DEVELOPMENTAL DYNAMICS 239:688–702, 2010

Characterization of the Neuroligin Gene Family Expression and Evolution in Zebrafish

Alberto Rissone,[1,2]* Lorenzo Sangiorgio,[3] Matteo Monopoli,[4] Monica Beltrame,[5] Ileana Zucchi,[3] Federico Bussolino,[1,2†] Marco Arese,[1,2†] and Franco Cotelli[4†]

Neuroligins constitute a family of transmembrane proteins localized at the postsynaptic side of both excitatory and inhibitory synapses of the central nervous system. They are involved in synaptic function and maturation and recent studies have linked mutations in specific human Neuroligins to mental retardation and autism. We isolated the human Neuroligin homologs in *Danio rerio*. Next, we studied their gene structures and we reconstructed the evolution of the Neuroligin genes across vertebrate phyla. Using reverse-transcriptase polymerase chain reaction, we analyzed the expression and alternative splicing pattern of each gene during zebrafish embryonic development and in different adult organs. By in situ hybridization, we analyzed the temporal and spatial expression pattern during embryonic development and larval stages and we found that zebrafish Neuroligins are expressed throughout the nervous system. Globally, our results indicate that, during evolution, specific subfunctionalization events occurred within paralogous members of this gene family in zebrafish. *Developmental Dynamics 239:688–702, 2010.* © 2009 Wiley-Liss, Inc.

Key words: Neuroligin; zebrafish; alternative splicing; evolution

Accepted 20 November 2009

INTRODUCTION

Neuroligins are a family of synaptic type I transmembrane proteins that consist of a large extracellular portion similar to cholinesterases, a glycosylated linker sequence, a transmembrane domain, and a short C-terminal tail containing a type I PDZ-binding motif (Ichtchenko et al., 1995). Neuroligins belong to a family of molecules containing the cholinesterase-like domain, called cholinesterase-like adhesion molecules (CLAMs), which include glutactin, neurotactin, and gliotactin (Gilbert and Auld, 2005). However, unlike cholinesterases, Neuroligins lack one of the residues of the catalytic triad located within the extracellular esterase-like domain, which renders them enzymatically inactive. Thus, instead of mediating enzyme/substrate interplay, the Neuroligins' cholinesterase-like domain is involved in protein–protein interaction.

Neuroligin proteins have been identified in different species from invertebrates to human (Ichtchenko et al., 1995; Bolliger et al., 2001; Gilbert et al., 2001; Kwon et al., 2004; Biswas et al., 2008). The mouse genome encodes four Neuroligin family members while five different genes (*NLGN1, NLGN2, NLGN3, NLGN4X*, and *NLGN4Y*) are present in the human genome (Ichtchenko et al., 1996; Bolliger et al., 2001, 2008; Paraoanu et al., 2006). The extracellular portion of mammalian Neuroligins contains two conserved regions subjected to alternative splicing events, indicated as Site A and B (Ichtchenko et al., 1995, 1996; Bolliger et al., 2001; Paraoanu et al.,

Additional Supporting Information may be found in the online version of this article.

[1]Department of Oncological Sciences, University of Torino School of Medicine, Candiolo, Italy
[2]IRCC, Institute for Cancer Research and Treatment, University of Torino School of Medicine, Candiolo, Italy
[3]Istituto di Tecnologie Biomediche, Consiglio Nazionale delle Ricerche, Segrate, Italy
[4]Department of Biology, Università degli Studi di Milano, Milano, Italy
[5]Department of Biomolecular Sciences and Biotechnology, Università degli Studi di Milano, Milano, Italy

Grant sponsor: Associazione Italiana per la Ricerca sul Cancro (AIRC); Grant Sponsor: Fondazione CRT-Torino; Grant Sponsor: "Ministero della Salute"; Grant Sponsor: Fondazione Cariplo; Grant numbers: 2006.0807; Grant sponsor: Progetto Cariplo N.O.B.E.L.; Grant sponsor: Regione Piemonte (Ricerca Scientifica Applicata 2004 , A17).

†Federico Bussolino, Marco Arese, and Franco Cotelli contributed equally to the project.
***Correspondence to: Alberto Rissone, Department of Oncological Sciences, University of Torino, School of Medicine, 10060 Candiolo (TO), Italy. E-mail: alberto.rissone@unito.it**

DOI 10.1002/dvdy.22196
Published online 23 December 2009 in Wiley InterScience (www.interscience.wiley.com).

2006). Usually, Neuroligin 1 and Neuroligin 3 can present three different isoforms at site A: without inserts, with insert A1/A2, or with inserts A1+A2. Neuroligin 2 and Neuroligins 4 can only present the insert A2. Finally, only Neuroligin 1 can be alternatively spliced at Site B.

Neuroligins are localized at the postsynaptic side of both excitatory and inhibitory synapses of the central nervous system (CNS). They bind to presynaptic alpha and beta forms of Neurexins (Ichtchenko et al., 1995, 1996; Missler et al., 1998; Boucard et al., 2005) and the alternative splicing of both protein families controls their binding affinities (Boucard et al., 2005; Chih et al., 2006; Comoletti et al., 2006; Fabrichny et al., 2007; Koehnke et al., 2008a,b; Shen et al., 2008). The current data suggest that Neurexin–Neuroligin binding is governed by a complex code that is based on the type of isoforms and splice variants involved, calcium binding, and glycosylation (Sudhof, 2008).

Neuroligin 1 expression is mainly restricted to the CNS, while Neuroligins 2–4 present a broader expression pattern (Song et al., 1999; Philibert et al., 2000; Bolliger et al., 2001; Gilbert et al., 2001; Kang et al., 2004; Suckow et al., 2008). Some lines of evidence suggest that the expression of Neuroligins may differ between species. For example, rat Neuroligin 3 is mainly detected in brain, whereas human Neuroligin 3 and 4 are also expressed in different peripheral tissues (Ichtchenko et al., 1996; Nemeth et al., 1999; Philibert et al., 2000; Bolliger et al., 2001).

Numerous studies investigated the synaptic roles of Neuroligins and Neurexins. Although different works suggested their involvement in synapses formation (Scheiffele et al., 2000; Dean et al., 2003; Graf et al., 2004; Prange et al., 2004; Chubykin et al., 2005; Levinson et al., 2005), other studies and the analysis of knockout mice proposed that Neuroligins, along with the alpha-forms of Neurexins, are involved in synaptic function and maturation (Missler et al., 2003; Sara et al., 2005; Varoqueaux et al., 2006; Chubykin et al., 2007). Triple knockout mice lacking Nlgn1–3 die shortly after birth and present normal synapse numbers with an apparently normal ultrastructure (Varoqueaux et al., 2006). Moreover, the importance of Neuroligins in proper brain function is documented by the fact that mutations in the human Neuroligin 3 and Neuroligin 4 genes appear to be involved in mental retardation and autism (Jamain et al., 2003; Chih et al., 2004; Comoletti et al., 2004; Laumonnier et al., 2004; Talebizadeh et al., 2006).

In this study, we describe the isolation and characterization of Neuroligin genes in *Danio rerio*. The zebrafish Neuroligin gene family includes seven genes very similar to their human homologs, suggesting that, during evolution, they have been subjected to strong evolutionary pressure in order to preserve their function. Through a phylogenomic analysis of the intron/exon structure, we reconstructed the evolution of Neuroligin genes in vertebrates. Our data highlighted the presence of eleven intron gains and a unique intron loss event, mostly involving major branching points in the vertebrate tree such as the origin of mammals and teleosts. Moreover, our data confirm previous independent studies on vertebrate genomes indicating that intron gain is more prevalent than intron loss. Through Reverse-transcriptase polymerase chain reaction (RT-PCR)-based methods, we analyzed the alternative splicing pattern during embryo development and in adult organs finding that: (1) in adult fishes Neuroligins are expressed in many different organs, (2) in many cases different organs present specific alternative splicing patterns, and (3) subfunctionalization events occurred differentiating the expression pattern and the alternative splicing regulation of paralogous genes. Finally, gene expression analyses by whole mount in situ hybridization showed that Neuroligins are widely expressed in the brain of developing embryos. In particular, almost all genes share a similar expression pattern in the midbrain and the hindbrain. Nevertheless, we also find evidence of a differential expression between paralogous genes.

RESULTS AND DISCUSSION

Molecular Cloning and Characterization of the Zebrafish Neuroligins

The sequences of the human Neuroligins were assembled using ENSEMBL and VEGA and then used as queries in BLAST searches at the Zebrafish Genome Browser (www.ensembl.org/Danio_Rerio) and in public EST databases available at the NCBI (www.ncbi.nlm.nih.gov/BLAST). We found seven different genes homologous to human Neuroligins in different Linkage Groups (LGs) of the zebrafish genome. Using in silico cloning and rapid amplification of cDNA ends (RACE) techniques, we assembled the complete coding sequences (CDS) of zebrafish Neuroligins. Finally, we confirmed each CDS by RT-PCR and sequencing.

According to Zebrafish Nomenclature Guidelines (www.zfin.org), we designed these new genes as: *nlgn1* (partially located on LG11 and LG2 in Ensembl Zv8 assembly), *nlgn2a* (LG7), *nlgn2b* (LG10), *nlgn3a* (LG5), *nlgn3b* (LG14), *nlgn4a* (LG1), and *nlgn4b* (LG9). In zebrafish, Neuroligin 2–4 genes are duplicated; only the Neuroligin 1 gene seems to be present in a single copy and, notwithstanding extensive database searches, we did not find any traces of a duplicate form of this gene. The lack of a Neuroligin 1 paralogous gene in other teleosts such as *G. aculeatus, O. latipes, T. nigrovirids*, and *T. rubripes* (see below) strongly suggests that, after the whole genome duplication (Postlethwait, 2007), one copy quickly disappeared.

