Evolutionary and Expression Analyses Show Co-option of *khdrbs* Genes for Origin of Vertebrate Brain

Su Wang[1,2], Qingyun Yang[1,2], Ziyue Wang[1,2], Shuoqi Feng[1,2], Hongyan Li[1,2], Dongrui Ji[1,2]* and Shicui Zhang[1,2]*

[1] *Laboratory for Evolution and Development, Institute of Evolution and Marine Biodiversity, Ocean University of China, Qingdao, China,* [2] *Laboratory for Evolution and Development, Department of Marine Biology, Ocean University of China, Qingdao, China*

Genes generated by whole genome duplications (WGD) can be co-opted by changing their regulation process or altering their coding proteins, which has been shown contributable to the emergence of vertebrate morphological novelties such as vertebrate cartilage. Mouse *khdrbs* genes, differing from its invertebrate orthologs, were mainly expressed in brain, hinting that *khdrbs* gene family as a member of genetic toolkit may be linked to vertebrate brain development. However, the evolutionary relationship between *khdrbs* gene family and vertebrate brain development is unclear. First, we analyzed the evolutionary history of *khdrbs* gene family in metazoans, and then investigated their expression patterns during early development and in adulthood of zebrafish. We found that the duplication of *khdrbs* gene family by WGD took place in zebrafish, and all zebrafish *khdrbs* genes were predominantly expressed in the substructures of brain during early development. Given the expression of invertebrate *khdrbs* gene in germ line, the distinct expression domains of zebrafish *khdrbs* genes in brain suggested that the duplicated *khdrbs* genes are co-opted for promoting the evolutionary origin of vertebrate brain.

Keywords: khdrbs, co-option, evolution, brain development, zebrafish

INTRODUCTION

One of the key questions of evolution is to answer how the morphological complexity takes place among extant animals including vertebrates. The subphylum Vertebrata is characterized by the appearance of critical morphological innovations such as an elaborate segmented brain, neural crest cells, neurogenic placodes, and endoskeleton (Shimeld and Holland, 2000). Several decades of developmental genetic studies have led to the important discovery that most animals from different taxa share a bunch of regulatory and tissue-specific genes known as "genetic toolkit," which controls animal body pattern (Carroll, 2000). Therefore, the molecular phylogeny of these toolkit genes will provide us an insight into the evolutionary origin of morphological novelties. Whole genome duplication (WGD) could not only enrich the genomic complexity, but also generate new functions

of duplicated genes by co-option (Ohno, 1970; True and Carroll, 2002; Dehal and Boore, 2005; Conant and Wolfe, 2008; Cañestro et al., 2013). WGD occurred around the origin of vertebrate lineage, has been shown to be related to vertebrate morphological innovations (Manzanares et al., 2000; True and Carroll, 2002). For example, the origin of vertebrate cartilage was found to be attributable to the duplication of chordate fibrillar *collagen* genes (Wada et al., 2006; Zhang et al., 2006), and the emergence of vertebrate vascular vessels to the duplication of *kank* genes (Hensley et al., 2016).

The family of KHDRBS protein (KH Domain-containing, RNA Binding, and Signal transduction associated protein), is characterized by the GSG (GRP33/SAM68/GLD-1) domain, also named STAR domain, which includes a single KH domain flanked by the conserved N-terminal QUA1 and C-terminal QUA2 (Di Fruscio et al., 1999; Lukong and Richard, 2003). KHDRBS proteins bind RNA through their KH domains, and participate in signal transduction (Frisone et al., 2015; Ehrmann et al., 2016). In nematodes, GLD-1, the ortholog of KHDRBS, was found to be localized in the germ cell cytoplasm, and indispensable for oogenesis and meiotic prophase progression, while it shows little roles in the male germ line or soma, although it can stimulate sex determination of males in the hermaphrodite germ line (Jones and Schedl, 1995; Lee and Schedl, 2001, 2010). In contrast, NSR, the ortholog of KHDRBS of fruit fly, was found to be predominantly localized in the nuclei of primary spermatocytes, and necessary for male reproduction by regulating some male fertility genes (Ding et al., 2010). In vertebrates, there are three members of KHDRBS, KHDRBS1 (also called Sam68), KHDRBS2 (also called SLM1), and KHDRBS3 (also called SLM2) which share many features such as RNA binding and signal transduction. Interestingly, *khdrbs1* knockout mice showed impaired fertility in males as a result of mRNA translational regulation defect during spermiogenesis (Paronetto et al., 2009; Ehrmann and Elliott, 2010; Frisone et al., 2015), and KHDRBS3 was also involved in spermatogenesis via directly binding to the genes essential for male gametogenesis (Zhang et al., 2009). This function of KHDRBS1 and KHDRBS3 in mice is apparently analogous to their invertebrate orthologs GLD-1 and NSR in term of regulation of gametogenesis. On the contrary, the three members of vertebrate *khdrbs* family appear to be expressed more specifically in the brain. In mouse, *khdrbs1* and *khdrbs3* were predominantly expressed in the brain and testis, but *khdrbs2* was exclusively expressed in the brain (Ehrmann et al., 2016). Moreover, although *khdrbs1* KO male mice showed male infertility, behavioral deficits and poor motor control (Ehrmann et al., 2013), *khdrbs2* KO mice as well as *khdrbs1* and *khdrbs2* double KO mice both exhibited defects in cerebellar morphogenesis (Iijima et al., 2014), and *khdrbs3* KO mice displayed synaptic plasticity and behavioral defects (Traunmüller et al., 2016). These data together suggest that *khdrbs* gene family, as a member of genetic toolkit, may be linked to vertebrate brain development. However, the evolutionary relationship between *khdrbs* gene family and vertebrate brain development is still a mystery.

The aim of this study is thus to answer this question by taking advantage of the zebrafish (*Danio rerio*) model, which has a brain resembling that of humans in both basic structures and functional capacities (Tropepe and Sive, 2003). We first analyzed the molecular evolution of *khdrbs* gene family in representative metazoan taxa, and then examined the expression patterns of *khdrbs* during early development and in adulthood of zebrafish. We found that *khdrbs* gene family was expanded by WGD in zebrafish, and all zebrafish *khdrbs* genes were predominantly expressed in the substructures of brain during early development. Given that these substructures are vertebrate-specific trait, the distinct expression domains of *khdrbs* genes in zebrafish suggested that *khdrbs* gene family was co-opted for vertebrate brain development after WGD events around the split of Vertebrata.

MATERIALS AND METHODS

Zebrafish Strain and Embryos

The AB strain zebrafish were cultured at $28 \pm 1°C$. Embryos were cultured in E3 medium consisting of 5 mM NaCl, 0.17 mM KCl, 0.33 mM $CaCl_2$, and 0.33 mM $MgSO_4$. For whole-mount *in situ* hybridization (WISH), 0.0045% 1-phenyl-2-thiourea was added into E3 medium to prevent embryos from pigmentation started 24 h post-fertilization (hpf). Different stages of embryos were sorted and fixed following the guide of Kimmel et al. (1995).

Sequence Retrieval and Bioinformatics Analysis

KHDRBS protein sequences of most major metazoan taxa were obtained from NCBI and Ensembl database. Other KHDRBS protein sequences were identified by a BLASTp search with a query sequence (human KHDRBS1 protein). The sequences used were listed in Supplementary Table 1. Sequence alignment was performed by ClustalW method in MegAlign v7.1.0. To acquire the best evolutionary model for phylogenetic inference, we used default parameters in MEGA7 to conduct a best-fit protein model test. According to calculated BIC (Bayesian information criterion) scores for every model, we chose LG + G model which had the lowest score for further phylogenetic analysis. After that, rooted phylogenetic trees were constructed by Bayesian analysis and maximum likelihood (ML) method. These two methods were conducted with MrBayes v3.2.6 and PHyML website, respectively. ML phylogenetic analysis used LG + G model and set bootstrap as 1,000 replicates. Bayesian analysis was performed with following parameters: ngen = 2,000,000, nruns = 2, nchains = 4, aamodel = fixed (LG), rates = gamma, samplefreq = 1,000, burninfrac = 0.25. Finally, phylogenetic trees obtained were viewed and modified with FigTree v1.4.2. Synteny data of *khdrbs* genes were collected using tools available from Ensembl and NCBI database. Gene structural features of those chosen transcripts were analyzed using data from Ensembl database. Characteristics of the protein domain were predicted by SMART website.

Gene Cloning, WISH, Cryosection, and qRT-PCR

Fragments of zebrafish *khdrbs* genes were amplified with specific primers (**Table 1**) that were designed using Primer Premier 5.0 based on existing sequences from NCBI. The purified PCR products were sub-cloned into vector pGEM-T, which was sequenced to verify inserts orientation.

