Comparative Biochemistry and Physiology, Part B xxx (2016) xxx–xxx

Contents lists available at ScienceDirect

Comparative Biochemistry and Physiology, Part B

journal homepage: www.elsevier.com/locate/cbpb

Genome-wide identification, evolution of chromobox family genes and their expression in Nile tilapia

Xing-Yong Liu, Xian-Bo Zhang, Ming-Hui Li, Shu-Qing Zheng, Zhi-Long Liu, Yun-Ying Cheng, De-Shou Wang *

Key Laboratory of Freshwater Fish Reproduction and Development (Ministry of Education), Key Laboratory of Aquatic Science of Chongqing, School of Life Sciences, Southwest University, 400715 Chongqing, PR China

ARTICLE INFO

Article history:
Received 2 May 2016
Received in revised form 18 August 2016
Accepted 2 September 2016
Available online xxxx

Keywords:
Chromobox family
Evolution
Tilapia
Temporal and spatial expression profiles
Gonad

ABSTRACT

Chromobox (Cbx) family proteins are transcriptional repressors that involved in epigenetic and developmental processes. In this study, comprehensive analyses of *Cbxs* were performed using available genome databases from representative animal species. The *Cbx* family were originated from one Polycomb (Pc) gene like the yeast Pc, which duplicated into two and gave rise to the Pc and the Heterochromatin protein 1 (Hp1) identified in invertebrates from protozoon to lancelet. Rapid expansion of *Cbx* family members was observed in vertebrates as ~8 (5 Pc and 3 Hp1) were identified in spotted gar, coelacanth and tetrapods. Further expansion of the members to ~14 (9 Pc and 5 Hp1) was observed in teleosts due to the third round genome duplication (3R). Based on transcriptome data from eight adult tilapia tissues, most of the *Cbxs* were found to be dominantly expressed in the brain, testis, ovary and heart. Analyses of the gonadal transcriptome data from four developmental stages revealed that all *Cbxs* were expressed in both ovary and testis except *Cbx7b*, with significant increase of the total and average RPKM from 5 to 90 dah (days after hatching). By in situ hybridization, the three most highly and sexual dimorphically expressed *Cbx* genes in gonads, *Cbx1b*, *Cbx3a* and *Cbx5*, were found to be expressed in phase I and II oocytes of the ovary, and in secondary spermatocytes (*Cbx1b* and *Cbx3a*) and spermatids (*Cbx5*) of the testis. Our results revealed the evolution of *Cbx* genes and indicated a potential role of *Cbxs* in epigenetic regulation of gametogenesis.

1. Introduction

Chromobox (*Cbx*) family members are transcription repressors which keep the repressive state of target genes (Ma et al., 2014). All Cbx family members have a conserved N-terminal Chromodomain which is a three beta strands and a helix containing domain present in proteins that are involved in chromatin organization. They are further divided into two groups in mammals: 1) CBX1, CBX3 and CBX5, also known as Heterochromatin protein 1β (Hp1β), Hp1γ and Hp1α, respectively, all having a C-terminal Chromo-shadow-domain; 2) CBX2, CBX4, CBX6, CBX7 and CBX8, all having a C-terminal Polycomb Repressor (PcR) box, also known as Polycomb group (PcG) proteins (Wotton and Merrill, 2007). The Chromodomain, Chromo-shadow-domain and PcR box are relatively conserved from fruit fly (*Drosophila melanogaster*) to human (*Homo sapiens*) including teleosts (Senthilkumar and Mishra, 2009). In general, CBX proteins physically interact with tri-methylated histone via their Chromodomains, Chromo-shadow-domain and the PcR box are responsible for the repressive role of CBX proteins (Aasland and Stewart, 1995; Muller and Verrijzer, 2009). Chromodomain of fruit fly PcG exhibits preferential binding to H3K27me3 whereas Chromodomain of Hp1 recognizes H3K9Me3 mark (Fischle et al., 2003). Mammalian PcG members differentially bind to methylated histone, CBX2 and CBX7 bind to both H3K9Me3 and H3K27Me3 whereas CBX4 shows strong affinity for H3K9Me3 (Bernstein et al., 2006a, 2006b).

