Marine Genomics 1 (2008) 69–78

Contents lists available at ScienceDirect

Marine Genomics

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

Comparative analysis of teleost fish genomes reveals preservation of different ancient clock duplicates in different fishes

Han Wang *

Department of Zoology and Stephenson Research and Technology Center, University of Oklahoma, Norman, OK 73019, USA

ARTICLE INFO

Article history:
Received 17 January 2008
Received in revised form 21 June 2008
Accepted 30 June 2008

Keywords:
Circadian rhythms
Conserved synteny
dN/dS
Relative evolutionary rates
Genome duplication and differential gene loss

ABSTRACT

Clock (*Circadian locomotor output cycle kaput*) was the first vertebrate circadian clock gene identified in a mouse forward genetics mutagenesis screen. It encodes a bHLH-PAS protein that is highly conserved throughout evolution. Tetrapods also have the second *Clock* gene, *Clock2* or *Npas2* (*Neuronal PAS domain protein 2*). Conversely, the fruit fly, an invertebrate, has only one *clock* gene. Interrogation of the five teleost fish genome databases revealed that the zebrafish and the Japanese pufferfish (fugu) each have three *clock* genes, whereas the green spotted pufferfish (tetraodon), the Japanese medaka fish and the three-spine stickleback each have two *clock* genes. Phylogenetic and splice site analyses indicated that zebrafish and fugu each have two *clock1* genes, *clock1a* and *clock1b* and one *clock2*; tetraodon also have *clock1a* and *clock1b* but do not have *clock2*; and medaka and stickleback each have *clock1b* and one *clock2*. Genome neighborhood analysis further showed that *clock1a*/*clock1b* in zebrafish, fugu and tetraodon is an ancient duplicate. While the dN/dS ratios of these three fish *clock* duplicates are all <1, indicating that purifying selection has acted upon them; the Tajima relative rate test showed that all three fish *clock* duplicates have asymmetric evolutionary rates, implicating that one of these duplicates have been under positive selection or relaxed functional constraint. These results support the view that teleost fish *clock* genes were generated from an ancient genome-wide duplication, and differential gene loss after the duplication resulted in retention of different ancient duplicates in different teleost fishes, which could have contributed to the evolution of the distinct fish circadian clock mechanisms.

1. Introduction

Circadian rhythms display in almost all living things from cyanobacteria to humans at literally every level from molecular to organismal (Pittendrigh, 1993; Dunlap, 1999; Bell-Pedersen et al., 2005). The generation and maintenance of circadian rhythmicity are regulated and controlled by the transcription/translation negative feedback loop composed of a group of circadian clock genes and their proteins (Dunlap, 1999; Reppert and Weaver, 2001; Panda et al., 2002a,b; Bell-Pedersen et al., 2005; Hardin, 2005; Ko and Takahashi, 2006; Yu and Hardin, 2006). Forward genetics mutagenesis screens, and subsequently molecular cloning, biochemical, physiological and behavioral characterizations have been instrumental in identifying these canonical circadian clock genes (Konopka and Benzer, 1971; Bargiello et al., 1984; Reddy et al., 1984; Takahashi et al., 1994; Dunlap, 1999). *Clock* (*Circadian locomotor output cycle kaput*) identified as the first vertebrate circadian clock gene in a mouse (*Mus musculus*) forward genetics mutagenesis screen encodes a bHLH (basic helix-loop-helix)-PAS (Per-Arnt-Sim) protein (Vitaterna et al., 1994; Antoch et al., 1997; King et al., 1997). The second copy of *Clock* gene, *i.e.*, *Clock2* or *Npas2* (Neuronal PAS domain protein 2), also exists in tetrapod genomes (King et al., 1997; Rutter et al., 2001). On the other hand, the fruit fly (*Drosophila melanogaster*), an invertebrate, possesses only one *clock* gene (Allada et al., 1998). In fly and mouse, CLOCK and CYCLE in fly or BMAL1 (Brain and muscle ARNT like protein 1) in mouse, another bHLH-PAS protein, form the CLOCK:CYCLE or CLOCK:BMAL1 heterodimer through their PAS domains (Allada et al., 1998; Darlington et al., 1998; Hogenesch et al., 1998; Dunlap, 1999; Panda et al., 2002a,b; Hardin, 2005; Ko and Takahashi, 2006; Yu and Hardin, 2006). Through binding the E-boxes (CACGTG) in the regulatory regions of *Per* (*Period*), *tim* (*timeless*), *Cry* (*Cryptochrome*) and other clock-controlled genes (CCGs), the CLOCK:CYCLE or CLOCK:BMAL1 heterodimer drives transcription of these genes (Allada et al., 1998; Darlington et al., 1998; Hogenesch et al., 1998; Panda et al., 2002a,b; Hardin, 2005; Ko and Takahashi, 2006; Wijnen et al., 2006; Yu and Hardin, 2006). Thus the molecular machinery underlying the time-keeping mechanisms has been remarkably conserved throughout