Digoxigenin (DIG)-labeled *khdrbs* antisense riboprobes were synthesized with linearized vectors (digested by *Nco*I restriction enzyme) and Sp6 RNA polymerase through *in vitro* transcription, while synthesis of sense riboprobes used *Sal*I restriction enzyme and T7 RNA polymerase. WISH experimental procedures followed the protocol described by Thisse and Thisse (2008). After staining, the embryos were fixed with 4% paraformaldehyde, rinsed with 70% ethanol and then mounted in glycerin for imaging. The embryos were also washed with PBS, soaked in 30% sucrose (diluted in PBS) overnight and cryosectioned.

Quantitative real-time PCR (qRT-PCR) was used to test the expression patterns of *khdrbs* in the different tissues of adult zebrafish. Total RNAs were extracted from the tissues gill, eye, brain, intestine, liver, heart, muscle, skin, spleen, testis, and ovary with TRIzol^{TM} (Invitrogen) and purified using Total RNA Kit I (OMEGA Bio-Tek). For each tissue, 1 μg RNA was used for next reverse transcription. The cDNAs were reverse transcribed using M-MLV reverse transcriptase (TaKaRa) and Oligo (dT) primers as guided by the manufacturer's instructions. *β-actin* and *EF1-α* were chosen as control to standardize the results by eliminating variations in mRNA and cDNA quantity and quality. qRT-PCR was conducted using the gene-specific primers (**Table 2**) on the ABI 7500 real-time PCR system (Applied Biosystem) with the 2× SYBR Premix Ex Taq^{TM} Kit (TaKaRa). A total of 0.5 μl cDNA was used as template in each replicate. Reaction conditions were: 95°C for 15 s as stage 1, then 40 cycles of 95°C for 15 s, 60°C for 15 s, and 72°C for 35 s as stage 2. The expression level of *khdrbs* genes relative to that of the housekeeping genes *β-actin* and *EF1-α* was calculated by the comparative threshold cycle (CT) method ($2^{-\Delta\Delta Ct}$). The experiments were performed in triplicate, and each replicate was from three experiments, i.e., the indicated replicates were both technical and biological replicates.

RESULTS

Three Groups of *khdrbs* Genes in Vertebrates

Recent studies have revealed that changes of key genetic toolkit at cellular and developmental level can generate novel characters in morphological structures. To better understand, it we need to dig into the evolutionary history of these critical genes. Considering that DNA sequences are more vulnerable and variable to selection pressure during long evolutionary history, we used KHDRBS protein sequences of major metazoan taxa

TABLE 1 | Primers used for specific antisense probe.

Gene	GenBank accession no.	Forward primer (5′–3′) Reverse primer (5′–3′)
khdrbs1a	NM_130925.1	CCTCAGGGCAGCACCATCAA
		CGCCACCTTCCCAGTCATCTT
khdrbs1b	NM_213235.1	TTAACTTTGTTGGGAAGATCCTTG
		GCTGCTTGTGGCTGACTGTAGTA
khdrbs2	NM_001077290.1	AAAAGTTTGAAGGCGACGAGTTG
		CTTGTGCGTGGCTTGCAGTG
khdrbs3	XM_009298280.2	TGTGCTTATTGAGGTGTTTGCTCC
		CGCTTTGTTGCGGTTGTTTGT

TABLE 2 | Primers used for qRT-PCR.

Gene	GenBank accession no.	Forward primer (5′–3′) Reverse primer (5′–3′)
khdrbs1a	NM_130925.1	CCTCAGGGCAGCACCATCAA
		TCCATACCTTCCACAGGCATCA
khdrbs1b	NM_213235.1	GAAAGGATTCTCATACCAGTCAAACA
		CAGGAGCAAACACCTCAATAAACAC
khdrbs2	NM_001077290.1	CAAGGGTTCAATGAGGGATAAAGG
		TGTGGAGGTCAGCCGCAATC
khdrbs3	XM_009298280.2	TCCATGTGCTTATTGAGGTGTTTGC
		TGTTCCCGGCGCTGATGTT
β-actin	NM_131031.1	GGTATTGTGATGGACTCTGGTGAT
		TCGGCTGTGGTGGTGAAG
EF1-α	NM_131263.1	GACTGGTGTCCTCAAGCCTGGTATG
		CAATCCAGCACTGGGGCGTAAC

for later analyses. We acquired multiple KHDRBS protein sequences mainly from NCBI and Ensembl. These sequences were aligned by ClustalW algorithm for next phylogeny analyses. We constructed phylogenetic trees by MrBayes (**Figure 1A**) and ML methods (**Figure 1B**). These two phylogenetic trees were generally consistent with each other. The analyses revealed that KHDRBS was traced back to *Trichoplax adhaerens* and *Hydra vulgaris*, the basal metazoans.