The multialignment of zebrafish and human Neuroligins amino acid sequences (Fig. 1) shows that they present common features. Similarly to human Neuroligins, zebrafish proteins are composed of an N-terminal signal peptide of variable length, an esterase-like domain encompassing almost the entire extracellular portion, a short linker just upstream the transmembrane domain, and a cytosolic region. As observed in mammalian Neuroligins, the highest degree of sequence conservation can be found in the esterase-like domain, in the transmembrane domain, and in the PDZ binding domain, which mediates their interaction with the scaffold protein PSD-95 (Hata et al., 1996; Barrow et al., 2009). Notably, the linker between the esterase-like domain and the transmembrane region is different among all the vertebrate Neuroligins. Like mammalian Neuroligins, each zebrafish protein presents a substitution (Gly for Ser) of one residue of the catalytic

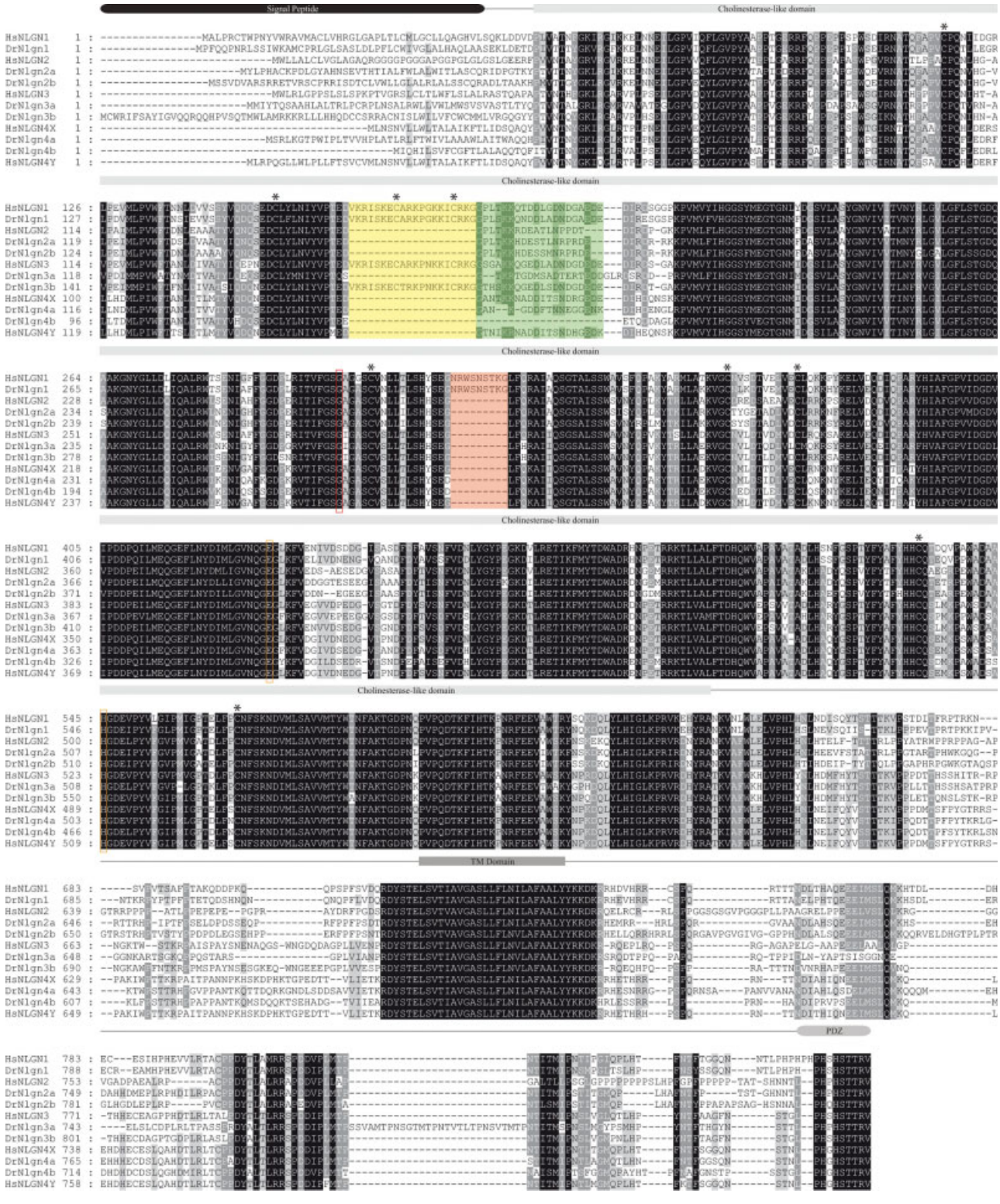

Fig. 1. Multiple alignment of human and zebrafish Neuroligin proteins. The amino acid sequences of human and zebrafish Neuroligins were aligned using AlignX of Vector NTI Advance 10.1.1 based on the Clustal W algorithm. Subsequently, the alignment was edited with GeneDoc using the conservation mode for shading. Consensus residues were assigned based on the number of occurrences of the character in the column. Different levels of shading were set to 100% conserved (black shading), 80% or greater conserved (dark grey shading), 60% or greater conserved (grey shading), and less than 60% conserved (no shading). The Ser, Glu/Asp, and His catalytic triad of esterases are boxed (in all Neuroligins the Ser that is replaced by Gly is marked in the with a red box). Alternative splice sites are color shaded in yellow (site A1), green (site A2), and red (site B). With the exception of *nlgn3a* and *nlgn4b*, which lack alternatively spliced exons at Site A, the alternative splicing mechanism is totally conserved. Conserved Cys residues are marked by asterisks. Predicted features of human Neuroligin 1 (HsNLGN1) structure were obtained using PROSITE and are depicted above the amino acid sequences. TM Domain, transmembrane domain; PDZ, PDZ interaction site. Hs and Dr, *Homo sapiens* and *Danio rerio*, respectively.

TABLE 1. Percentage of Identity and Conservation Between *Homo sapiens* and *Danio rerio* Homologs[a]

	DnNlgn1	DnNlgn2a	DnNlgn2b	DnNlgn3a	DnNlgn3b	DnNlgn4a	DnNlgn4b
HsNLGN1	**83 (91)**	67 (80)	65 (72)	63 (76)	69 (80)	71 (81)	71 (81)
HsNLGN2	64 (76)	**75 (83)**	**73 (82)**	60 (74)	62 (75)	62 (73)	65 (77)
HsNLGN3	69 (80)	64 (76)	64 (76)	**74 (84)**	**84 (90)**	71 (82)	70 (83)
HsNLGN4X	72 (82)	67 (79)	66 (78)	66 (80)	72 (83)	**84 (91)**	**78 (87)**
HsNLGN4Y	69 (80)	64 (77)	62 (76)	64 (78)	70 (83)	81 (89)	77 (86)

[a]Numbers in bold indicate identity (and conservation) percentages among each couple of homologs.

triad of esterases and it is, therefore, catalytically inactive. Moreover, the positions of all cysteine residues involved, through disulphide bridges, in the correct folding of these proteins are completely maintained in zebrafish proteins (asterisks in Fig. 1).

Table 1 shows the percentage of amino acid identity and conservation between mammalian and zebrafish homologs gathered from the multialignment in Figure 1. The identity and conservation values are higher than 70 and 80%, respectively, further supporting the identity of the cloned genes. As we previously observed comparing zebrafish and human Neurexin proteins (Rissone et al., 2007), also zebrafish Neuroligins are very similar to their human homologs, notwithstanding the evolutionary distance from the common ancestor (Aparicio et al., 2002). This high degree of sequence similarity can be explained by the presence of positive evolutionary pressure acting on both Neurexins and Neuroligins strongly suggesting a conservation of their functions across vertebrate evolution. Remarkably, both Neuroligins 4 (a and b) are more similar to human Neuroligin 4X than Neuroligin 4Y. The lack of one or more homologue of human *NLGN4Y* is coherent with the assumption that Neuroligin 4Y is a primate-specific gene that originated from a recent duplication (Sudhof, 2008). It is remarkable that, in spite of their global similarity, the different Neuroligins (from *Danio rerio* to *Homo sapiens*) are on the average only 60–70% identical (see Supporting Information Table S1, which is available online). This suggests an appreciable evolutionary distance among different genes (Nlgn1, Nlgn2, Nlgn3, and Nlgn4s) of the same family and it suggests a probable functional diversification.

To further support the data obtained by the multialignment, we performed phylogenetic (Fig. 2) and syntenic analyses (Supporting Information Fig. S1). The full-length nucleotide and amino acid sequences of zebrafish Neuroligins were used as queries to find homologous genes in other vertebrate and invertebrate species by performing BlastN and TBlastN searches in different genomic databases. In Supporting Information Table S7, we list all the accession numbers of the sequences used for phylogenetic tree construction. In Figure 2, we present a rooted neighbour-joining (NJ) tree obtained by aligning multiple protein sequences of members of the Neuroligin protein family from different species. The topology of the tree is coherent with the known relationship between the different taxa and is supported by robust bootstrap values in almost all the nodes. All vertebrate Neuroligins are divided in four groups that correspond to the four different types of Neuroligins. The putative invertebrate orthologs, used as outgroup, represent the root of the tree. Subsequently, we used the MultiContigView tool of ENSEMBL Genome Browser (Hubbard et al., 2005) in order to analyse the neighbouring genomic regions of each zebrafish Neuroligin with the mapped human or mouse genome (Supporting Information Fig. S1). We found conserved genes in genomic flanking regions around each zebrafish homolog. Notably, in the case of Neuroligins 4a and 4b, we observed syntenic genes only in flanking regions of human Neuroligin 4X but not in Neuroligin 4Y, confirming that *NLGN4Y* is specific to primates.

Taken together, phylogenetic and syntenic analyses confirm that all the zebrafish Neuroligin genes derived from a whole genome duplication event occurred at the base of the teleost radiation at approximately 350 Mya (Postlethwait, 2007). Both paralogous genes survived during evolution with the exception of one duplicate of Neuroligin 1, which has been lost before the first speciation events within the teleosts.

Gene Structure of the Zebrafish Neuroligins

We determined the exon–intron structure of zebrafish Neuroligin genes by mapping each CDS to the genomic sequence available in the ENSEMBL genomic database and then we compared them to the human genes. Supporting Information Tables S2–S5 present the results of this analysis. Globally, all the splicing borders present canonical GT and AG donor and acceptor sequences and, with specific exceptions discussed below, all the exons display the same protein reading frame, are identical or very similar in size, and encode the same protein region.

Among all the zebrafish Neuroligins, *nlgn1, nlgn3b*, and *nlgn4a* are the most similar to human genes. Indeed, all the exon sizes are identical to the human counterpart, with the exception of the first and the last exons, which present different sizes in all the zebrafish Neuroligins. In particular, *nlgn1* presents a high level of sequence conservation in exonic borders and intronic flanking regions of exons 2, 3, and 5, which are alternatively spliced at sites A1, A2, and B, respectively (Supporting Information Tables S2, S6). The exonic and intronic flanking regions of alternatively spliced exons can present numerous cis-regulatory elements that serve as either splicing enhancers or silencers (Ladd and Cooper, 2002; Black, 2003; Chen and Manley, 2009). As shown in Supporting Information Table S6, intronic flanking regions of alternative splicing sites

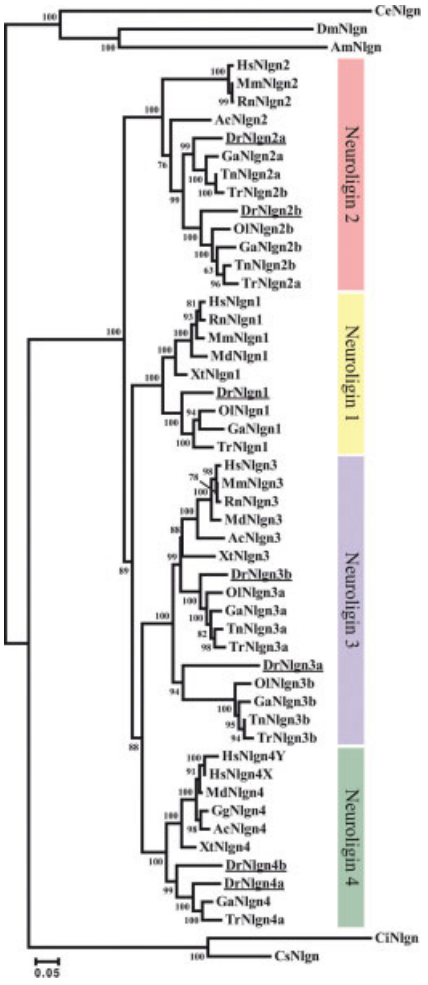

Fig. 2. Phylogenetic relationship of Neuroligin genes. Evolutionary comparison of different members of Neuroligin protein family represented in a phylogenetic rooted tree generated using MEGA 4.1 program with a Neighbor-Joining (NJ) method and 1,000 bootstrap replicates. The zebrafish Neuroligins are underlined. Branch lengths are measured in terms of amino acid substitutions, with the scale indicated below the tree. Numbers at nodes indicate percent of bootstrap probabilities. Ac, *Anolis carolinensis*; Am, *Apis mellifera*; Ce, *Caenorhabditis elegans*; Ci, *Ciona intestinalis*; Cs, *Ciona savignyi*; Dm, *Drosophila melanogaster*; Dr, *Danio rerio*; Ga, *Gasterosteus aculeatus*; Gg, *Gallus gallus*; Hs, *Homo sapiens*; Md, *Monodelphis domestica*; Mm, *Mus musculus*; Ol, *Oryzias latipes*; Rn, *Rattus norvegicus*; Tn, *Tetraodon nigroviridis*; Tr, *Takifugu rubripes*; Xl, *Xenopus laevis*; Xt, *Xenopus tropicalis*.

