub.2018.08.061)
60. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990 Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410. (doi:10.1016/S0022-2836(05)80360-2)
61. Edgar RC. 2004 MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113. (doi:10.1186/1471-2105-5-113)
62. Kumar S, Stecher G, Li M, Niyaz C, Tamura K. 2018 MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549. (doi:10.1093/molbev/msy096)
63. Darriba D, Taboada GL, Doallo R, Posada D. 2011 ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics* **27**, 1164–1165. (doi:10.1093/bioinformatics/btr088)
64. Ronquist F *et al.* 2012 MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542. (doi:10.1093/sysbio/sys029)
65. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015 IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* **32**, 268–274. (doi:10.1093/molbev/msu300)
66. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018 UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522. (doi:10.1093/molbev/msx281)
67. Miller MA, Pfeiffer W, Schwartz T. 2010 Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *2010 Gateway Computing Environments Workshop (GCE)*, pp. 1–8.
68. Kumar M *et al.* 2020 ELM—the eukaryotic linear motif resource in 2020. *Nucleic Acids Res.* **48**, D296–D306. (doi:10.1093/nar/gkz1030)

69. Price MG, Davis CF, Deng F, Burgess DL. 2005 The alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionate receptor trafficking regulator 'stargazin' is related to the claudin family of proteins by its ability to mediate cell–cell adhesion. *J. Biol. Chem.* **280**, 19 711–19 720. (doi:10.1074/jbc.M500623200)
70. Adato A *et al.* 2002 USH3A transcripts encode clarin-1, a four-transmembrane-domain protein with a possible role in sensory synapses. *Eur. J. Hum. Genet.* **10**, 339–350. (doi:10.1038/sj.ejhg.5200831)
71. Kato AS, Siuda ER, Nisenbaum ES, Brecht DS. 2008 AMPA receptor subunit-specific regulation by a distinct family of type II TARPs. *Neuron* **59**, 986–996. (doi:10.1016/j.neuron.2008.07.034)
72. Kim E, Sheng M. 2004 PDZ domain proteins of synapses. *Nat. Rev. Neurosci.* **5**, 771–781. (doi:10.1038/nrn1517)
73. Dakoji S, Tomita S, Karimzadegan S, Nicoll RA, Brecht DS. 2003 Interaction of transmembrane AMPA receptor regulatory proteins with multiple membrane associated guanylate kinases. *Neuropharmacology* **45**, 849–856. (doi:10.1016/s0028-3908(03)00267-3)
74. Ernst A, Appleton BA, Ivarsson Y, Zhang Y, Gfeller D, Wiesmann C, Sidhu SS. 2014 A structural portrait of the PDZ domain family. *J. Mol. Biol.* **426**, 3509–3519. (doi:10.1016/j.jmb.2014.08.012)
75. Díaz E *et al.* 2002 Molecular analysis of gene expression in the developing pontocerebellar projection system. *Neuron* **36**, 417–434. (doi:10.1016/s0896-6273(02)01016-4)
76. Emes RD, Grant SGN. 2012 Evolution of synapse complexity and diversity. *Annu. Rev. Neurosci.* **35**, 111–131. (doi:10.1146/annurev-neuro-062111-150433)
77. Bayés A, Collins MO, Croning MDR, van de Lagemaat LN, Choudhary JS, Grant SGN. 2012 Comparative study of human and mouse postsynaptic proteomes finds high compositional conservation and abundance differences for key synaptic proteins. *PLoS ONE* **7**, e46683. (doi:10.1371/journal.pone.0046683)
78. Emes RD *et al.* 2008 Evolutionary expansion and anatomical specialization of synapse proteome complexity. *Nat. Neurosci.* **11**, 799–806. (doi:10.1038/nn.2135)
79. Bayés A, van de Lagemaat LN, Collins MO, Croning MDR, Whittle IR, Choudhary JS, Grant SGN. 2011 Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nat. Neurosci.* **14**, 19–21. (doi:10.1038/nn.2719)
80. Zhu F *et al.* 2018 Architecture of the mouse brain synaptome. *Neuron* **99**, 781–799. (doi:10.1016/j.neuron.2018.07.007)
81. Cizeron M, Qiu Z, Koniaris B, Gokhale R, Komiyama NH, Fransén E, Grant SGN. 2020 A brainwide atlas of synapses across the mouse life span. *Science* **369**, 270–275. (doi:10.1126/science.aba3163)
82. Kasahara M. 2007 The 2R hypothesis: an update. *Curr. Opin. Immunol.* **19**, 547–552. (doi:10.1016/j.coi.2007.07.009)
83. Bayés À, Collins MO, Reig-Viader R, Gou G, Goulding D, Izquierdo A, Choudhary JS, Emes RD, Grant SGN. 2017 Evolution of complexity in the zebrafish synapse proteome. *Nat Commun.* **8**, 14613. (doi:10.1038/ncomms14613)
84. Nithianantharajah J *et al.* 2013 Synaptic scaffold evolution generated components of vertebrate cognitive complexity. *Nat. Neurosci.* **16**, 16–24. (doi:10.1038/nn.3276)
85. Chen S, Gouaux E. 2019 Structure and mechanism of AMPA receptor–auxiliary protein complexes. *Curr. Opin. Struct. Biol.* **54**, 104–111. (doi:10.1016/j.sbi.2019.01.011)
86. Maher GJ, Hilton EN, Urquhart JE, Davidson AE, Spencer HL, Black GC, Manson FD. 2011 The cataract-associated protein TMEM114, and TMEM235, are glycosylated transmembrane proteins that are distinct from claudin family members. *FEBS Lett.* **585**, 2187–2192. (doi:10.1016/j.febslet.2011.05.060)
87. Koopmans F *et al.* 2019 SynGO: An evidence-based, expert-curated knowledge base for the synapse. *Neuron* **103**, 217–234. (doi:10.1016/j.neuron.2019.05.002)