The invertebrate KHDRBS proteins formed a single clade, positioned at the base of vertebrate KHDRBS proteins. The overall phylogenetic relationship of KHDRBS well reflects the classic phylogeny of metazoan taxa. In contrast to only one KHDRBS group in invertebrates, there were three distinct clades of KHDRBS proteins in vertebrates, KHDRBS1, KHDRBS2, and KHDRBS3 (**Figure 1**), consistent with recent reports that humans, mice, and birds have three KHDRBS genes (Artzt and Wu, 2010). Apparently, the emergence of these three clades occurred around the split of subphylum Vertebrata, implicating that the three KHDRBS clades originated from sequential WGD events. Notably, absence of KHDRBS3 was also observed in Amazon molly (*Poecilia formosa*), suggesting that gene loss event may have happened during evolution process. In addition, KHDRBS protein sequences alignment of major vertebrates (**Figure 2**) showed that KHDRBS2 and KHDRBS3 shared higher identity (range from 44.5 to 69.8%) than that between KHDRBS1 and KHDRBS2 (range from 43.6 to 65.9%)
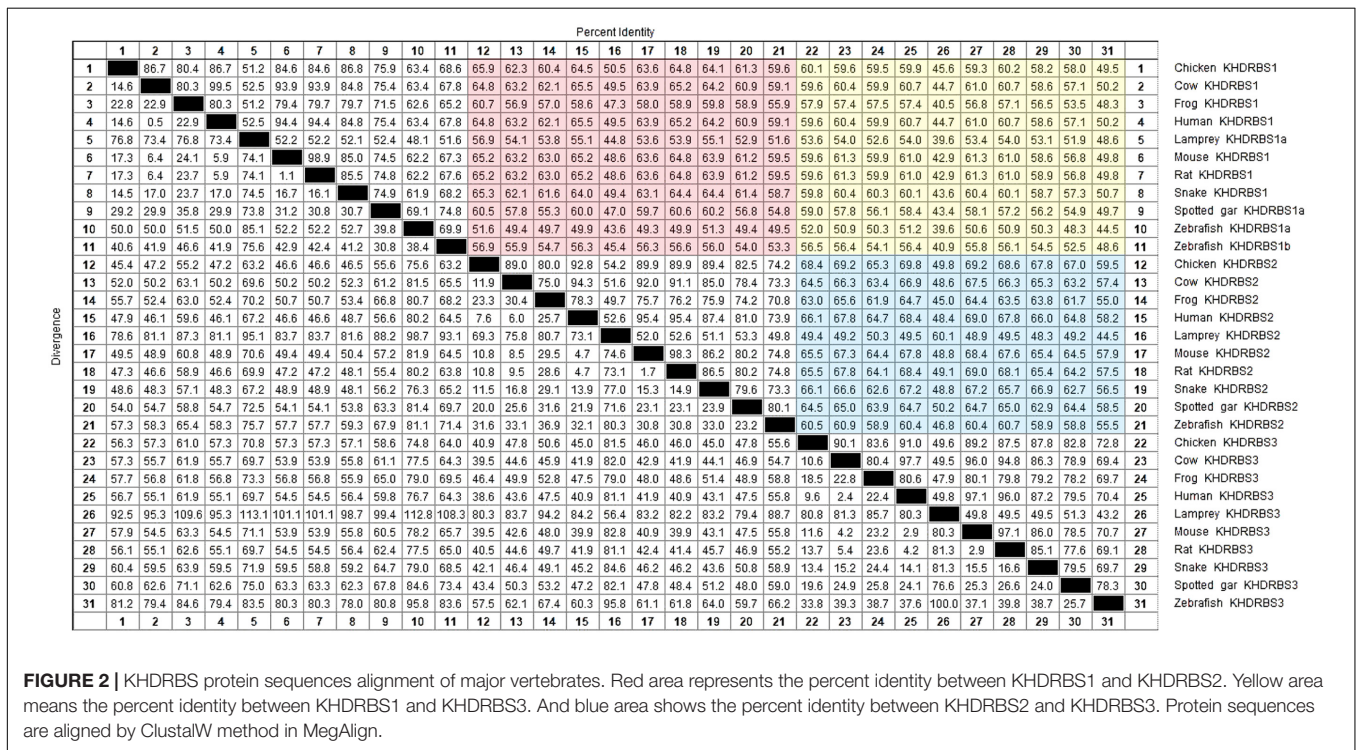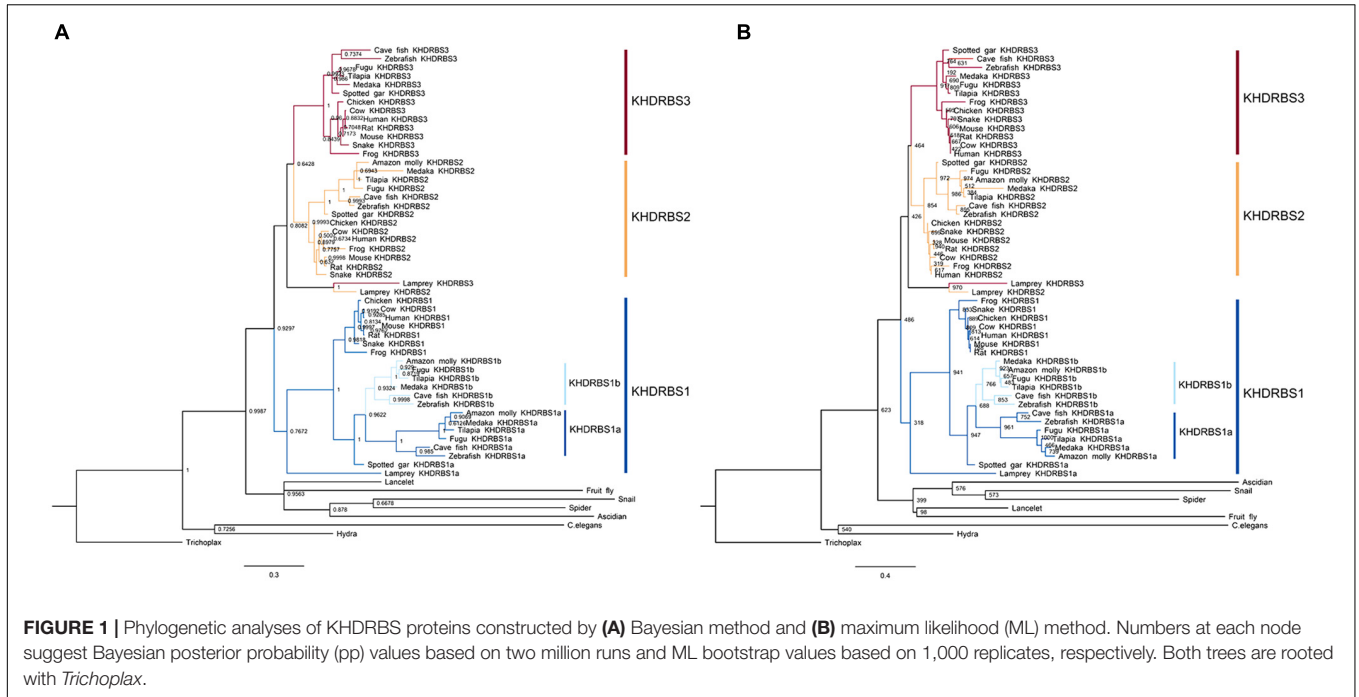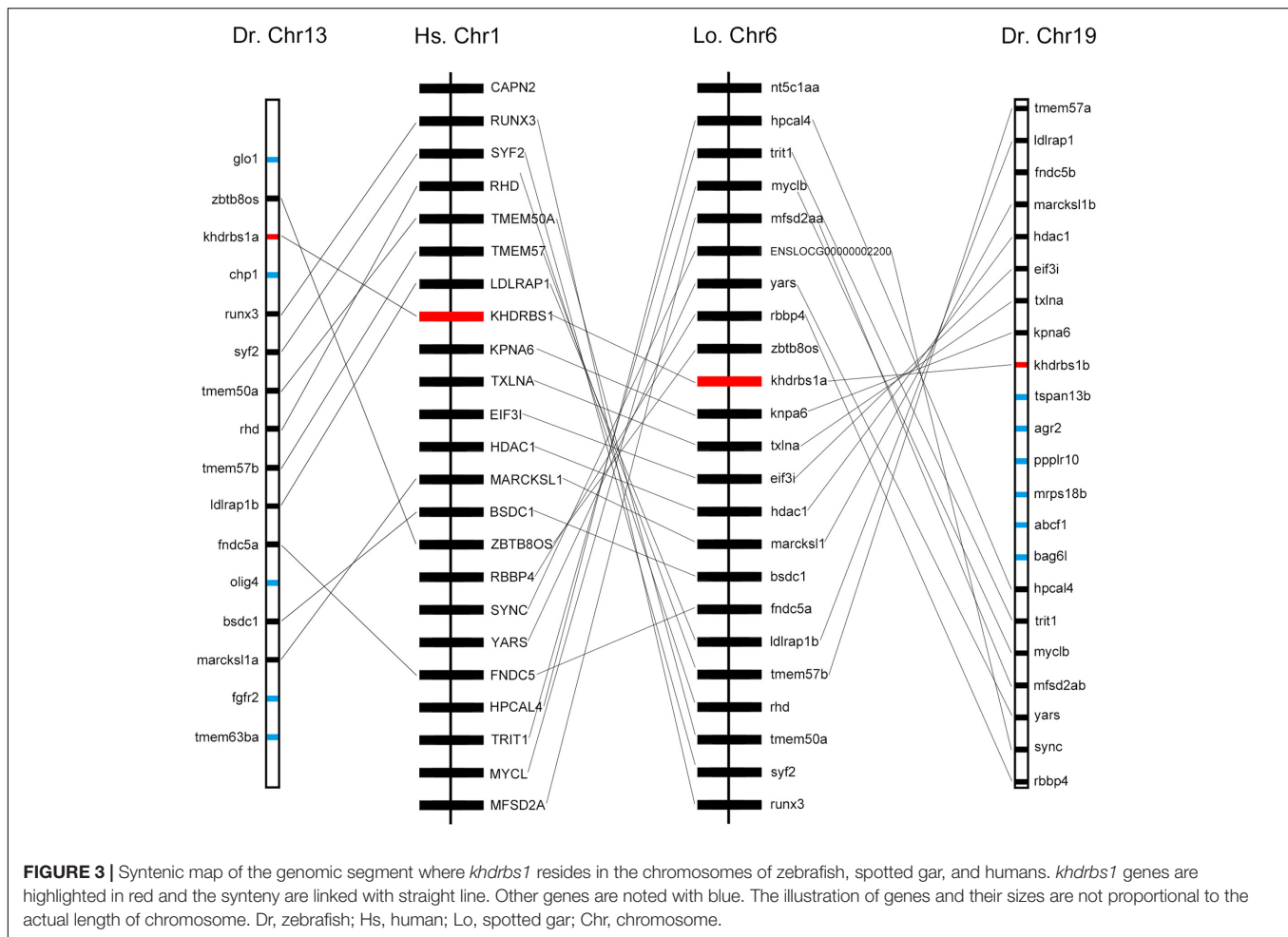
FIGURE 1 | Phylogenetic analyses of KHDRBS proteins constructed by **(A)** Bayesian method and **(B)** maximum likelihood (ML) method. Numbers at each node suggest Bayesian posterior probability (pp) values based on two million runs and ML bootstrap values based on 1,000 replicates, respectively. Both trees are rooted with *Trichoplax*.