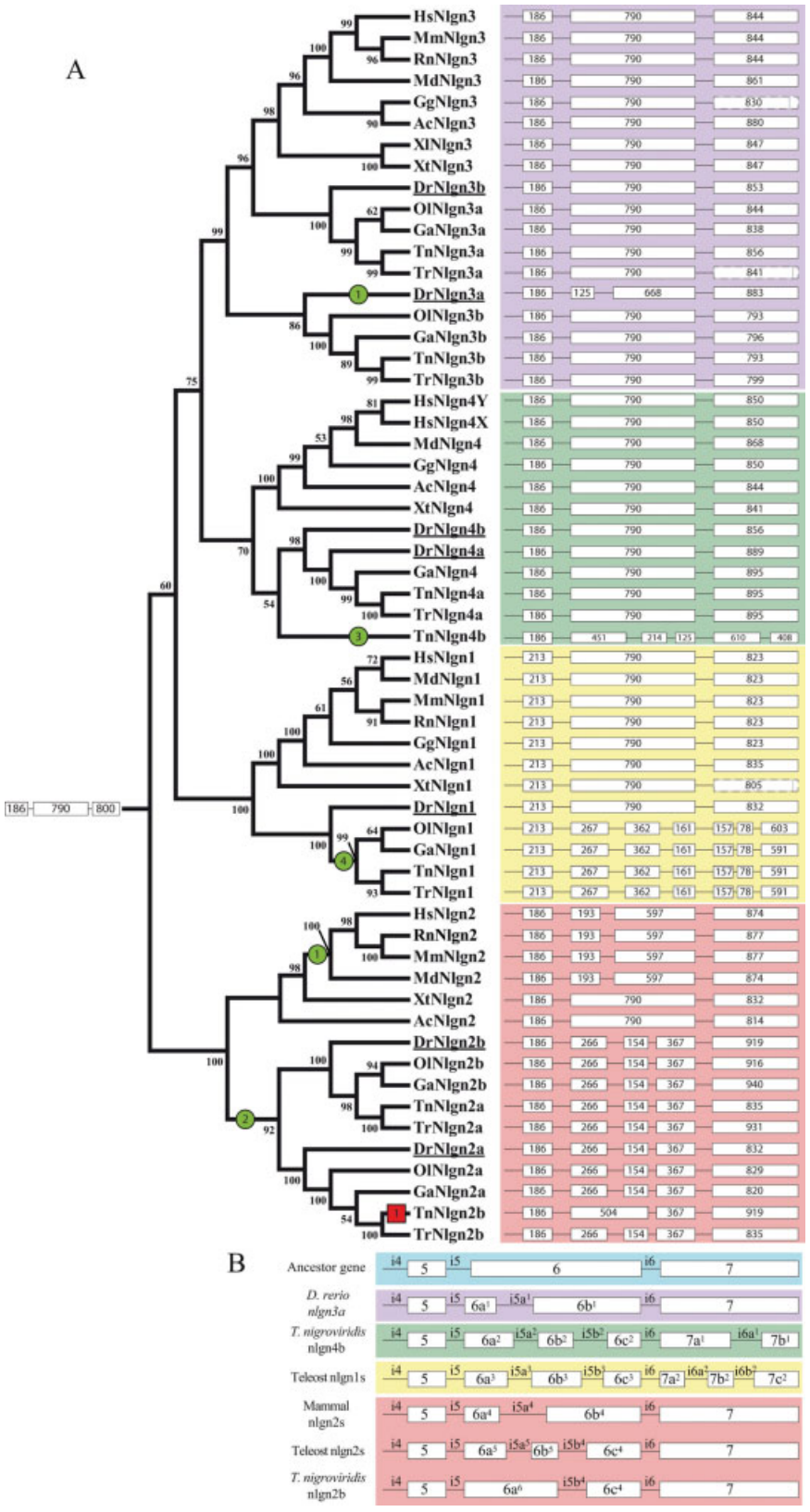

Fig. 3.

of Neuroligin 1 are slightly more conserved with respect to the other zebrafish Neuroligins, suggesting a possible conservation of regulatory elements involved in alternative splicing regulation.

Overall, these data suggest the existence of a strong selective pressure to preserve the original gene function due to the early loss of one copy after gene duplication.

Classical models indicate that the most common fate of a gene duplication event is nonfunctionalization.

Because of functional redundancy, most new gene copies tend to accumulate deleterious mutations and degenerate into pseudogenes before being, eventually, lost (Ohno, 1970, 1973; Holland et al., 1994; Sidow, 1996; Petrov and Hartl, 2000; Postlethwait, 2007). Alternative fates of duplicated genes are neofunctionalization and subfunctionalization. In neofunctionalization, one or both paralogous genes acquire new functional roles; in subfunctionalization, the genes divide their original functions (Postlethwait et al., 2004; Taylor and Raes, 2004; Postlethwait, 2007). In addition to gene duplication, alternative splicing is another major source of protein function diversity. When an alternatively spliced gene is duplicated, each copy can lose some alternative splicing isoforms due to the functional redundancy, or it can acquire new isoforms (Su et al., 2006). Therefore, alternative splicing can contribute also to the neo- or subfunctionalization of one or both newly duplicated genes.

nlgn3a and *nlgn4b*, which lost alternative splicing sites A1 and A2, respectively (see Supporting Information Tables S4, S5), can represent typical cases of subfunctionalization events. Their paralogous genes (*nlgn3b* and *nlgn4a*) present the exons lost by *nlgn3a* and *nlgn4b* and they produce the corresponding specific isoforms, as demonstrated by RT-PCR analyses (see Fig. 5A and B for further details).

Both *nlgn2* genes (*a* and *b*) are more divergent from the human gene. Exons 1–4 present similar features but, from exon 5 to exon 7, the genomic organization varies (see Supporting Information Table S3). In particular, exons 5–7 of both zebrafish *nlgn2*s seem to correspond to exons 5–6 of the human gene. Notably, the amino acid sequence of this terminal region of the protein is almost conserved. A similar case is represented by *nlgn3a* (see Supporting Information Table S4), where exons 6–7 correspond to a unique exon in human Neuroligin 3. The presence of a unique exon in *nlgn3b* suggests that in *nlgn3a* the exonic fragmentation is the result of a specific phenomenon of evolutionary divergence, which occurred after the complete duplication of zebrafish genome.

Analysis of Intron Loss/Gain Events During Evolution

In the last years, numerous and independent studies indicated that, in eukaryotes, intron evolution is a dynamic process and introns are gained and lost in different genomes in response to strong selective pressures (Jeffares et al., 2006). Therefore, at least a fraction of introns seems to be important for genome adaptation. Since the reconstruction of intron gain/loss events may provide valuable information to clarify evolutionary relationships within gene families and may provide insights about their possible functional implications (Coulombe-Huntington and Majewski, 2007), we compared the region corresponding to exons 5–7 of human Neuroligin 1 in different species of vertebrates. Furthermore, we reconstructed the possible intron gain/loss events along the evolutionary history of the Neuroligin gene family (Fig. 3A–B).

In Figure 3A we present a phylogenomic reconstruction of the abovementioned exonic region, and in Figure 3B we show a graphic representation of the same region limited to the genes subjected to intron gain/loss events with a classification of exons and introns. In our phylogenomic analysis, we postulated that the presence of a single exon 6 (of 790 bases) could represent the original condition of the hypothetical ancestral gene. This working hypothesis is essentially based on the following observations: (1) in most cases, this region of the CDS is encoded by a unique exon, (2) in almost all the species, the sum of the fragmented exons corresponds to the size of the unique exon form, and (3) the exon obtained by assembling all the fragmented exons presents similar borders and, if compared to the unique form of exon 6, it encodes the same protein region. As shown in Figure 3, using the mapped presence/absence of introns on the phylogenetic tree, we calculated the presence of 11 intron gains (green circles) and a unique intron loss event (red square). Intriguingly, Neuroligin 1 presents four intron gains only in teleosts (see Fig. 3A and B), with the only exception of *D. rerio*. This means that these events followed the zebrafish speciation and, consequently, since zebrafish diverged about 314–332 Myr ago from the other teleost species (Kasahara et al., 2007), they are relatively old. Moreover, they were completely fixed during evolution. Although it is not possible to exclude eventual intron loss events in zebrafish, our hypothesis remains the most parsimonious.

The Neuroligins 2 present by far the most fragmented exon structure. In our analysis, only *X. tropicalis* and *A. carolinensis* present a unique exon 6 as the supposed ancestor. In Neuroligins 2, we identified two intron gains (introns $i5a^5$ and $i5b^4$ in Fig. 3B) before the teleost radiation, followed by a specific intron loss in *T. nigroviridis* (resulting in the formation of exon $6a^6$) and, lastly, an independent intron gain event probably at the base of mammals ($i5a^4$). It is important to note that although exons $6a^3$ and $6a^5$ of teleost nlgn1s and nlgn2s, respectively, present very similar sizes (267 vs. 266 bp), introns $i5a^3$ and $i5a^5$

Fig. 3. Evolutionary history of the Neuroligin gene structure. **A**: Phylogenetic analysis (left) and structural comparison (right) of the evolutionary history of the last exons at the 3′ end of each Neuroligin gene in different species. A NJ phylogenetic tree with 1,000 bootstrap replicates has been used to reconstruct the evolution of vertebrate Neuroligins. The branch length is not proportional to genetic distance. Numbers at nodes indicate bootstrap values. Potential intron loss events are marked by a red square, putative intron gains are marked by a green circle. White boxes indicate exons, black lines indicate line introns. The numbers inside each exon indicate its size in base pairs, while a dashed line highlights when an incomplete sequence of the exon has been used. Exons and introns are not depicted in scale. Ac, *A. carolinensis*; Dr, *D. rerio*; Ga, *G. aculeatus*; Gg, *G. gallus*; Hs, *H. sapiens*; Md, *M. domestica*; Mm, *M. musculus*; Ol, *O. latipes*; Rn, *R. norvegicus*; Tn, *T. nigroviridis*; Tr, *T. rubripes*; Xl, *X. laevis*; Xt, *X. tropicalis*. **B**: Graphic representation of the exonic structure of genes subjected to intron gain/loss events. Exons 5–7 of different Neuroligins involved in intron gain/loss events are represented by white boxes, the introns by black lines. Starting from the hypothetical structure of the ancestor gene, we marked with different names the non-orthologous introns and exons. Exons and introns are not depicted to scale.

cannot be considered orthologous introns subjected to intron sliding events because they present different phases (Garcia-Espana et al., 2009). Therefore, our data indicate that, during evolution, two different introns were independently inserted in almost the same exonic position, suggesting that this region could represent a hotspot of intron gain.