Several lines of studies indicated that all *Cbx* genes are involved in the regulation of heterochromatin, gene expression, and developmental programs (Fischle et al., 2003; Ren and Kerppola, 2011). PcG proteins were widely recognized in mammals for their roles in a variety of biological processes, such as cell cycle control, cell fate decision, X-inactivation and epigenetic regulation (Aloia et al., 2013; Dietrich et al., 2007; Bernstein et al., 2006a, 2006b; Plath et al., 2004). Additionally, evidences also suggested that *Cbxs* involved in sex determination and differentiation, and gonadal development in mammals (Katoh-Fukui et al., 1998; Biason-Lauber et al., 2009). Targeted deletion of *Cbx2* in mouse (*Mus musculus*) resulted in homeotic transformations of the axial skeleton, growth retardation and male-to-female sex reversal (Baumann and de la Fuente, 2011). Two point mutations (P98L and R443P) in the *CBX2* gene resulted in normal bilateral ovaries and female

* Corresponding author at: Key Laboratory of Freshwater Fish Reproduction and Development (Ministry of Education), Key Laboratory of Aquatic Science of Chongqing, School of Life Sciences, Southwest University, Chongqing, China.
E-mail address: wdeshou@swu.edu.cn (D.-S. Wang).

http://dx.doi.org/10.1016/j.cbpb.2016.09.001
1096-4959/

phenotype (with internal and external genitalia) in a girl with 46, XY karyotype (Biason-Lauber et al., 2009). *Cbx3* is known to be required for embryonic kidney development and mitotic cell division in cells, normal ovarian response in women (Dihazi et al., 2015; Leonard et al., 2015; Tsui et al., 2014). However, these studies have so far been restricted to expression or function of single *Cbx* in limited tissues or at limited stages of development. Furthermore, the expression profiles of the majority of *Cbx* members in different tissues and at different stages of development in teleosts are still unknown. In addition, much efforts have been devoted to human, mouse and fruit fly, while it remains unclear if PcGs and Hp1s originated from the same ancestor gene and evolved independently in the animal kingdom.

Extensive studies on the repertoires of PcG proteins have indicated increasing *Cbx* members from tetrapods to teleosts (Wotton and Merrill, 2007). The whole genome duplication (WGD) is considered to be the driving force behind the expansion of gene families. The ancestral genomes of all teleosts underwent two older WGD events common to all vertebrates (Dehal and Boore, 2005; Van de Peer et al., 2009) and a fish specific WGD, termed the third round genome duplication (3R) (Hoegg et al., 2004; Meyer and Van de Peer, 2005). Some teleosts, like rainbow trout (*Oncorhynchus mykiss*) and common carp (*Cyprinus carpio*), even underwent a fourth round of genome duplication (4R) (Berthelot et al., 2014; Xu et al., 2014). However, comprehensive analyses are deficient on the number and evolution of *Cbx* family members in vertebrates, especially in teleosts. With the significant improvement in genome sequencing quality and bioinformatics methods, we speculate that more *Cbxs* could be isolated from non-mammalian vertebrates than reported. Additionally, genome-wide investigations of invertebrate *Cbxs* are still lacking. On the other hand, recently, genomes of more and more fishes, such as the elephant shark (*Callorhinchus milii*, a chondrichthyan) (Venkatesh et al., 2014), coelacanth (*Latimeria chalumnae*, an early sarcopterygian) (Amemiya et al., 2013), spotted gar (*Lepisosteus oculatus*, a non-teleost actinopterygian) (Amores et al., 2011; Braasch et al., 2016), common carp (a teleost underwent 4R) (Xu et al., 2014), have been sequenced and published. The available genome sequences of the species mentioned above provide new resources to understand the evolution of this family.

The Nile tilapia (*Oreochromis niloticus*) is an important farmed fish with an XX-XY sex determination system. The availability of the whole genome sequence of tilapia and tissue transcriptomes, together with its gonadal transcriptomes at different developmental stages (Brawand et al., 2014; Tao et al., 2013), made it an excellent model for genome-wide identification, tissue distribution and gonadal expression profiles investigation for *Cbx* family.

Given the significance of *Cbxs* in diverse biological processes, as well as our main interests in sex determination and differentiation in fish, here we report the first genome-wide identification of *Cbxs* from representative animal species, their chromosomal location, phylogeny, synteny and spatial and temporal expression profiles in Nile tilapia. Our results provided a framework of expanding process of *Cbx* genes during evolution and certainly contributed to the understanding of their roles in teleosts.