Percent Identity

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31		
1	■	86.7	80.4	86.7	51.2	84.6	84.6	86.8	75.9	63.4	68.6	65.9	62.3	60.4	64.5	50.5	63.6	64.8	64.1	61.3	59.6	60.1	59.6	59.5	59.9	45.6	59.3	60.2	58.2	58.0	49.5	Chicken KHDRBS1	1
2	14.6	■	80.3	99.5	52.5	93.9	93.9	84.8	75.4	63.4	67.8	64.8	63.2	62.1	65.5	49.5	63.9	65.2	64.2	60.9	59.1	59.6	60.4	59.9	60.7	44.7	61.0	60.7	58.6	57.1	50.2	Cow KHDRBS1	2
3	22.8	22.9	■	80.3	51.2	79.4	79.7	79.7	71.5	62.6	65.2	60.7	56.9	57.0	58.6	47.3	58.0	58.9	59.8	58.9	55.9	57.9	57.4	57.5	57.4	40.5	56.8	57.1	56.5	53.5	48.3	Frog KHDRBS1	3
4	86.7	0.5	22.9	■	52.5	94.4	94.4	84.8	75.4	63.4	67.8	64.8	63.2	62.1	65.5	49.5	63.9	65.2	64.2	60.9	59.1	59.6	60.4	59.9	60.7	44.7	61.0	60.7	58.6	57.1	50.2	Human KHDRBS1	4
5	76.8	73.4	76.8	73.4	■	52.2	52.2	52.1	52.4	48.1	51.6	56.9	54.1	53.8	55.1	44.8	53.6	53.9	55.1	52.9	51.6	53.6	54.0	52.6	54.0	39.6	53.4	54.0	53.1	51.9	48.6	Lamprey KHDRBS1a	5
6	17.3	6.4	24.1	5.9	74.1	■	98.9	85.0	74.5	62.2	67.3	65.2	63.2	65.2	48.6	63.6	64.8	63.9	61.2	59.5	59.6	61.3	59.9	61.0	42.9	61.3	61.0	58.6	56.8	49.8	Mouse KHDRBS1	6
7	17.3	6.4	23.7	5.9	74.1	1.1	■	85.5	74.8	62.2	67.6	65.2	63.2	63.0	65.2	48.6	63.6	64.8	63.9	61.2	59.5	59.6	61.3	59.9	61.0	42.9	61.3	61.0	58.6	56.8	49.8	Rat KHDRBS1	7
8	14.5	17.0	23.7	17.0	74.5	16.7	16.1	■	74.9	61.9	68.2	65.3	62.1	61.6	64.0	49.4	63.1	64.4	64.4	61.4	58.7	59.8	60.4	60.3	60.1	43.6	60.4	60.1	58.7	57.3	50.7	Snake KHDRBS1	8
9	29.2	29.9	35.8	29.9	73.8	31.2	30.8	30.7	■	69.1	74.8	60.5	57.8	55.3	60.0	47.0	59.7	60.6	60.2	56.8	54.8	57.8	56.1	58.4	43.4	58.1	57.2	56.2	54.9	49.7	Spotted gar KHDRBS1a	9
10	50.0	50.0	51.5	50.0	85.1	52.2	52.2	52.7	39.8	■	69.9	51.6	49.4	49.7	49.9	43.6	49.3	49.9	51.3	49.4	49.5	52.0	50.9	50.3	51.2	39.6	50.6	50.9	50.3	48.3	44.5	Zebrafish KHDRBS1a	10
11	40.6	41.9	46.6	41.9	75.6	42.9	42.4	41.2	30.8	38.4	■	56.9	55.9	54.7	56.3	45.4	56.6	56.0	54.0	53.3	56.5	56.4	54.1	56.4	40.9	55.8	56.1	54.5	52.5	48.6	Zebrafish KHDRBS1b	11
12	45.4	47.2	55.2	47.2	63.2	46.6	46.6	46.5	55.6	75.6	63.2	■	89.0	80.0	92.8	54.2	89.9	89.9	89.4	82.5	74.2	68.4	69.2	65.3	69.8	49.8	69.2	68.6	67.8	67.0	59.5	Chicken KHDRBS2	12
13	52.0	50.2	63.1	50.2	69.6	50.2	50.2	52.3	61.2	81.5	65.5	11.9	■	75.0	94.3	51.6	92.0	91.1	85.0	78.4	73.3	64.5	66.3	63.4	66.9	48.6	67.5	66.3	65.3	63.2	57.4	Cow KHDRBS2	13
14	55.7	52.4	63.0	52.4	70.2	50.7	50.7	53.4	66.8	80.7	68.2	23.3	30.4	■	78.3	49.7	75.7	76.2	75.9	74.2	70.8	63.0	65.6	61.9	64.7	45.0	64.4	63.5	63.8	61.7	55.0	Frog KHDRBS2	14
15	47.9	46.1	59.6	46.1	67.2	46.6	46.6	48.7	56.6	80.2	64.5	7.6	6.0	25.7	■	52.6	95.4	95.4	87.4	81.0	73.9	66.1	67.8	64.7	68.4	48.4	69.0	67.8	66.0	64.8	58.2	Human KHDRBS2	15
16	78.6	81.1	87.3	81.1	95.1	83.7	83.7	81.6	88.2	98.7	93.1	69.3	75.8	80.7	73.1	■	52.0	52.6	51.1	53.3	49.8	49.4	49.2	50.3	49.5	60.1	48.9	49.5	48.3	49.2	44.5	Lamprey KHDRBS2	16
17	49.5	48.9	60.8	48.9	70.6	49.4	49.4	50.4	57.2	81.9	64.5	10.8	8.5	29.5	4.7	74.6	■	98.3	86.2	80.2	74.8	65.5	67.8	64.4	67.8	48.8	68.4	67.6	65.4	64.5	57.9	Mouse KHDRBS2	17
18	47.3	46.6	58.9	46.6	69.9	47.2	47.2	48.1	55.4	80.2	63.8	10.0	9.5	28.6	4.7	73.1	1.7	■	86.5	80.2	74.8	65.5	67.8	64.1	68.4	49.1	69.0	68.1	65.4	64.2	57.5	Rat KHDRBS2	18
19	48.6	48.3	57.1	48.3	67.2	48.9	48.9	48.1	56.2	76.3	65.2	11.5	16.8	29.1	13.9	77.0	15.3	14.9	■	79.6	73.3	66.1	66.6	62.6	67.2	48.8	67.2	65.7	66.9	62.7	56.5	Snake KHDRBS2	19
20	54.0	54.7	58.8	54.7	72.5	54.1	54.1	53.8	63.3	81.4	69.7	20.0	25.6	31.6	21.9	71.6	23.1	23.1	23.9	■	80.1	64.5	65.0	63.9	64.7	50.2	64.7	65.0	62.9	64.4	58.5	Spotted gar KHDRBS2	20
21	57.3	58.3	65.4	58.3	75.7	57.7	57.7	57.7	59.3	67.9	81.1	71.4	31.6	33.1	36.9	32.1	80.3	30.8	30.8	33.0	23.2	■	60.5	60.9	58.9	60.4	46.8	60.4	60.7	58.9	58.8	55.5	Zebrafish KHDRBS2	21
22	56.3	57.3	61.0	57.3	70.8	57.3	57.3	57.1	58.6	74.8	64.0	40.9	47.8	50.6	45.0	81.5	46.0	46.0	45.0	47.8	55.6	■	90.1	83.6	91.0	49.6	89.2	87.5	87.8	82.8	72.8	Chicken KHDRBS3	22
23	57.3	55.7	61.9	55.7	69.7	53.9	53.9	55.8	61.1	77.5	64.3	39.5	44.6	45.9	41.9	82.0	42.9	41.9	44.1	46.9	54.7	10.6	■	80.4	97.7	49.5	96.0	94.8	86.3	78.9	69.4	Cow KHDRBS3	23
24	57.7	56.8	61.8	56.8	73.3	56.8	56.8	55.9	65.0	79.0	69.5	46.4	49.9	52.8	47.5	79.0	48.0	48.6	51.4	48.9	58.8	18.5	22.8	■	80.6	47.9	80.1	79.8	79.2	78.2	69.7	Frog KHDRBS3	24
25	56.7	56.1	61.9	55.1	69.7	54.5	54.5	56.4	59.8	76.7	64.3	38.6	43.6	47.5	40.9	81.1	41.9	40.9	43.1	47.5	55.8	9.6	2.4	22.4	■	49.8	97.1	96.0	87.2	79.5	70.4	Human KHDRBS3	25
26	92.5	95.3	109.6	95.3	113.1	101.1	101.1	98.7	99.4	112.8	108.3	80.3	83.7	94.2	84.2	56.4	83.2	82.2	83.2	79.4	88.7	80.8	81.3	85.7	80.3	■	49.8	49.5	49.5	51.3	43.2	Lamprey KHDRBS3	26
27	57.9	54.5	63.3	54.5	71.1	53.9	53.9	55.8	60.5	78.2	65.7	39.5	42.6	48.0	39.9	82.8	40.9	39.9	43.1	47.5	55.8	11.6	4.2	23.2	2.9	80.3	■	97.1	86.0	78.5	70.7	Mouse KHDRBS3	27
28	56.1	55.1	62.6	55.1	69.7	54.5	54.5	56.4	62.4	77.5	65.0	40.5	44.6	49.7	41.9	81.1	41.4	41.4	45.7	49.9	58.3	13.7	5.4	23.6	4.2	81.3	2.9	■	85.1	77.6	69.1	Rat KHDRBS3	28
29	60.4	59.5	63.9	59.5	71.9	59.5	58.8	59.2	64.7	79.0	68.5	42.1	46.4	49.1	45.2	84.6	46.2	46.2	43.6	50.8	58.9	13.4	15.2	24.4	14.1	81.3	15.5	16.6	■	79.5	69.7	Snake KHDRBS3	29
30	60.8	62.6	71.1	62.6	75.0	63.3	63.3	62.3	67.8	84.6	73.4	43.4	50.3	53.2	47.2	82.1	47.8	48.4	51.2	48.0	59.0	19.6	24.9	25.8	24.1	76.6	25.3	26.6	24.0	■	78.3	Spotted gar KHDRBS3	30
31	81.2	79.4	84.6	79.4	83.5	80.3	80.3	78.0	80.8	95.8	83.6	57.5	62.1	67.4	60.3	95.8	61.1	61.8	64.0	59.7	66.2	33.8	39.3	38.7	37.6	100.0	37.1	39.8	38.7	25.7	■	Zebrafish KHDRBS3	31
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31		