Finally, the gene structures of Neuroligin 3 and 4 appear more conserved during evolution. Only limited and independent events occurred in two different fish species. As previously noted, an intron gain occurred only in zebrafish *nlgn3a* (the insertion of intron $i5a^1$ with the formation of exons $6a^1$ and $6b^1$) and three different events in *T. nigroviridis* Neuroligin 4b.

Taken together, these results suggest that major branching events in the vertebrate tree (like the origin of mammals and teleosts) seem to be correlated with intron gain events. Furthermore, our results are very consistent with the following independent results: (1) genome-wide informatics studies suggest that intron gain is more prevalent than intron loss (Babenko et al., 2004; Kumar and Hedges, 2005; Roy and Gilbert, 2005a,b), (2) the fish lineage has gained many introns after it diverged from the ancestor of the mammalian lineage (Venkatesh et al., 1999), and (3) intron gain is extremely rare in mammals (Roy et al., 2003).

Although in almost all the highlighted cases for Neuroligin genes, intron gain/loss events do not induce variations in the amino acid sequence and protein structure, they represent strong selective events. Thus, a possible explanation is that intron gain/loss events could represent evolutionary events, which can result in a different gene regulation of distinct members of the same gene family. Indeed, some introns are known to enhance or be necessary for normal levels of mRNA transcription, processing, and transport. Moreover, they can encode a variety of untranslated RNAs including microRNAs, small nucleolar RNAs, and guide RNAs for RNA editing (Jeffares et al., 2006). From this point of view, the presence of the vast majority of these events in the most ancient species analyzed (teleosts) suggests a possible correlation between intron gain/loss events and the evolutionary age of the species. It is interesting to note that, at least in vertebrates, all the Neuroligin intron gain/loss events are localized in the last exons of CDS, which encode part of the extracellular region and the intracellular portion of the proteins. A possible explanation is that this exonic region represents the majority of the CDS (~62%) and as a result it displays an increased probability of being subjected to these events. However, the possibility that this increased frequency is caused by the presence of specific nucleotide sequences favoring gain/loss events (as suggested by independent insertion of introns $i5a^3$ and $i5a^5$ in teleost nlgn1 and nlgn2, respectively) cannot be excluded and it requires further analyses.

Correlation of Intron Gain/Loss Events and Secondary Protein Structure

Experimental evidence indicates that a non-random tendency exists for introns to be located in interdomain regions of proteins (Patthy, 1999; Liu and Altman, 2003) and that introns have a propensity to avoid secondary structure elements such as alpha-helices and beta-strands (Contreras-Moreira et al., 2003). Therefore, we investigated the possible correlation between intron gain/loss events and the secondary structure of Neuroligin proteins.

In Figure 4, we aligned the amino acid sequences of the Neuroligins subjected to intron gain/loss events (see also Fig. 3B) and we compared their different exonic structures to the secondary structure elements of crystallized Neuroligins (Fabrichny et al., 2007; Koehnke et al., 2008a). In Neuroligins, most of the intron gain/loss events are localized outside or at the borders of secondary structures and only few events are found inside alpha-helices. This tendency could be explained by the purifying effects of natural selection, as a result of the chance of disrupting the alpha-helices. Mammalian Neuroligins 2 (represented in Fig. 4 by the mouse sequence) and teleost Neuroligin 1 (represented by the fugu sequence) present a unique intronic insertion (respectively, introns $i5a^4$ and $i5b^3$ in Fig. 3B) in an interdomain region. Moreover, all the teleost Neuroligins 2 display an intron ($i5b^4$) at the border of one of the two alpha-helices involved in Neuroligin dimerization (α14) and *T. nigroviridis* Neuroligin 4b presents an intron immediately before alpha-helix 17 (indicated as α17 in Fig. 4).

In a few cases, the intron insertions involve specific secondary structure elements. In *D. rerio nlgln3a*, the intron $i5a^1$, which divides exon 6 in exons $6a^1$ and $6b^1$ (Fig. 3B), was inserted inside the coding sequence for alpha-helix 9 (α9 in Fig. 4). Moreover, in teleost Neuroligins 1–2 two different intron insertions occurred ($i5a^3$ and $i5a^5$, respectively) between α11 and β11. However, in all these cases, the intron gain did not produce any sort of protein sequence alteration. As shown in Figure 4, the region between α11 and β11 (residues QGEFLN, highlighted with a violet box in Fig. 4) represents the main binding site for beta-Neurexin 1 (Fabrichny et al., 2007). Notably, the specific intron loss event in *T. nigroviridis* produced the insertion in *nlgn2b* CDS of a specific sequence absent in all the other Neuroligins (see Fig. 4); this nucleotide sequence encodes 27 amino acids that break the main binding region (Q-GEFLN, where dash indicates the insertion). Nevertheless, surprisingly, the amino acid insertion terminates with a glutamine (Q), thus reforming the correct pair of residues (QG) involved in Ca^{2+} coordination at the interface between Neuroligin and beta-Neurexin 1 (Fabrichny et al., 2007). The effect of this sequence insertion on the Neuroligin structure and, therefore, on the Neurexin–Neuroligin binding requires further analyses.

Analysis of Alternative Splicing of Zebrafish Neuroligins in Adult Organs and During Embryonic Development

Mammalian Neuroligins present two different alternative splice sites (sites A and B) inside the esterase-like domain. In particular, each human Neuroligin gene encodes different isoforms at site A (A1 and A2), while only Neuroligin 1 is alternatively spliced at site B (see Fig. 1).

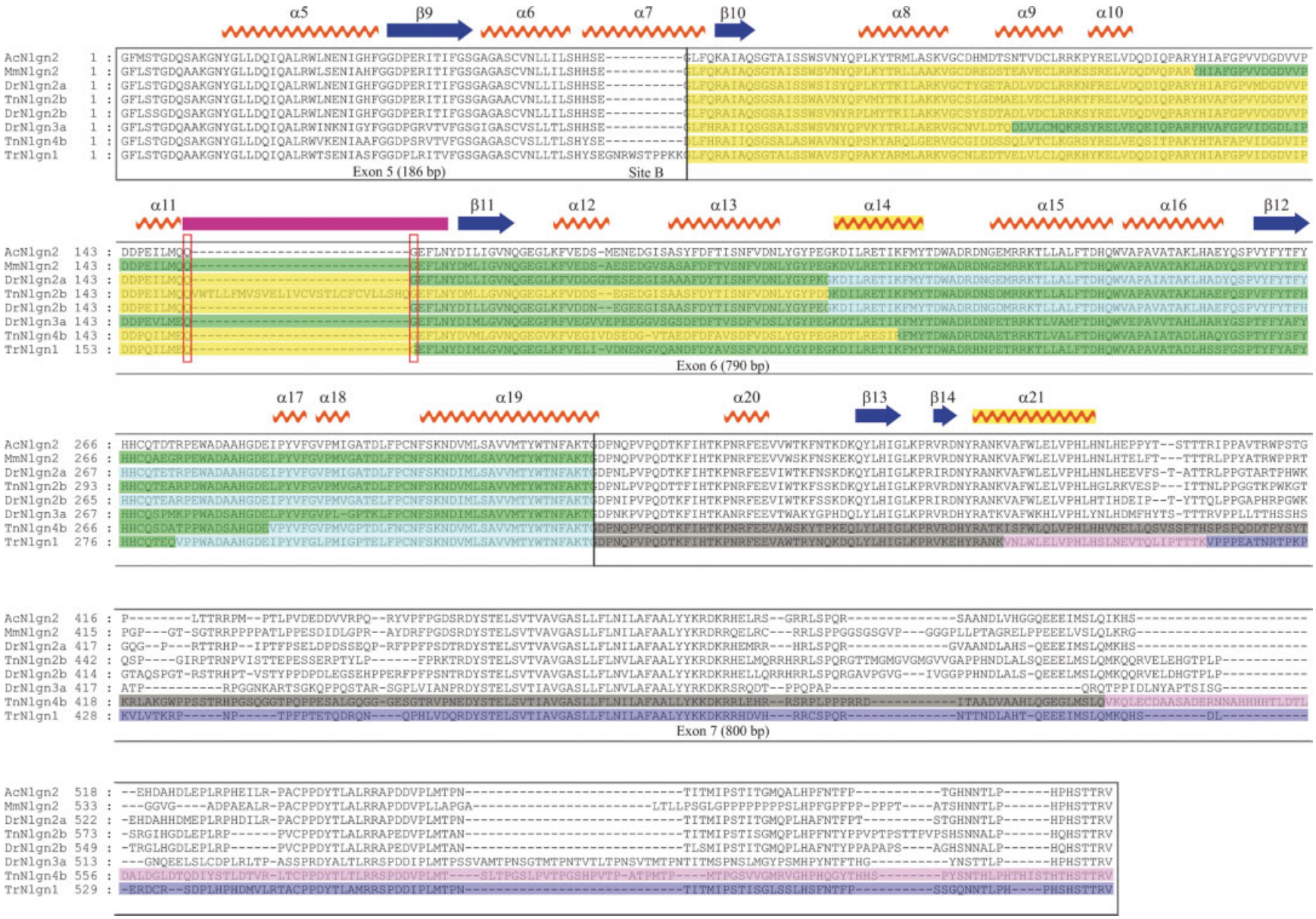

Fig. 4. Exonic and secondary structure comparison of the 3′ portion of Neuroligin genes. The amino acid sequences of the last exons of different Neuroligins were aligned using AlignX of Vector NTI Advance 10.1.1 based on the Clustal W algorithm. The exonic structure of the hypothetical ancestor gene is indicated with open boxes and exons are indicated below the sequences. Site B indicates the alternative splicing site B of *T. rubripes* Neuroligin 1 (TrNlgn1). For each sequence, the exonic structure is indicated with colored shades when it differs from that of the ancestor gene. Each color represents a separate exon. Secondary structure elements of Neuroligins are depicted above the alignment; beta-strands are indicated with blue arrows and alpha-helices in red. The two helices involved in dimerization are highlighted in yellow, the purple box indicates the main region of interaction with beta-Neurexins, and the pair of residues involved in Ca^{2+} binding is boxed in red. Modified from Koehnke et al. (2008) and Fabrichny et al. (2007). Ac, *A. carolinensis*; Dr, *D. rerio* (zebrafish); Ga, *G. aculeatus*; Mm, *M. musculus*; Tn, *T. nigroviridis*; Tr, *T. rubripes*.