2. Materials and Methods

2.1. Identification of Cbxs from representative animal species

The Chromodomain is highly conserved in *Cbx* genes among different species. To insure the identification of all *Cbx* genes in each genome analyzed, the Chromodomain of zebrafish (*Danio rerio*) *Cbxs* was used as the query sequence to blast the genome sequences by tblastn ($E = 2e^{-5}$). The identified *Cbxs* were used to back search against the NCBI by blastx to reduce redundant matches. Cbx sequences of the yeast (*Saccharomyces cerevisiae*), trichomonad (*Trichomonas vaginalis*), nematode (*Caenorhabditis elegans*), polypus (*Hydra vulgaris*), bombyx (*Bombyx mori*), florida lancelet (*Branchiostoma floridae*), lamprey (*Lampetra japonica*), elephant shark, spotted gar, coelacanth, tilapia, zebrafish, stickleback (*Gasterosteus aculeatus*), medaka (*Oryzias latipes*), fugu (*Takifugu rubripes*), tetraodon (*Tetraodon nigroviridis*), xenopus (*Xenopus tropicalis*), lizard (*Anolis carolinensis*), chicken (*Gallus gallus*), mouse and human were collected from NCBI and Ensembl database. The genome sequence of common carp was downloaded from (http://www.carpbase.org/download_home.php). To avoid confusion, we kept their names in accordance with those in the zebrafish (Bertrand et al., 2007).

2.2. Phylogenetic analyses and genomic distribution of Cbxs

The multiple alignment software Bioedit was employed to align the Chromodomain of Cbx from all species analyzed. Then the Neighbour-joining (NJ) trees were constructed based on p-distances, with bootstrap 1000 replicates (Felsenstein, 1985) using the program MEGA version 5 (Tamura et al., 2011). Genomic distributions of *Cbxs* were performed using UCSC Blat search (http://www.genome.ucsc.edu/blat). The syntenic block was identified as described previously (Zhang et al., 2014).

2.3. Transcriptome analyses of expression profile of Cbxs in Nile tilapia

The brain, heart, liver, ovary, testis, kidney and muscle Illumina RNA-seq data of adult tilapia were downloaded from the NCBI Sequence Read Archive (SRA) (Accession codes were SRR391697, SRR391681, SRR391688, SRR391687, SRR391690, SRR391684 and SRR391702, respectively) (Brawand et al., 2014). The head kidney transcriptome was sequenced in our previous study (Cheng et al., 2015). *Cbxs* with the total RPKM < 10 in all eight tissues were considered as background expression.

Four pairs of RNA preparations from gonads of XX and XY tilapia at 5, 30, 90 and 180 dah (days after hatching) (corresponding to molecular sex determination and differentiation, the initiation of germ cell meiosis in the XX gonads, the initiation of germ cell meiosis in the XY gonads, vitellogenesis in the XX gonads or sperm maturation in the XY gonads, respectively) were sequenced using Illumina 2000 HiSeq technology in our previous study (Tao et al., 2013). The statistical criteria of "FDR (false discovery rate) ≤ 10^{-2}" and "$|\log_2(XX\_RPKM/XY\_RPKM)| \geq 1$" were used to identify XX/XY-enhanced *Cbxs* (Tao et al., 2013; Yuan et al., 2014). The *Cbxs*, which identified as XX/XY-enhanced *Cbxs* in at least three stages, were considered as XX/XY-dominant genes. The student *t*-test (Excel, Microsoft) was used to compare the expression difference of *Cbxs* between ovary and testis at the same stage (Table 2).

Table 1
The numbers of Cbx family members in some representative vertebrate species genomes, which supported 2R and 3R hypotheses. Q4

	Hp1s			PcGs					
	Cbx1	*Cbx3*	*Cbx5*	*Cbx2*	*Cbx4*	*Cbx6*	*Cbx7*	*Cbx8*	*Total*
Elephant shark	1	1	1	0	0	1	1	0	5
Spotted gar	1	2	1	1	1	0	0	1	7
Coelacanth	1	1	1	1	1	1	1	1	8
Xenopus	1	1	1	1	1	1	1	1	8
Chicken	1	1	1	1	1	1	1	1	8
Lizard	1	1	1	1	1	0	1	1	7
Mouse	1	1	1	1	1	1	1	1	8
Human	1	1	1	1	1	1	1	1	8
Stickleback	1	2	1	2	2	0	2	2	12
Medaka	2	2	1	2	2	1	2	2	14
Fugu	2	2	1	2	2	1	2	2	14
Tetraodon	2	2	1	2	2	1	2	2	14
Tilapia	2	2	1	2	2	1	2	2	14
Zebrafish	2	2	1	1	2	2	2	2	14
Common carp	3	2	2	2	3	2	2	2	18

Please cite this article as: Liu, X.-Y., et al., Genome-wide identification, evolution of chromobox family genes and their expression in Nile tilapia, Comp. Biochem. Physiol. B (2016), http://dx.doi.org/10.1016/j.cbpb.2016.09.001