(Divergence — lower-left triangle)

FIGURE 2 | KHDRBS protein sequences alignment of major vertebrates. Red area represents the percent identity between KHDRBS1 and KHDRBS2. Yellow area means the percent identity between KHDRBS1 and KHDRBS3. And blue area shows the percent identity between KHDRBS2 and KHDRBS3. Protein sequences are aligned by ClustalW method in MegAlign.

or KHDRBS1 and KHDRBS3 (range from 39.6 to 61.3%), which meant KHDRBS2 and KHDRBS3 formed a more close evolutionary relationship.

Duplication of *khdrbs1* Genes in Teleosts

We found that *khdrbs1* genes in teleost were clustered into two sub clades, as *khdrbs1a* and *khdrbs1b*. Compared to tetrapod *khdrbs1* genes, *khdrbs1a* and *khdrbs1b* formed a more close branch with a sister clade spotted gar (**Figure 1**). Spotted gar has an important evolutionary status because it diverged from teleost before gene duplication happened in teleost (Braasch et al., 2016). Therefore, this indicated that *khdrbs1a* and *khdrbs1b* emerged simultaneously within teleost lineage, i.e., they were generated by teleost-specific genome

FIGURE 3 | Syntenic map of the genomic segment where *khdrbs1* resides in the chromosomes of zebrafish, spotted gar, and humans. *khdrbs1* genes are highlighted in red and the synteny are linked with straight line. Other genes are noted with blue. The illustration of genes and their sizes are not proportional to the actual length of chromosome. Dr, zebrafish; Hs, human; Lo, spotted gar; Chr, chromosome.

duplication (Hoegg et al., 2004; Jaillon et al., 2004). As phylogenetics, albeit widely analyzed, may be still short of enough reliability to elucidate gene orthologous relationships due to sophisticated situations like gene loss or gene expansion in specific lineage, we thus performed a syntenic analysis among zebrafish *khdrbs1*, spotted gar *khdrbs1a* and human *khdrbs1* genes, which represented teleost and tetrapod, respectively (**Figure 3**), to further confirm the relationship between *khdrbs1a* and *khdrbs1b*. The results revealed that the flanking genes beside both *khdrbs1a* and *khdrbs1b* could be found in the neighboring region around spotted gar and human *khdrbs1*, which meant that *khdrbs1* in zebrafish was orthology to spotted gar and human *khdrbs1* and remained an evolutionarily conserved synteny.

To pinpoint the relationships between the duplications of teleost *khdrbs1* genes, we conducted protein domain analysis by SMART. We found that when compared with human KHDRBS1, both zebrafish KHDRBS1a and KHDRBS1b possessed similar protein domains, though they shared more similarities to spotted gar KHDRBS1a (**Figure 4A**). Additionally, the differences between genomic intron–exon structures of zebrafish *khdrbs1a* and *khdrbs1b* were also analyzed. Both *khdrbs1a* and *khdrbs1b* had an intron–exon structure similar to that of other species

(**Figure 4B**), with the highly conserved region within the KH domains (**Figure 4C**).

Similar but Distinct Expression Patterns of Zebrafish *khdrbs* Genes

WGD commonly leads to the functional specialization of paralogous genes, or gene co-option, and contributes to vertebrate morphological innovations. Recently, *khdrbs1a* and *khdrbs1b* were found both mainly expressed in the brain primordium of zebrafish embryos, hinting at the clue that *khdrbs* genes may play critical roles in vertebrate brain development. We thus hypothesized that investigating the gene expression patterns of the four zebrafish *khdrbs* genes might illuminate the origin of the vertebrate brain during evolution.

Zebrafish have two *khdrbs1* genes, *khdrbs1a* and *khdrbs1b*. Different from control group (Supplementary Figures 1A–F), both *khdrbs1a* and *khdrbs1b* were pan-expressed in the blastoderm and the embryo *per se* (**Figures 5A–E,A′–E′**). At 24 hpf, *khdrbs1a* expression was detected in the brain (**Figures 6A–D**), eye, spinal cord, lateral line primordium, gonad primordium, pectoral fin, and caudal fin (**Figures 5F–I**), while *khdrbs1b* expression was relatively weak in the brain (**Figures 6A′–D′**), eye, lateral line primordium, and spinal

FIGURE 4 | Bioinformatic analysis of KHDRBS1 proteins in cavefish, human, spotted gar, and zebrafish. **(A)** Protein domain structure predictions of KHDRBS1 by SMART. Purple bars represent low complexity regions. Vertical colored lines indicate the intron positions. KH domains are shown in oval shapes with "KH". **(B)** Intron–exon structures of *khdrbs1*. The solid boxes denote the gene coding regions, and the empty boxes mean untranscribed regions. Dotted lines represent the introns. **(C)** Protein sequences alignment of KHDRBS1. Protein sequences are aligned by ClustalW method in MegAlign. The highly conserved part is noted by red rectangles.

FIGURE 5 | Gene expression patterns of zebrafish *khdrbs1a* and *khdrbs1b* genes during early development detected by WISH. Stages of embryonic development: 0.45 hpf **(A,A′)**, 3 hpf **(B,B′)**, 6 hpf **(C,C′)**, 10 hpf **(D,D′)**, 14 hpf **(E,E′)**, 24 hpf **(F–I, F′–I′)**, 48 hpf **(J–M, J′–M′)**, and 72 hpf **(N–Q, N′–Q′)**. Expression sites are labeled with abbreviation as follows. cf, caudal fin; e, eye; gp, gonad primordium; hb, hindbrain; lln, lateral line neuromast; llp, lateral line primordium; mb, midbrain; ov, otic vesicle; pa, pharyngeal arch; pd, pronephric ducts; pt, pectoral fin; sp, spinal cord.

cord (**Figures 5F′–I′**). By 48 hpf (**Figures 5J–M**) and 72 hpf (**Figures 5N–Q**), *khdrbs1a* expression in the brain (**Figures 6E–L**) and pectoral fin enhanced, and the signal at the lateral line neuromast was evident. Additional expressions were also observed in the pharyngeal arch, pronephric ducts (**Figure 7**) and otic vesicle, but the expression signals in the spinal cord and caudal fin apparently faded. Compared to *khdrbs1a*, *khdrbs1b* was not expressed in the lateral line by 48 hpf (**Figures 5J′–M′**); instead, its expression was restricted to the brain (**Figures 6E′–H′**), eye, pharyngeal arch and pectoral fin, and this expression pattern remained till 72 hpf (**Figures 5N′–Q′**, **6I′–L′**).

We then examined the expression patterns of zebrafish *khdrbs2* and *khdrbs3* expression. At the early developmental stages, both *khdrbs2* and *khdrbs3*, like *khdrbs1*, were pan-expressed in the blastoderm and the embryo *per se* (**Figures 8A–E,A′–E′** and Supplementary Figures 1G–L). However, *khdrbs2* and *khdrbs3* genes both later generated more

specific expression patches in the brain, forming a sharp contrast to that of *khdrbs1*. By 24 hpf (**Figures 8F,I,L, 9A–D**), *khdrbs2* was expressed in the telencephalon, epiphysis, hypothalamus, hind rhombomeres, and spinal cord. Signals in these regions gradually expanded in the midbrain and hindbrain with larva development. At 48 hpf (**Figures 8G,J,M, 9E–H**) and 72 hpf (**Figures 8H,K,N, 9I–L**), expression of *khdrbs2* was primarily restricted to the tegmentum and tectum, respectively, and expression in the spinal cord disappeared. More interestingly, *khdrbs3* signal was observed in the telencephalon, diencephalon, midbrain, and hindbrain at 24 hpf (**Figures 8F′,I′,L′, 9A′–D′**), and then became localized intensely in the telencephalon, hypothalamus, tegmentum, and hind rhombomeres (**Figures 8G′,J′,M′, 9E′–H′**). By 72 hpf (**Figures 8H′,K′,N′, 9I′–L′**), *khdrbs3* expression spread out around the midbrain and hindbrain, with additional signals emerging in the retina and tectum.