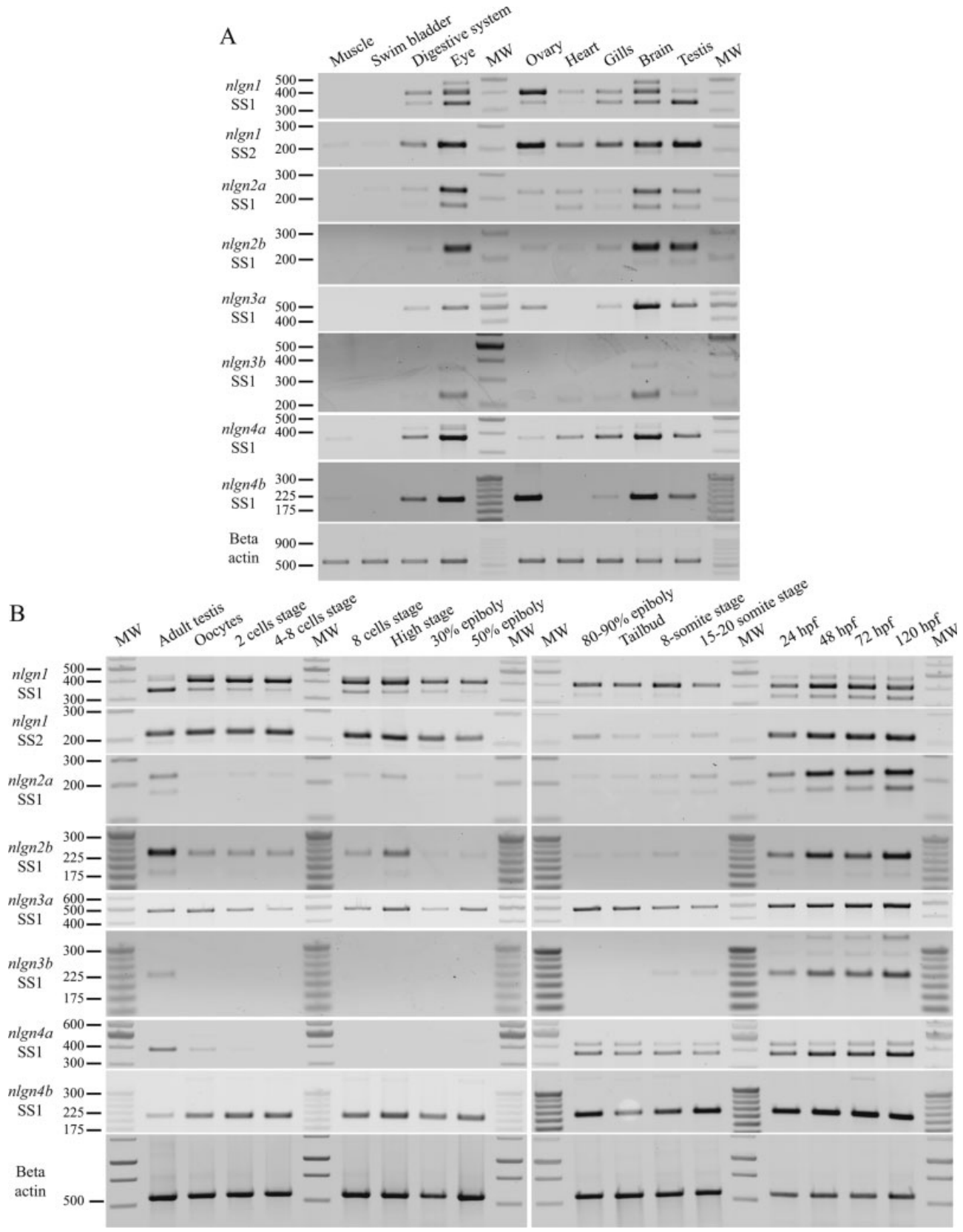

Fig. 5. Analysis of zebrafish Neuroligins splicing pattern during embryo development and in adult organs. RT-PCR products were generated from total RNA extracted from different adult organs, oocytes, and stages of embryonic development. Specific beta actin primers (Argenton et al., 2004) were used to check cDNA quality and possible genomic contamination. SS1 and SS2 indicate alternatively spliced sites of Neuroligin genes (Site A and Site B, respectively). Numbers on the left indicate the dimension in base pairs of the different molecular weights (MW). Control RT-PCR reactions without cDNA template were negative (data not shown).

Previous studies have already focused on the expression pattern of Neuroligin genes in different species (from mammals to birds), but few data concerning the alternative splicing pattern variation are available (Philibert et al., 2000; Bolliger et al., 2001; Chih et al., 2006; Suckow et al., 2008).

In order to analyse the Neuroligins expression and their alternative splicing pattern variation in zebrafish, we performed RT-PCR assays on cDNA obtained from different adult organs and developmental stages, using primers specific for zebrafish isoforms (Fig. 5A,B and Supporting Information Table S8). Zebrafish Neuroligin transcripts are detected in many adult organs with the exception of muscle and swim bladder where they are barely identifiable or completely absent (Fig. 5A). Notably, in some cases the alternative splicing pattern seems to be organ specific.

nlgn1 is strongly expressed in ovary, testis, brain, and eye samples. Ovary and testis have a specular pattern and, interestingly, only in brain and eye a third band corresponding to the isoform A1+A2 is detectable. At site B (indicated as SS2 in Fig. 5) only eye, brain, ovary, and testis present a band corresponding to the insert-minus forms, which is expressed at a low level.

nlgn2a and *nlgn2b* present a very similar expression pattern. Their expression is abundant in brain, eye, and testis where both forms (with or without exon A2) are present. The major difference between these two duplicated genes is that in *nlgn2a*, the short form seems to be more represented. Moreover, as shown in Supporting Information Table S6, both genes present a low percentage of sequence identity in intronic regions flanking the alternatively spliced exon at site A2. Together, these data indicate the presence of subfunctionalization events in the regulative mechanisms of alternative splicing of both genes.

The *nlgn3a* and *nlgn3b* genes represent another case of subfunctionalization event. While *nlgn3a* shows a unique band corresponding to insert-minus isoform, *nlgn3b* presents all the possible isoforms (compare eye and brain samples in Fig. 5A and 24–120-hpf stages in Fig. 5B).

With the exception of the swim bladder, the *nlgn4a* is expressed in all the organs tested. The splice site A of *nlgn4a* can produce two different isoforms (with or without insert A2), confirming that alternative splicing occurs at site A. Moreover, while the longer isoform is expressed only in specific organs, the isoform without the insert is shared by all the organs tested. *nlgn4b* insert-minus transcripts are present at low levels in digestive system, gills, and testis, while their expression is particularly enriched in eye, ovary, and brain. Overall, these data indicate that, in adult *Danio rerio*, Neuroligin genes are expressed and alternatively spliced in different organs inside and outside the Central Nervous System (CNS). Although their broad expression can indicate possible unknown functions, its real functional meaning requires further analyses.

Figure 5B shows the different isoforms expressed during zebrafish embryonic development. As the Neurexin genes (Rissone et al., 2007), many *nlgns* (*1*, *2a*, *2b*, *3a*, and *4b*) are expressed since the earliest stages of development and they increase their expression from 24 hr post fertilization (hpf).

As independently confirmed by an EST from ovary (Acc. Number: BI709920), all the *nlgn1* isoforms are maternally inherited. For *nlgn1* splice site A, the intermediate form (with inserts A1 or A2) is the more abundant throughout embryo development. The expression level of the other isoforms decreases from tailbud to 15–20 somite stage and afterwards it increases starting from 24 hpf. Concerning site B, both possible forms are present during development. Starting from late epiboly, the shorter form is no longer detected and, as observed in other vertebrate species (Sudhof, 2008), the longer form predominates (Fig. 5B).

The insert-plus isoform of *nlgn2a* is expressed, at very low levels, from the first stages of development. A lower band representing the insert-minus is detectable from the 8-somite stage and, in the last examined stages, both isoforms are more represented. *nlg2a* and *nlgn2b* present a similar expression during embryonic development and, as previously observed in adult organs, the alternative splicing pattern is different. In particular, the insert-minus isoform of *nlgn2b* is expressed at very low levels (Fig. 5B).

As previously mentioned, *nlgn3a* presents a unique isoform without insert, which is present at all developmental stages examined. Although the intensity of the band seems to vary at different stages, it is possible to observe a marked increase starting from 24 hpf. On the contrary, *nlgn3b* is not expressed until the 8-somite stage. Notably, starting from 24 hpf its expression increases and all the possible alternatively spliced isoforms are produced.

Finally, *nlgn4a* and *nlgn4b* display a complementary distribution. While *nlgn4a* is not expressed until the 50% epiboly stage (Fig. 5B), *nlgn4b* transcripts are maternally inherited and they are detectable up to 120 hpf. The complementary expression of duplicate *nlgn3s* and *nlgn4s* could be induced by further evolutionary phenomena targeting their promoter regions, which resulted in different temporal expressions.

The data presented so far indicate that, during evolution, duplicate Neuroligin genes underwent different evolutionary fates that differentiated their gene expression and alternative splicing regulation.

Analysis of Expression Patterns of Zebrafish Neuroligin Genes During Embryonic Development

We analyzed the spatial and temporal expression pattern of zebrafish Neuroligins during embryogenesis by whole mount in situ hybridization (WISH) from 48 hr post-fertilization (hpf) to 120 hpf. All the sense probes did not show any staining (data not shown). In general, Neuroligins are widely expressed throughout the brain in all the stages analyzed (Figs. 6–7).

At 48 hpf, *nlgn1* is mainly expressed in the mesencephalon and the rostral part of rhomboencephalon (Fig. 6A, B). Moreover, a discrete staining is weakly present also in the ventral diencephalon and telencephalon. Afterwards, a strong signal is visible throughout the whole brain and the retina (Fig. 6C,D) while the expression of *nlgn1* appears to fade at 120 hpf (Fig. 6E,F).