Finally, we carried out qRT-PCR to explore the expression profiles of *khdrbs* in adult zebrafish (**Figure 10** and
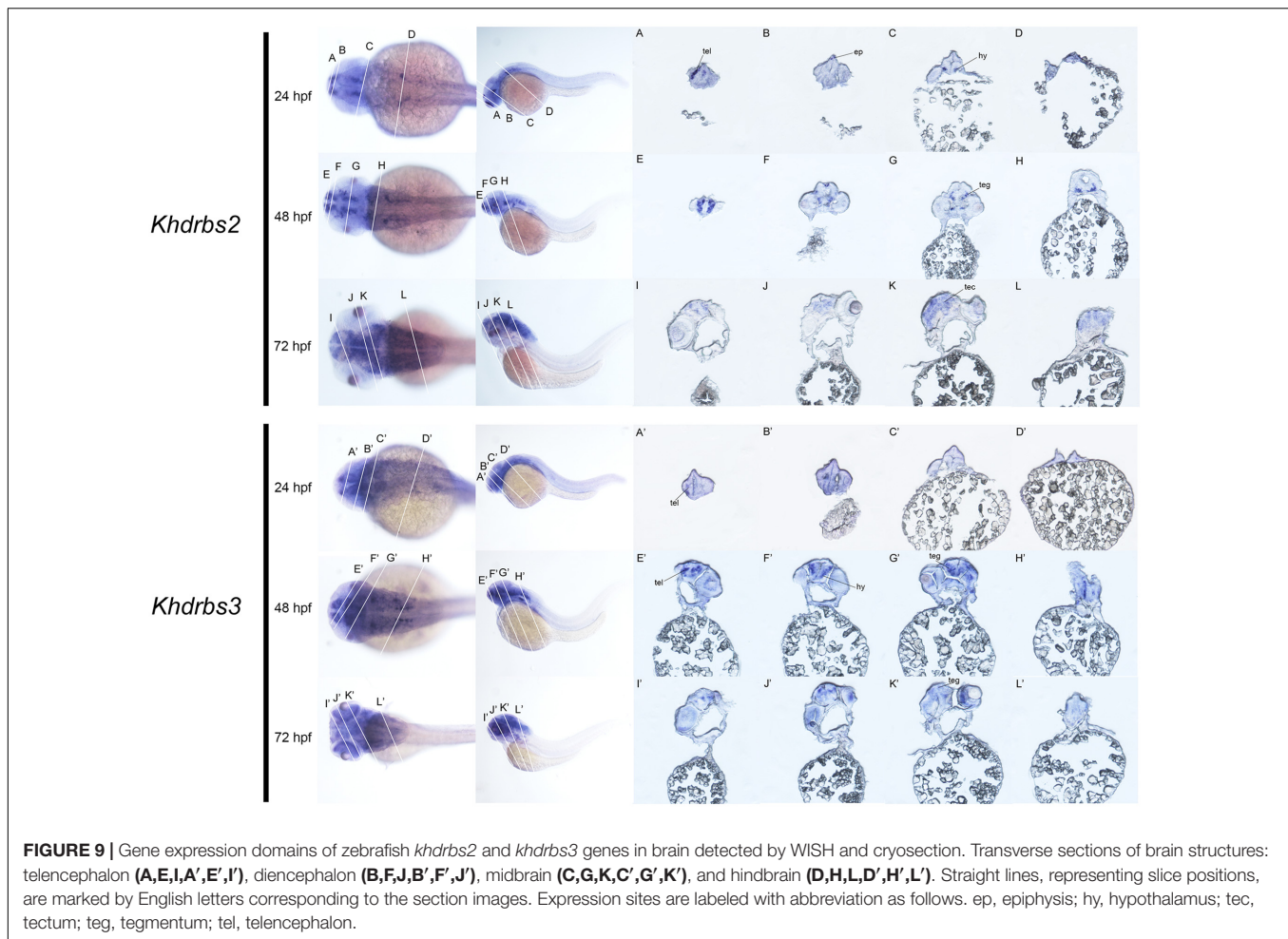
FIGURE 6 | Gene expression domains of zebrafish *khdrbs1a* and *khdrbs1b* genes in brain detected by WISH and cryosection. Transverse sections of brain structures: telencephalon **(E,I,E',I')**, diencephalon **(F,J,F',J')**, midbrain **(G,K,G',K')** and hindbrain **(C,D,H,L,C',D',H',L')**. Brain structure of zebrafish at 24 hpf **(A,B,A',B')**. Straight lines, representing slice positions, are marked by English letters corresponding to the section images. Expression sites are labeled with abbreviation as follows. lln, lateral line neuromast.

FIGURE 7 | Gene expression domains of zebrafish *khdrbs1a* gene in pronephric ducts detected by WISH and cryosection at 48 hpf. Straight lines represent slice positions **(A,B)**. Transverse sections of zebrafish pronephric ducts at 48 hpf **(C–E)**. Expression sites are labeled with abbreviation as follows. pd, pronephric ducts.

Supplementary Figure 2). As shown in **Figures 10A,B**, both *khdrbs1a* and *khdrbs1b* shared similarities in their expression profiles. They were both detected in all the tissues tested, with the highest expression levels in the brain. The obvious difference between *khdrbs1a* and *khdrbs1b* was that the latter gene had higher expression in the gonads. In contrast, the expression of

khdrbs2 and *khdrbs3* became more distinct: both *khdrbs2* and *khdrbs3* signal was predominantly detected in the brain, with lower expression in the eye and testis (**Figures 10C,D**). These indicated that these genes are mainly expressed in the brains of adult zebrafish in a tissue-specific fashion, consistent with their expression patterns in embryogenesis.

FIGURE 8 | Gene expression patterns of zebrafish *khdrbs2* and *khdrbs3* genes during early development detected by WISH. Stages of embryonic development: 0.45 hpf **(A,A′)**, 3 hpf **(B,B′)**, 6 hpf **(C,C′)**, 10 hpf **(D,D′)**, 14 hpf **(E,E′)**, 24 hpf **(F,I,L,F′,I′,L′)**, 48 hpf **(G,J,M,G′,J′,M′)**, and 72 hpf **(H,K,N,H′,K′,N′)**. Expression sites are labeled with abbreviation as follows. di, diencephalon; ep, epiphysis; hb, hindbrain; hr, hind rhombomeres; hy, hypothalamus; mb, midbrain; re, retina; sp, spinal cord; tec, tectum; teg, tegmentum; tel, telencephalon.

DISCUSSION

One goal for evolutionary biology is to decipher the mystery underlying morphological novelties and complexity of extant organisms, especially vertebrates. As the largest and most diverse lineage in chordate, Vertebrata is characterized by several key morphological characteristics such as a tripartite brain.

Such morphological characteristics are usually originated from variation of the related genetic toolkits. Hence, expression pattern and evolutionary analyses of key genes in these toolkits can deepen our understanding of the emergence of morphological innovations during evolution. In this study, we first analyzed the molecular evolution of *khdrbs* gene family, and then investigated zebrafish *khdrbs* genes expression

FIGURE 9 | Gene expression domains of zebrafish *khdrbs2* and *khdrbs3* genes in brain detected by WISH and cryosection. Transverse sections of brain structures: telencephalon **(A,E,I,A′,E′,I′)**, diencephalon **(B,F,J,B′,F′,J′)**, midbrain **(C,G,K,C′,G′,K′)**, and hindbrain **(D,H,L,D′,H′,L′)**. Straight lines, representing slice positions, are marked by English letters corresponding to the section images. Expression sites are labeled with abbreviation as follows. ep, epiphysis; hy, hypothalamus; tec, tectum; teg, tegmentum; tel, telencephalon.

patterns during embryonic development and in adulthood of zebrafish. The results indicated that after vertebrate WGD, *khdrbs* gene family was co-opted for brain development, and the duplication and diversification of *khdrbs* genes may promote the origin of vertebrate elaborate brains during evolution.

Evolutionary History of *khdrbs* Genes

The *khdrbs* genes are evolutionarily conserved. We found that *khdrbs* date far back to the basal metazoans such as placozoa and cnidarian. Except for fruit fly, all the invertebrates analyzed possess a single *khdrbs* gene in each species. In fruit fly, there are five *khdrbs* genes, including *qkr54B*, *qkr58E-1*, *qkr58E-2*, *qkr58E-3*, and *nsr* (Ding et al., 2010; Volk, 2010), which are all localized on a region of the chromosome 2R, and clustered into one branch (Supplementary Figure 3), suggesting that they were generated by local tandem gene duplication of the ancestral *khdrbs* gene. Based on their expression patterns and tandem duplicated gene traits (Fan et al., 2008; Zhou et al., 2008; Volk, 2010), we speculated that these five genes perform similar biological functions. In contrast, three *khdrbs* genes, *khdrbs1*, *khdrbs2*, and *khdrbs3*, are identified in all the vertebrates examined, including lamprey (Xu et al., 2016). The

birth of these three *khdrbs* gene clades coincides with the origin of vertebrates. As there existed two rounds (2R) of WGD happened in the vertebrate ancestors, it is thus expected that four *khdrbs* genes would be generated by these WGD. Interestingly, there are only three *khdrbs* genes identified so far in vertebrates. It is thus highly likely that only one copy out of the two *khdrbs* genes generated by the first round WGD underwent further duplication, thereby leading to the appearance of three *khdrbs* genes in vertebrates. This seems supported by the facts that KHDRBS2 and KHDRBS3 were grouped together, forming a sub-clade in the phylogenetic trees (**Figure 1**), and shared higher identity (**Figure 2**). However, the possibility cannot be ruled out that this may be due to gene loss as observed in Amazon molly. Additionally, we found duplication of *khdrbs1* in teleost lineage. This duplication happened between the split of spotted gar and bony fish, agreeing with the third rounds (3R) of WGD in bony fish (Braasch et al., 2016).

On the basis of our analyses on *khdrbs* gene family evolution and the mainstream opinions in WGD, we propose an evolutionary model for *khdrbs* gene family, as depicted in **Figure 11**. In brief, an invertebrate *khdrbs* gene underwent duplication once, and one copy of the resulting genes, as

FIGURE 10 | Gene expression patterns of zebrafish **(A)** *khdrbs1a*, **(B)** *khdrbs1b*, **(C)** *khdrbs2*, and **(D)** *khdrbs3* in different tissues. *β-actin* is used as internal control for normalization. Relative expression data is calculated by the method of $2^{-\Delta\Delta Ct}$. Vertical bars represent the mean ± standard deviation (SD) ($n = 3$). Data are from three independent experiments which were performed in triplicate.

FIGURE 11 | Molecular evolution of *khdrbs* genes and the origin of vertebrate tripartite brain. The evolutionary relationships of metazoan and Vertebrata are based on our analyses above and recent phylogenomic analyses. Two red bars represent the 2R WGD events happening around the origin of vertebrate linage. The yellow one indicates teleost specific WGD event. Those blue rectangles represent the process of *khdrbs* gene duplication and diversification. The branch lengths are not proportional to the time of diversification.

exemplified by *khdrbs1*, remained unchanged, while the other copy underwent duplication once more, thereby creating the vertebrate *khdrbs1*, *khdrbs2*, and *khdrbs3*. In teleost, the *khdrbs1* was further duplicated through bony fish specific WGD, forming *khdrbs1a* and *khdrbs1b*. In some species like Amazon molly, *khdrbs3* is lost in a lineage-specific manner.

Distinct and Overlapping Expression Profiles of Zebrafish *khdrbs* Genes

Expression pattern of genes is often associated with their functions in early development. We found all the zebrafish *khdrbs* genes were primarily expressed in the brain, while *khdrbs1a* was expressed in extra domains like the lateral line primordium, gonad primordium, and pronephric ducts. In adult fish, all the *khdrbs* also showed strong expression in the brain. Additionally, *khdrbs1a*, *khdrbs1b*, and *khdrbs3* were found expressed abundantly in the gonads. These overlapping but distinct expression patterns may be a reflection of the sub-functionalization and/or neo-functionalization of gene duplicates after WGD. The split of bony fish from tetrapod happened around 450 million years ago. The fact that all the four *khdrbs* genes are retained in bony fish since the split may be because the KHDRBS proteins are signal-transduction related molecules (Putnam et al., 2008). Retained *khdrbs* genes may enrich their functions through changes in their expression patterns, which may then be co-opted in regulation network to generate physiological and morphological evolution.

Co-option of Vertebrate *khdrbs* Genes for Brain Development

Species diverges from common ancestors through changes in their DNA. Duplicated genes that impart new roles from their ancestors can be co-opted by changing their regulation process or altering their coding proteins to promote developmental and physiological features (Carroll, 2000; True and Carroll, 2002; Conant and Wolfe, 2008). A complex tripartite brain especially telencephalon is a key trait of vertebrates (Murakami et al., 2005; Sugahara et al., 2013, 2016). The highly plastic vertebrate brain influences all the basic biological functions, advance order processing and behavior control (Koscik and Tranel, 2012). As the most important part in organisms, however, it is hard to trace brain evolution because brains themselves are rarely fossilized. Hence, to explore the evolutionary origin of vertebrate brain needs to rely on molecular evolution and developmental studies.

Currently, mouse *khdrbs* genes were found expressed primarily in the brain, differing from that of the invertebrate orthologs, and *khdrbs* KO mice exhibited brain defects, indicating that *khdrbs* genes may be in the genetic toolkit correlated to vertebrate brain development (Ehrmann et al., 2013). In invertebrates, *khdrbs* genes were found expressed specifically in germ line (Lee and Schedl, 2001; Ding et al., 2010), so

as zebrafish *khdrbs1a*, which was also expressed in the gonad primordium. Thus we surmise that one of the functions about *khdrbs* genes is to regulate gonad genesis. In addition to expression in gonads, zebrafish *khdrbs* genes were also found expressed in new domains such as the substructures of brain, especially in telencephalon. Telencephalon is an utterly vertebrate innovation (Murakami et al., 2005; Šestak and Domazet-Lošo, 2015). Considering that KHDRBS proteins act as splicing regulator molecules to control alternative splicing, and the brains have the largest amount of alternative splicing in the body, this distinct expression pattern of *khdrbs* genes gives a hint that the duplicated *khdrbs* genes were co-opted for motivating the evolutionary origin of vertebrate brain. Further studies on the expression and functions of *khdrbs* genes in other representative species within vertebrate lineage will better clarify the roles of KHDRBS played in brain evolution.

ETHICS STATEMENT

All the zebrafish used in the experiments were treated in accordance with the guidelines of the Laboratory Animal Administration Law of China, with the permit number SD2007695 approved by the Ethics Committee of the Laboratory Animal Administration of Shandong Province.

AUTHOR CONTRIBUTIONS

SZ and DJ conceived and designed the experiment. SW, SF, and QY conducted the experiment. SW, HL, and DJ analyzed the data. SW, QY, and ZW prepared the figures and the tables. All the authors were involved in discussion and editing of the manuscript.

FUNDING

ACKNOWLEDGMENT

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2017.00225/full#supplementary-material

REFERENCES

Artzt, K., and Wu, J. I. (2010). "Star Trek," in *Post-Transcriptional Regulation by STAR Proteins*, ed. T. Volk (New York, NY: Springer Science & Business Media), 1–24.

Braasch, I., Gehrke, A. R., Smith, J. J., Kawasaki, K., Manousaki, T., Pasquier, J., et al. (2016). The spotted gar genome illuminates vertebrate evolution and facilitates human-to-teleost comparisons. *Nat. Genet.* 48, 427–437. doi: 10.1038/ng.3526

Cañestro, C., Albalat, R., Irimia, M., and Garcia-Fernàndez, J. (2013). Impact of gene gains, losses and duplication modes on the origin and diversification of vertebrates. *Semin. Cell Dev. Biol.* 24, 83–94. doi: 10.1016/j.semcdb.2012.12.008

Carroll, S. B. (2000). Endless forms: the evolution of gene regulation and morphological diversity. *Cell* 101, 577–580. doi: 10.1016/S0092-8674(00)80868-5

Conant, G. C., and Wolfe, K. H. (2008). Turning a hobby into a job: how duplicated genes find new functions. *Nat. Rev. Genet.* 9, 938–950. doi: 10.1038/nrg2482

Dehal, P., and Boore, J. L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate. *PLOS Biol.* 3:e314. doi: 10.1371/journal.pbio.0030314

Di Fruscio, M., Chen, T., and Richard, S. (1999). Characterization of Sam68-like mammalian proteins SLM-1 and SLM-2: SLM-1 is a Src substrate during mitosis. *Proc. Natl. Acad. Sci. U.S.A.* 96, 2710–2715. doi: 10.1073/pnas.96.6.2710

Ding, Y., Zhao, L., Yang, S., Jiang, Y., Chen, Y., Zhao, R., et al. (2010). A young *Drosophila* duplicate gene plays essential roles in spermatogenesis by regulating several Y-linked male fertility genes. *PLOS Genet.* 6:e1001255. doi: 10.1371/journal.pgen.1001255