In all the analyzed stages, the expression pattern of *nlgn2a* is quite similar to that of *nlgn1* with the exception of the retina at 72 hpf (compare Fig. 6D and J). *nlgn2b* is widely detected at stage 48 hpf in the brain. Notably, compared to *nlgn2a*, the

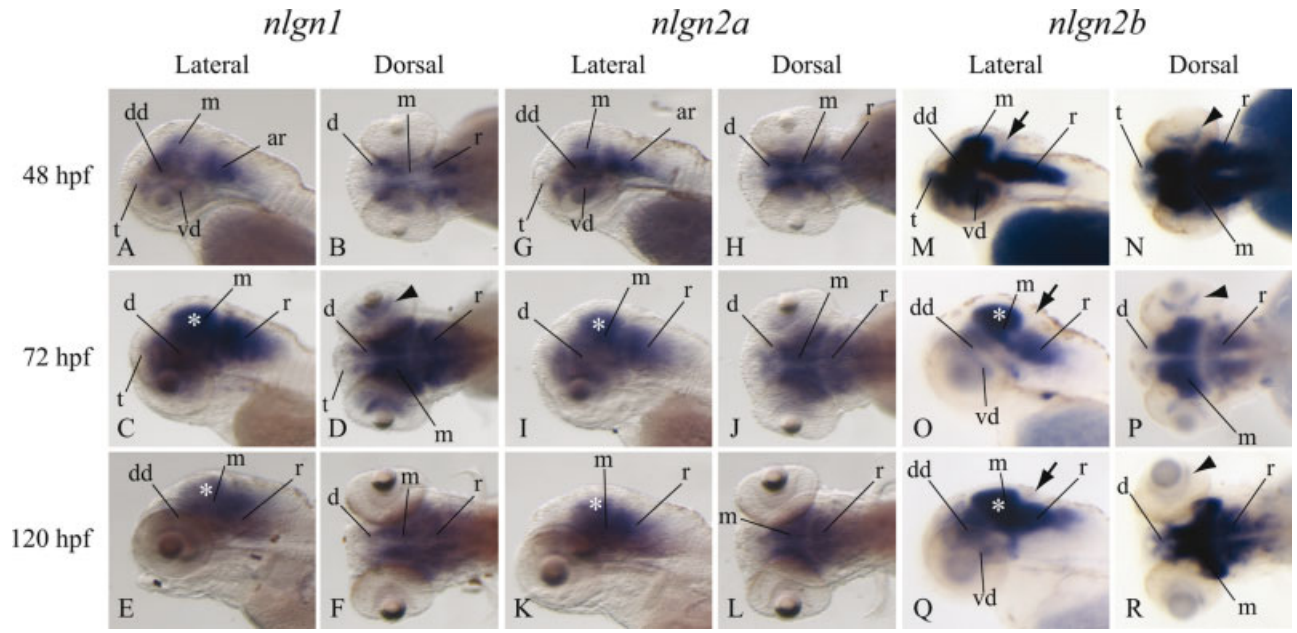

Fig. 6.

staining in the telencephalon is wider and, intriguingly, a positive signal is also present in the retina (black arrowheads in Fig. 6N, P, and R). Later during development (72 and 120 hpf), the mesencephalon and the rhomboencephalon are strongly stained, while only discrete regions within the diencephalon show a positive signal.

The two Neuroligin 3 paralogs display a rather distinct expression pattern (Fig. 7A–L). *nlgn3a* mRNA is detected at 48 hpf in discrete regions of the telencephalon, diencephalon, and rhomboencephalon. At later stages, positive staining is widespread in brain, with discrete regions visible in the telencephalon and diencephalon, while a faint signal is present also in the retina (black arrowhead in Fig. 7F). On the contrary, at 48 and 120 hpf *nlgn3b* presents a more restricted expression pattern (Fig. 7G,H and K,L), while at 72 hpf it is expressed throughout the brain (Fig. 7I,J). Notably, the retina is completely unstained in all the analyzed stages.

At 48 hpf, *nlgn4a* mRNA is detectable in discrete areas of the telencephalon, diencephalon, and rhomboencephalon (Fig. 7M,N). At 72 hpf, its expression expands also to mesencephalon, while at 120 hpf the ventral mesencephalon, the anterior rhomboencephalon, and the diencephalon are stained (Fig. 7O–R).

For *nlgn4b*, a strong signal is visible in all the analyzed stages. At 48 hpf,

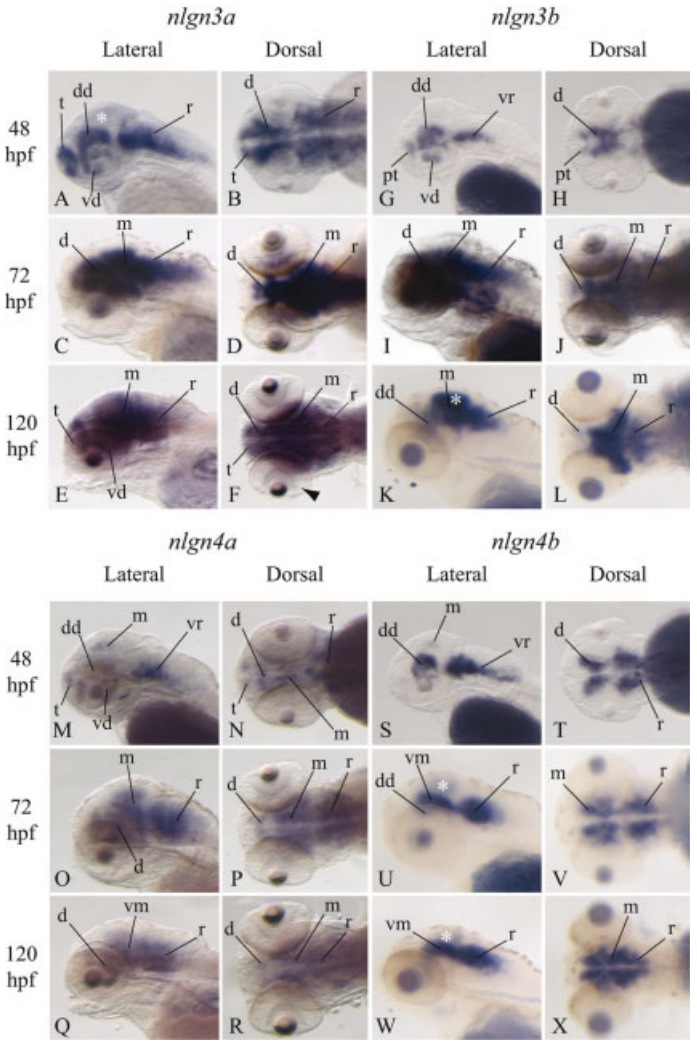

Fig. 7.

the staining is present in the diencephalon and rhomboencephalon (Fig. 7S,T); starting from 72 hpf, the mRNA is mainly detectable in the ventral mesencephalon and in the anterior rhomboencephalon (Fig. 7U–X). At 72 hpf, a faint signal is also visible in the dorsal diencephalon (Fig. 7U).

Taken together, these data indicate that, starting from 48 hpf, several Neuroligins are widely expressed in the brain of developing embryos. Although almost all Neuroligins shared a similar expression pattern in some regions of the brain as, for example, the midbrain and the hindbrain, in some cases we highlighted a differential expression between paralogous genes.

For example, *nlgn2b* and *nlgn3a* are expressed in the retina (from 48 hpf and at 120 hpf, respectively) and at 48 hpf *nlgn4b* is strongly expressed only in the dorsal diencephalon and ventral hindbrain. Finally, *nlgn3b* at 48 and 120 hpf presents a very restricted staining pattern.

Overall, RT-PCR and WISH analyses indicate that, during evolution, different subfunctionalization events differentiated the gene expression and the alternative splicing pattern of paralogous Neuroligins in zebrafish.

EXPERIMENTAL PROCEDURES

Zebrafish Maintenance

Zebrafish were raised and maintained under standard laboratory conditions as described in the Zebrafish Book (Westerfield, 2000) and staged according to Kimmel et al. (1995). Beginning from 24 hpf, embryos were cultured in fishwater containing 0.003% 1-phenyl-2-thiouera (PTU) to prevent pigmentation and 0.01% methylene blue to prevent fungal growth.

Databases and Bioinformatic Analysis of Data

The following genome assemblies were searched with the tools available from ENSEMBL: *Homo sapiens* (GRCh37, Feb. 2009), *Mus musculus* (NCBI m37, Apr. 2007), *Danio rerio* (Zv8, Dec. 2008), *Caenorhabditis elegans* (WS200, Jan. 2009), *Tetraodon nigroviridis* genome (TETRAODON 8.0, Mar. 2007), *Takifugu rubripes* (FUGU 4.0, June 2005), *Orizyas latipes* (HdrR, Oct. 2005), *Gasterosteus aculeatus* (BROAD S1, Feb. 2006), *Gallus. gallus* (WASHUC2, May 2006), *Anolis carolinensis* (AnoCar1.0, Feb. 2007), *Monodelphis domestica* (monDom5, Oct. 2006), *Rattus norvegicus* (RGSC 3.4, Dec. 2004), *Xenopus tropicalis* (JGI 4.1, Aug. 2005), *Ciona savignyi* (CSAV 2.0, Oct. 2005), and *Ciona intestinalis* genome (JGI version 2.0). Multi-alignment of human and zebrafish Neuroligins was performed with AlignX of VECTOR NTI Advance 10.1.1 (Invitrogen Corporation, Carlsbad, CA) and edited with GeneDoc version 2.7 (Nicholas, 1997). Structural features of amino acid sequences of human Neuroligins were predicted using release 20.52 (dated 28 July 2009) of PROSITE (http://www.expasy.org/prosite/).

Phylogenetic Analyses

Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 4.1, build number 4103 (Tamura et al., 2007). Amino acid sequences of Neuroligins from organisms representing different taxa were aligned in MEGA 4.1 using Clustal W algorithm with a Gonnet protein weight matrix. A rooted phylogenetic tree was built using a Neighbor-Joining method. Bootstrap analyses for 1,000 cycles were used to assess the strength of the topologies.

Cloning of Zebrafish Neurexins and Neuroligins

Public databases were searched to find genes and their genomic environments: Ensembl database http://www.ensembl.org/), NCBI (http://www.ncbi.nlm.nih.gov/). For 5′ RACE technique, we used two different kits: FirstChoice® RLMRACE Kit (Ambion, Austin, TX) and 5′ RACE System for Rapid Amplification of cDNA Ends (Invitrogen) following the manufacturer's instructions. The cDNA sequences obtained with PCR methods were compared with the genomic sequences to identify the splice sites.

Syntenic Analysis

To find evidence for the conservation of synteny, we compared genomic regions neighbouring the zebrafish Neuroligins to the genes neighbouring human and mouse Neuroligins. Putative orthologs for each zebrafish gene were located on the human and mouse map using the MultiContigView comparative tools in

Fig. 6. Temporal and spatial expression patterns of *nlgn1, nlgn2a*, and *nlgn2b*. Whole mount in situ hybridization assays were performed on embryos at 48, 72, and 120 hr post-fertilization (hpf). **A–F**: *nlgn1* is mainly expressed in midbrain, hindbrain, and also in discrete regions of the diencephalon. At 72 hpf, a strong signal is visible throughout the whole brain. **G–L**: The expression of *nlgn2a* mRNA is similar to *nlgn1*. **M–R**: *nlgn2b* is widely detected at stage 48 hpf in the brain (with the exception of cerebellum, black arrows), and, unlike *nlgn2a*, in the retina (black arrowheads). From 72 hpf, the signal is detectable mainly in midbrain (especially in the *tectum opticum*, white asterisks), hindbrain, and discrete regions of the forebrain. Embryos are mounted with anterior to the left and dorsal up. ar, anterior rhomboencephalon; d, diencephalon; dd, dorsal diencephalon; m, mesencephalon; r, rhomboencephalon; t, telencephalon; vd, ventral diencephalon.