Ehrmann, I., Dalgliesh, C., Liu, Y., Danilenko, M., Crosier, M., Overman, L., et al. (2013). The tissue-specific RNA binding protein T-STAR controls regional splicing patterns of *Neurexin* pre-mRNAs in the brain. *PLOS Genet.* 9:e1003474. doi: 10.1371/journal.pgen.1003474

Ehrmann, I., and Elliott, D. J. (2010). "Expression and functions of the star proteins Sam68 and T-STAR in mammalian spermatogenesis," in *Post-Transcriptional Regulation by STAR Proteins*, ed. T. Volk (New York, NY: Springer Science & Business Media), 67–81. doi: 10.1007/978-1-4419-7005-3_5

Ehrmann, I., Fort, P., and Elliott, D. J. (2016). STARs in the CNS. *Biochem. Soc. Trans.* 44, 1066–1072. doi: 10.1042/BST20160084

Fan, C., Chen, Y., and Long, M. (2008). Recurrent tandem gene duplication gave rise to functionally divergent genes in *Drosophila*. *Mol. Biol. Evol.* 25, 1451–1458. doi: 10.1093/molbev/msn089

Frisone, P., Pradella, D., Di Matteo, A., Belloni, E., Ghigna, C., and Paronetto, M. P. (2015). SAM68: signal transduction and RNA metabolism in human cancer. *Biomed Res. Int.* 2015:528954. doi: 10.1155/2015/528954

Hensley, M. R., Cui, Z., Chua, R. F., Simpson, S., Shammas, N. L., Yang, J. Y., et al. (2016). Evolutionary and developmental analysis reveals KANK genes were co-opted for vertebrate vascular development. *Sci. Rep.* 6:27816. doi: 10.1038/srep27816

Hoegg, S., Brinkmann, H., Taylor, J. S., and Meyer, A. (2004). Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. *J. Mol. Evol.* 59, 190–203. doi: 10.1007/s00239-004-2613-z

Iijima, T., Iijima, Y., Witte, H., and Scheiffele, P. (2014). Neuronal cell type-specific alternative splicing is regulated by the KH domain protein SLM1. *J. Cell Biol.* 204, 331–342. doi: 10.1083/jcb.201310136

Jaillon, O., Aury, J. M., Brunet, F., Petit, J. L., Stange-Thomann, N., Mauceli, E., et al. (2004). Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 431, 946–957. doi: 10.1038/nature03025

Jones, A. R., and Schedl, T. (1995). Mutations in gld-1, a female germ cell-specific tumor suppressor gene in *Caenorhabditis elegans*, affect a conserved domain also found in Src-associated protein Sam68. *Genes Dev.* 9, 1491–1504. doi: 10.1101/gad.9.12.1491

Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B., and Schilling, T. F. (1995). Stages of embryonic development of the zebrafish. *Dev. Dyn.* 203, 253–310. doi: 10.1002/aja.1002030302

Koscik, T. R., and Tranel, D. (2012). Brain evolution and human neuropsychology: the inferential brain hypothesis. *J. Int. Neuropsychol. Soc.* 18, 394–401. doi: 10.1017/S1355617712000264

Lee, M. H., and Schedl, T. (2001). Identification of in vivo mRNA targets of GLD-1, a maxi-KH motif containing protein required for *C. elegans* germ cell development. *Genes Dev.* 15, 2408–2420. doi: 10.1101/gad.915901

Lee, M. H., and Schedl, T. (2010). "C. elegans star proteins, GLD-1 and ASD-2, regulate specific RNA targets to control development," in *Post-Transcriptional Regulation by STAR Proteins*, ed. T. Volk (New York, NY: Springer Science & Business Media), 106–122.

Lukong, K. E., and Richard, S. (2003). Sam68, the KH domain-containing superSTAR. *Biochim. Biophys. Acta* 1653, 73–86. doi: 10.1016/j.bbcan.2003.09.001

Manzanares, M., Wada, H., Itasaki, N., Trainor, P. A., Krumlauf, R., and Holland, P. W. (2000). Conservation and elaboration of Hox gene regulation during evolution of the vertebrate head. *Nature* 408, 854–857. doi: 10.1038/35048570

Murakami, Y., Uchida, K., Rijli, F. M., and Kuratani, S. (2005). Evolution of the brain developmental plan: insights from agnathans. *Dev. Biol.* 280, 249–259. doi: 10.1016/j.ydbio.2005.02.008

Ohno, S. (1970). *Evolution by Gene Duplication*. New York, NY: Springer Science and Business Media. doi: 10.1007/978-3-642-86659-3

Paronetto, M. P., Messina, V., Bianchi, E., Barchi, M., Vogel, G., Moretti, C., et al. (2009). Sam68 regulates translation of target mRNAs in male germ cells, necessary for mouse spermatogenesis. *J. Cell Biol.* 185, 235–249. doi: 10.1083/jcb.200811138

Putnam, N. H., Butts, T., Ferrier, D. E., Furlong, R. F., Hellsten, U., Kawashima, T., et al. (2008). The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453, 1064–1071. doi: 10.1038/nature06967

Šestak, M. S., and Domazet-Lošo, T. (2015). Phylostratigraphic profiles in zebrafish uncover chordate origins of the vertebrate brain. *Mol. Biol. Evol.* 32, 299–312. doi: 10.1093/molbev/msu319

Shimeld, S. M., and Holland, P. W. (2000). Vertebrate innovations. *Proc. Natl. Acad. Sci. U.S.A.* 97, 4449–4452. doi: 10.1073/pnas.97.9.4449

Sugahara, F., Murakami, Y., Adachi, N., and Kuratani, S. (2013). Evolution of the regionalization and patterning of the vertebrate telencephalon: what can we learn from cyclostomes? *Curr. Opin. Genet. Dev.* 23, 475–483. doi: 10.1016/j.gde.2013.02.008

Sugahara, F., Pascual-Anaya, J., Oisi, Y., Kuraku, S., Aota, S. I., Adachi, N., et al. (2016). Evidence from cyclostomes for complex regionalization of the ancestral vertebrate brain. *Nature* 531, 97–100. doi: 10.1038/nature16518

Thisse, C., and Thisse, B. (2008). High-resolution in situ hybridization to whole-mount zebrafish embryos. *Nat. Protoc.* 3, 59–69. doi: 10.1038/nprot.2007.514

Traunmüller, L., Gomez, A. M., Nguyen, T. M., and Scheiffele, P. (2016). Control of neuronal synapse specification by a highly dedicated alternative splicing program. *Science* 352, 982–986. doi: 10.1126/science.aaf2397

Tropepe, V., and Sive, H. L. (2003). Can zebrafish be used as a model to study the neurodevelopmental causes of autism? *Genes Brain Behav.* 2, 268–281. doi: 10.1034/j.1601-183X.2003.00038.x

True, J. R., and Carroll, S. B. (2002). Gene co-option in physiological and morphological evolution. *Annu. Rev. Cell Dev. Biol.* 18, 53–80. doi: 10.1146/annurev.cellbio.18.020402.140619

Volk, T. (2010). "Drosophila star proteins," in *Post-Transcriptional Regulation by STAR Proteins*, ed. T. Volk (New York, NY: Springer Science and Business Media), 93–105. doi: 10.1007/978-1-4419-7005-3_7

Wada, H., Okuyama, M., Satoh, N., and Zhang, S. (2006). Molecular evolution of fibrillar collagen in chordates, with implications for the evolution of vertebrate skeletons and chordate phylogeny. *Evol. Dev.* 8, 370–377. doi: 10.1111/j.1525-142X.2006.00109.x

Xu, Y., Zhu, S. W., and Li, Q. W. (2016). Lamprey: a model for vertebrate evolutionary research. *Zool. Res.* 37, 263–269. doi: 10.13918/j.issn.2095-8137.2016.5.263

Zhang, G., Miyamoto, M. M., and Cohn, M. J. (2006). Lamprey type II collagen and *Sox9* reveal an ancient origin of the vertebrate collagenous skeleton. *Proc. Natl. Acad. Sci. U.S.A.* 103, 3180–3185. doi: 10.1073/pnas.0508313103

Zhang, L. Y., Zeng, M., Chen, P., Sun, H. Q., Tao, D. C., Liu, Y. Q., et al. (2009). Identification of messenger RNA substrates for mouse T-STAR. *Biochemistry* 74, 1270–1277. doi: 10.1134/S0006297909110145

Zhou, Q., Zhang, G., Zhang, Y., Xu, S., Zhao, R., Zhan, Z., et al. (2008). On the origin of new genes in *Drosophila. Genome Res.* 18, 1446–1455. doi: 10.1101/gr.076588.108