Fig. 7. Analysis of *Neuroligin 3a, 3b, 4a*, and *4b* expression patterns. In situ hybridization staining in 48-, 72-, and 120-hr post-fertilization (hpf) zebrafish embryos. **A–F**: *nlgn3a* mRNA is detected at 48 hpf in discrete regions of the telencephalon, dorsal and ventral diencephalon and rhomboencephalon, but no signal is present in the *tectum opticum* (white asterisks). At later stages, a widespread staining is visible in the brain, with the exception of the telencephalon at 72 hpf. **G–L**: At 48 hpf, *nlgn3b* mRNA is detectable in the posterior telencephalon, dorsal and ventral diencephalon, and the ventral portion of the rhomboencephalon. At 72 hpf, a widespread staining is visible in the brain, except for the telencephalon. At 120 hpf, the expression is restricted to the dorsal part of the diencephalon, to the midbrain (especially in the *tectum opticum*, white asterisk) and the anterior hindbrain. **M–R**: *nlgn4a* mRNA is weakly detectable at 48 hpf in discrete regions of brain. From 72 hpf onwards, its expression covers midbrain and the rostral hindbrain, with a faint signal in diencephalon. **S–X**: At 48 hpf, *nlgn4b* mRNA is strongly detectable in the dorsal diencephalon and ventral rhomboencephalon. From 72 hpf, a strong signal is present in the ventral mesencephalon and anterior rhomboencephalon. A very weak signal is also visible in the *optic tectum* (white asterisks) in every stage analyzed. Embryos are mounted with anterior to the left and dorsal up. d, diencephalon; dd, dorsal diencephalon; m, mesencephalon; pt, posterior telencephalon; r, rhomboencephalon; t, telencephalon; vd, ventral diencephalon; vm, ventral mesencephalon; vr, ventral rhomboencephalon.

release 49 of ENSEMBL Genome Browser (Hubbard et al., 2005).

Reverse Transcription-Polymerase Chain Reaction (RT-PCR)

Total RNAs were prepared from different zebrafish adult organs (muscle, swim bladder, digestive system, eye, ovary, heart, gills, brain, and testis), oocytes, and embryos at different developmental stages using the Totally RNA Isolation Kit (Ambion) or the RNAgents Total RNA Isolation System (Promega, Madison, WI), treated with DNase I RNase-free (Roche) to avoid possible contamination from genomic DNA and then reverse transcribed using Superscript II (Invitrogen) and Oligo dT primers or the ImProm-II Reverse Transcription System (Promega) and Random primers. The cDNAs were then subjected to PCR amplification using specific primers (see Supporting Information Table S8) using Expand High Fidelity Taq polymerase (Roche) or GOTaq (Promega) following the manufacturer's instructions. When possible, all primer pairs have been designed on different exons to avoid the amplification of DNA contaminations eventually present in cDNA preparations. Control PCR experiments with samples prepared without reverse transcriptase were performed to ensure that genomic DNA contamination did not contribute to the PCR amplification (data not shown). The PCRs consisted of an initial denaturation of the samples at 95°C for 3 min, followed by 35 cycles. Each cycle consisted of a denaturation step at 95°C for 30s, a 30-s annealing step at the temperatures specified in Supporting Information Tables S8, and an extension step at 72°C for a time depending on fragment length. A final extension cycle of 10 min at 72°C was added to each PCR. Products were then separated on agarose gels at different concentration (from 1 to 3% maximum, based on the fragments length), visualised by ethidium bromide staining and then scanned with a Typhoon 8600 (Molecular Dynamics, Sunnyvale, CA). Nucleotide sequences of PCR products were determined by cloning in pCRII system (Invitrogen) or pGEM-T and pGEM-T Easy Vectors systems (Promega) followed by sequencing of both strands (PRIMM). A fragment of zebrafish β-actin cDNA was amplified by PCR (35 cycles) as an internal control for the quality of cDNA using a couple of primers that demonstrate the lack of genomic contamination in our RNA preparations (Argenton et al., 2004).

Accession Numbers

The accession numbers of zebrafish Neuroligin family members and all the other sequences used in phylogenetic analyses are listed in Supporting Information Table S7.

Gene-Specific Primer Sequences

The sequences of primers used in alternative splicing pattern analysis are listed in Supporting Information Table S8.

In Situ Hybridization

For Neuroligins probes preparation, gene-specific fragments were amplified by RT-PCR on suitable template. All the primers were designed to cover the 3′end of the coding sequence and part of the 3′UTR of the gene of interest. PCR products were cloned into the pGEM-T Easy Vector (Promega) or pCRII-TOPO (Invitrogen) and recombinant plasmids were sequenced on both strands. MAXIscript SP6/T7 Kits (Ambion) were used to synthesize antisense or sense RNA probes. All embryos used for whole-mount in situ (WISH) hybridization were fixed for 2 hr in 4% paraformaldehyde/phosphate buffered saline, rinsed with PBS-Tween, dehydrated in 100% methanol, and stored at −20°C until processed for WISH. WISH assays were carried out according to Thisse et al. (1993), modifying the protocol on purpose. Wild type embryos were hybridized by using digoxigenin-11-UTP (Boehringer, Roche) in vitro–labeled riboprobes at a concentration of 0.5–1 ng/μl. Hybridization was carried out at 68°C. Hybridized probes were then detected by using an anti-digoxigenin antibody conjugated to alkaline phosphatase (AP, Boehringer-Roche) at a 1:5,000 dilution and nitroblue tetrazolium/5-bromo-4-chloro-3-indolyl phosphate (NBT/BCIP; Promega) was used as the substrate for AP. Stained embryos were dehydrated and stored in methanol.

ACKNOWLEDGMENTS

We thank all of the members of Prof. Cotelli's lab for helpful advice and expertise. Special thanks go to Giulio Pavesi and David Horner for helpful discussions and to Erica Bresciani, Manuela Marai, Luca Del Giacco, and Roberto Marotta for a critical reading of the manuscript. This study was supported by the Associazione Italiana per la Ricerca sul Cancro (AIRC), Regione Piemonte (Ricerca Finalizzata 2007, 2008, 2009; Ricerca industriale e competitiva 2006, grant PRESTO; Ricerca Tecnologie convergenti 2007, grant PHOENICS; Piattaforme tecnologiche per le biotecnologie, grant Druidi); Fondazione CRT- Torino, and "Ministero della Salute" (Programma Ricerca Oncologica 2006, Ricerca Finalizzata 2006); Fondazione Cariplo (grant number 2006.0807) and Progetto Cariplo N.O.B.E.L. (biological and molecular characterization of cancer stem cells), and Regione Piemonte (Ricerca Scientifica Applicata 2004, A17).

REFERENCES

Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A, Gelpke MD, Roach J, Oh T, Ho IY, Wong M, Detter C, Verhoef F, Predki P, Tay A, Lucas S, Richardson P, Smith SF, Clark MS, Edwards YJ, Doggett N, Zharkikh A, Tavtigian SV, Pruss D, Barnstead M, Evans C, Baden H, Powell J, Glusman G, Rowen L, Hood L, Tan YH, Elgar G, Hawkins T, Venkatesh B, Rokhsar D, Brenner S. 2002. Whole-genome shotgun assembly and analysis of the genome of Fugu rubripes. Science 297: 1301–1310.

Argenton F, Giudici S, Deflorian G, Cimbro S, Cotelli F, Beltrame M. 2004. Ectopic expression and knockdown of a zebrafish sox21 reveal its role as a transcriptional repressor in early development. Mech Dev 121:131–142.

Babenko VN, Rogozin IB, Mekhedov SL, Koonin EV. 2004. Prevalence of intron gain over intron loss in the evolution of paralogous gene families. Nucleic Acids Res 32:3724–3733.

Barrow SL, Constable JR, Clark E, El-Sabeawy F, McAllister AK, Washbourne P. 2009. Neuroligin1: a cell adhesion molecule that recruits PSD-95 and NMDA receptors by distinct mechanisms during synaptogenesis. Neural Dev 4:17.

Biswas S, Russell RJ, Jackson CJ, Vidovic M, Ganeshina O, Oakeshott JG, Claudianos C. 2008. Bridging the synaptic gap: neuroligins and neurexin I in *Apis mellifera*. PLoS One 3:e3542.

Black DL. 2003. Mechanisms of alternative pre-messenger RNA splicing. Annu Rev Biochem 72:291–336.

Bolliger MF, Frei K, Winterhalter KH, Gloor SM. 2001. Identification of a novel neuroligin in humans which binds to PSD-95 and has a widespread expression. Biochem J 356:581–588.

Bolliger MF, Pei J, Maxeiner S, Boucard AA, Grishin NV, Sudhof TC. 2008. Unusually rapid evolution of Neuroligin-4 in mice. Proc Natl Acad Sci USA 105:6421–6426.

Boucard AA, Chubykin AA, Comoletti D, Taylor P, Sudhof TC. 2005. A splice code for trans-synaptic cell adhesion mediated by binding of neuroligin 1 to alpha- and beta-neurexins. Neuron 48: 229–236.

Chen M, Manley JL. 2009. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat Rev Mol Cell Biol 10:741–754.

Chih B, Afridi SK, Clark L, Scheiffele P. 2004. Disorder-associated mutations lead to functional inactivation of neuroligins. Hum Mol Genet 13:1471–1477.

Chih B, Gollan L, Scheiffele P. 2006. Alternative splicing controls selective trans-synaptic interactions of the neuroligin-neurexin complex. Neuron 51:171–178.

Chubykin AA, Liu X, Comoletti D, Tsigelny I, Taylor P, Sudhof TC. 2005. Dissection of synapse induction by neuroligins: effect of a neuroligin mutation associated with autism. J Biol Chem 280: 22365–22374.

Chubykin AA, Atasoy D, Etherton MR, Brose N, Kavalali ET, Gibson JR, Sudhof TC. 2007. Activity-dependent validation of excitatory versus inhibitory synapses by neuroligin-1 versus neuroligin-2. Neuron 54:919–931.

Comoletti D, De Jaco A, Jennings LL, Flynn RE, Gaietta G, Tsigelny I, Ellisman MH, Taylor P. 2004. The Arg451Cys-neuroligin-3 mutation associated with autism reveals a defect in protein processing. J Neurosci 24:4889–4893.

Comoletti D, Flynn RE, Boucard AA, Demeler B, Schirf V, Shi J, Jennings LL, Newlin HR, Sudhof TC, Taylor P. 2006. Gene selection, alternative splicing, and post-translational processing regulate neuroligin selectivity for beta-neurexins. Biochemistry 45:12816–12827.

Contreras-Moreira B, Jonsson PF, Bates PA. 2003. Structural context of exons in protein domains: implications for protein modelling and design. J Mol Biol 333:1045–1059.

Coulombe-Huntington J, Majewski J. 2007. Characterization of intron loss events in mammals. Genome Res 17: 23–32.

Dean C, Scholl FG, Choih J, DeMaria S, Berger J, Isacoff E, Scheiffele P. 2003. Neurexin mediates the assembly of presynaptic terminals. Nat Neurosci 6: 708–716.

Fabrichny IP, Leone P, Sulzenbacher G, Comoletti D, Miller MT, Taylor P, Bourne Y, Marchot P. 2007. Structural analysis of the synaptic protein neuroligin and its beta-neurexin complex: determinants for folding and cell adhesion. Neuron 56:979–991.

Garcia-Espana A, Mares R, Sun TT, Desalle R. 2009. Intron evolution: testing hypotheses of intron evolution using the phylogenomics of tetraspanins. PLoS One 4:e4680.

Gilbert MM, Auld VJ. 2005. Evolution of clams (cholinesterase-like adhesion molecules): structure and function during development. Front Biosci 10:2177–2192.

Gilbert M, Smith J, Roskams AJ, Auld VJ. 2001. Neuroligin 3 is a vertebrate gliotactin expressed in the olfactory ensheathing glia, a growth-promoting class of macroglia. Glia 34:151–164.

Graf ER, Zhang X, Jin SX, Linhoff MW, Craig AM. 2004. Neurexins induce differentiation of GABA and glutamate postsynaptic specializations via neuroligins. Cell 119:1013–1026.

Hata Y, Butz S, Sudhof TC. 1996. CASK: a novel dlg/PSD95 homolog with an N-terminal calmodulin-dependent protein kinase domain identified by interaction with neurexins. J Neurosci 16:2488–2494.

Holland PW, Garcia-Fernandez J, Williams NA, Sidow A. 1994. Gene duplications and the origins of vertebrate development. Dev Suppl:125–133.

Hubbard T, Andrews D, Caccamo M, Cameron G, Chen Y, Clamp M, Clarke L, Coates G, Cox T, Cunningham F, Curwen V, Cutts T, Down T, Durbin R, Fernandez-Suarez XM, Gilbert J, Hammond M, Herrero J, Hotz H, Howe K, Iyer V, Jekosch K, Kahari A, Kasprzyk A, Keefe D, Keenan S, Kokocinsci F, London D, Longden I, McVicker G, Melsopp C, Meidl P, Potter S, Proctor G, Rae M, Rios D, Schuster M, Searle S, Severin J, Slater G, Smedley D, Smith J, Spooner W, Stabenau A, Stalker J, Storey R, Trevanion S, Ureta-Vidal A, Vogel J, White S, Woodwark C, Birney E. 2005. Ensembl 2005. Nucleic Acids Res 33:D447–453.

Ichtchenko K, Hata Y, Nguyen T, Ullrich B, Missler M, Moomaw C, Sudhof TC. 1995. Neuroligin 1: a splice site-specific ligand for beta-neurexins. Cell 81: 435–443.

Ichtchenko K, Nguyen T, Sudhof TC. 1996. Structures, alternative splicing, and neurexin binding of multiple neuroligins. J Biol Chem 271:2676–2682.

Jamain S, Quach H, Betancur C, Rastam M, Colineaux C, Gillberg IC, Soderstrom H, Giros B, Leboyer M, Gillberg C, Bourgeron T. 2003. Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. Nat Genet 34:27–29.

Jeffares DC, Mourier T, Penny D. 2006. The biology of intron gain and loss. Trends Genet 22:16–22.

Kang HS, Lee CK, Kim JR, Yu SJ, Kang SG, Moon DH, Lee CH, Kim DK. 2004. Gene expression analysis of the pro-oestrous-stage rat uterus reveals neuroligin 2 as a novel steroid-regulated gene. Reprod Fertil Dev 16:763–772.

Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y, Jindo T, Kobayashi D, Shimada A, Toyoda A, Kuroki Y, Fujiyama A, Sasaki T, Shimizu A, Asakawa S, Shimizu N, Hashimoto S, Yang J, Lee Y, Matsushima K, Sugano S, Sakaizumi M, Narita T, Ohishi K, Haga S, Ohta F, Nomoto H, Nogata K, Morishita T, Endo T, Shin IT, Takeda H, Morishita S, Kohara Y. 2007. The medaka draft genome and insights into vertebrate genome evolution. Nature 447:714–719.

Kimmel CB, Ballard WW, Kimmel SR, Ullmann B, Schilling TF. 1995. Stages of embryonic development of the zebrafish. Dev Dyn 203:253–310.

Koehnke J, Jin X, Budreck EC, Posy S, Scheiffele P, Honig B, Shapiro L. 2008a. Crystal structure of the extracellular cholinesterase-like domain from neuroligin-2. Proc Natl Acad Sci USA 105: 1873–1878.

Koehnke J, Jin X, Trbovic N, Katsamba PS, Brasch J, Ahlsen G, Scheiffele P, Honig B, Palmer AG 3rd, Shapiro L. 2008b. Crystal structures of beta-neurexin 1 and beta-neurexin 2 ectodomains and dynamics of splice insertion sequence 4. Structure 16:410–421.

Kumar S, Hedges SB. 2005. Pushing back the expansion of introns in animal genomes. Cell 123:1182–1184.

Kwon JY, Hong M, Choi MS, Kang S, Duke K, Kim S, Lee S, Lee J. 2004. Ethanol-response genes and their regulation analyzed by a microarray and comparative genomic approach in the nematode *Caenorhabditis elegans*. Genomics 83:600–614.

Ladd AN, Cooper TA. 2002. Finding signals that regulate alternative splicing in the post-genomic era. Genome Biol 3: reviews0008.

Laumonnier F, Bonnet-Brilhault F, Gomot M, Blanc R, David A, Moizard MP, Raynaud M, Ronce N, Lemonnier E, Calvas P, Laudier B, Chelly J, Fryns JP, Ropers HH, Hamel BC, Andres C, Barthelemy C, Moraine C, Briault S. 2004. X-linked mental retardation and autism are associated with a mutation in the NLGN4 gene, a member of the neuroligin family. Am J Hum Genet 74:552– 557.

Levinson JN, Chery N, Huang K, Wong TP, Gerrow K, Kang R, Prange O, Wang YT, El-Husseini A. 2005. Neuroligins mediate excitatory and inhibitory synapse formation: involvement of PSD-95 and neurexin-1beta in neuroligin-induced synaptic specificity. J Biol Chem 280:17312–17319.

Liu S, Altman RB. 2003. Large scale study of protein domain distribution in the context of alternative splicing. Nucleic Acids Res 31:4828–4835.

Missler M, Fernandez-Chacon R, Sudhof TC. 1998. The making of neurexins. J Neurochem 71:1339–1347.

Missler M, Zhang W, Rohlmann A, Kattenstroth G, Hammer RE, Gottmann K, Sudhof TC. 2003. Alpha-neurexins couple Ca2+ channels to synaptic vesicle exocytosis. Nature 423:939–948.

Nemeth AH, Nolte D, Dunne E, Niemann S, Kostrzewa M, Peters U, Fraser E, Bochukova E, Butler R, Brown J, Cox RD, Levy ER, Ropers HH, Monaco AP, Muller U. 1999. Refined linkage disequilibrium and physical mapping of the gene locus for X-linked dystonia-parkinsonism (DYT3). Genomics 60: 320–329.

Nicholas, KB, Nicholas, HB Jr. 1997. GeneDoc: a tool for editing and annotating multiple sequence alignents.

Ohno S. 1970. Evolution by gene duplication. New York: Springer-Verlag.

Ohno S. 1973. Ancient linkage groups and frozen accidents. Nature 244:259–262.

Paraoanu LE, Becker-Roeck M, Christ E, Layer PG. 2006. Expression patterns of neurexin-1 and neuroligins in brain and retina of the chick embryo: Neuroligin-3 is absent in retina. Neurosci Lett 395:114–117.

Patthy L. 1999. Genome evolution and the evolution of exon-shuffling: a review. Gene 238:103–114.

Petrov DA, Hartl DL. 2000. Pseudogene evolution and natural selection for a compact genome. J Hered 91:221–227.

Philibert RA, Winfield SL, Sandhu HK, Martin BM, Ginns EI. 2000. The structure and expression of the human neuroligin-3 gene. Gene 246:303–310.

Postlethwait JH. 2007. The zebrafish genome in context: ohnologs gone missing. J Exp Zool B Mol Dev Evol 308: 563–577.

Postlethwait J, Amores A, Cresko W, Singer A, Yan YL. 2004. Subfunction partitioning, the teleost radiation and the annotation of the human genome. Trends Genet 20:481–490.

Prange O, Wong TP, Gerrow K, Wang YT, El-Husseini A. 2004. A balance between excitatory and inhibitory synapses is controlled by PSD-95 and neuroligin. Proc Natl Acad Sci USA 101:13915–13920.

Rissone A, Monopoli M, Beltrame M, Bussolino F, Cotelli F, Arese M. 2007. Comparative genome analysis of the neurexin gene family in *Danio rerio*: insights into their functions and evolution. Mol Biol Evol 24:236–252.

Roy SW, Gilbert W. 2005a. The pattern of intron loss. Proc Natl Acad Sci USA 102:713–718.

Roy SW, Gilbert W. 2005b. Rates of intron loss and gain: implications for early eukaryotic evolution. Proc Natl Acad Sci USA 102:5773–5778.

Roy SW, Fedorov A, Gilbert W. 2003. Large-scale comparison of intron positions in mammalian genes shows intron loss but no gain. Proc Natl Acad Sci USA 100:7158–7162.

Sara Y, Biederer T, Atasoy D, Chubykin A, Mozhayeva MG, Sudhof TC, Kavalali ET. 2005. Selective capability of SynCAM and neuroligin for functional synapse assembly. J Neurosci 25:260–270.

Scheiffele P, Fan J, Choih J, Fetter R, Serafini T. 2000. Neuroligin expressed in nonneuronal cells triggers presynaptic development in contacting axons. Cell 101:657–669.

Shen KC, Kuczynska DA, Wu IJ, Murray BH, Sheckler LR, Rudenko G. 2008. Regulation of neurexin 1beta tertiary structure and ligand binding through alternative splicing. Structure 16: 422–431.

Sidow A. 1996. Gen(om)e duplications in the evolution of early vertebrates. Curr Opin Genet Dev 6:715–722.

Song JY, Ichtchenko K, Sudhof TC, Brose N. 1999. Neuroligin 1 is a postsynaptic cell-adhesion molecule of excitatory synapses. Proc Natl Acad Sci USA 96: 1100–1105.

Su Z, Wang J, Yu J, Huang X, Gu X. 2006. Evolution of alternative splicing after gene duplication. Genome Res 16: 182–189.

Suckow AT, Comoletti D, Waldrop MA, Mosedale M, Egodage S, Taylor P, Chessler SD. 2008. Expression of neurexin, neuroligin, and their cytoplasmic binding partners in the pancreatic beta-cells and the involvement of neuroligin in insulin secretion. Endocrinology 149: 6006–6017.

Sudhof TC. 2008. Neuroligins and neurexins link synaptic function to cognitive disease. Nature 455:903–911.

Talebizadeh Z, Lam DY, Theodoro MF, Bittel DC, Lushington GH, Butler MG. 2006. Novel splice isoforms for NLGN3 and NLGN4 with possible implications in autism. J Med Genet 43:e21.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599.

Taylor JS, Raes J. 2004. Duplication and divergence: the evolution of new genes and old ideas. Annu Rev Genet 38: 615–643.

Thisse C, Thisse B, Schilling TF, Postlethwait JH. 1993. Structure of the zebrafish snail1 gene and its expression in wild-type, spadetail and no tail mutant embryos. Development 119:1203–1215.

Varoqueaux F, Aramuni G, Rawson RL, Mohrmann R, Missler M, Gottmann K, Zhang W, Sudhof TC, Brose N. 2006. Neuroligins determine synapse maturation and function. Neuron 51:741–754.

Venkatesh B, Ning Y, Brenner S. 1999. Late changes in spliceosomal introns define clades in vertebrate evolution. Proc Natl Acad Sci USA 96:10267–10271.

Westerfield M. 2000. The zebrafish book. A guide for the laboratory use of zebrafish (*Danio rerio*). Eugene: University of Oregon Press.