Abbreviations: Clock, *Circadian locomotor output cycle kaput*; *Bmal1*, *Brain and muscle ARNT like protein 1*; bHLH, basic Helix-loop-Helix; PAS, Period-Aryl hydrocarbon receptor nuclear translocator-Single mind; *Per*, *Period*, *tim*, *timeless*; *Cry*, *Cryptochrome*; CCGs, clock-controlled genes; SCN, suprachiasmatic nuclei; dN, the numbers of nonsynonymous nucleotide substitutions per nonsynonymous site; and dS, the numbers of synonymous nucleotide substitutions per synonymous site.

* Tel.: +1 405 325 9324; fax: +1 405 325 3442.
E-mail address: hwang@ou.edu.

1874-7787/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.margen.2008.06.003

evolution (Dunlap, 1999; Panda et al., 2002a,b; Bell-Pedersen et al., 2005; Yu and Hardin, 2006).

Teleost fishes, a group of vertebrates, display enormous biodiversity in numbers of species, morphology and behaviors (Nelson, 2006). Studies of developmentally important genes such as *Hox* genes suggest that gene and genome duplication likely contribute to the explosive radiation of teleost fishes by generating the teleost morphological complexity (Amores et al., 1998; Meyer and Schartl, 1999; Postlethwait, 2007). However, the magnificent teleost biodiversity also includes diverse behaviors such as mating, egg-laying and migration, which are all temporally regulated. An understanding of the evolution of teleost circadian clock genes and the teleost circadian clock mechanisms should help unravel the evolutionary mechanisms underlying the tremendous teleost explosion (Wang, 2008). There are three *clock* genes that have been identified and studied in zebrafish (*Danio rerio*), a freshwater teleost fish (Whitmore et al., 1998; Ishikawa et al., 2002). Screening an embryonic cDNA library using the mouse *Clock* cDNA probe identified a homologous *clock* gene as the first zebrafish circadian clock gene (Whitmore et al., 1998). An additional two zebrafish *clock* genes were isolated by RT-PCR (Reverse Transcription-Polymerase Chain Reaction) with degenerate primers (Ishikawa et al., 2002). However, no *clock* genes from the Japanese pufferfish (fugu) (*Takifugu rubripes*), the spotted green pufferfish (tetraodon) (*Tetraodon nigroviridis*), the Japanese medaka fish (*Oryzias latipes*) or the three-spine stickleback (*Gasterosteus aculeatus*) have been studied so far.

Recent releases of the genome sequences of zebrafish, fugu (Aparicio et al., 2002), tetraodon (Jaillon et al., 2004), medaka (Kasahara et al., 2007), and stickleback allow for comparative genomic analysis of circadian clock genes (Wang, 2008). Previously, I used a combination of phylogenetic, splice site and syntenic analyses to examine the evolution of teleost fish *period* (*per*), another canonical circadian clock gene (Wang, 2008). Zebrafish have two *per1* genes, *per1a* and *per1b*, one *per2* and one *per3*; medaka, pufferfish and tetraodon each have two *per2* genes, *per2a* and *per2b*, one *per1* and one *per3*; and stickleback have *per2a*, *per2b*, and one *per1* but do not have *per3* (Wang, 2008). Zebrafish have preserved the *per1a*/*per1b* ancient duplicate, whereas medaka, fugu, tetraodon and stickleback have maintained the *per2a*/*per2b* ancient duplicate (Wang, 2008). Here I study the evolutionary history of *clock* genes in teleost fish genomes. How many *clock* genes in these five teleost fish genomes? How were the extra copies of fish *clock* genes generated? Are there different *clock* duplicates in different fishes? How have fish *clock* duplicates evolved? Careful investigation of these questions not only helps us obtain an understanding of the evolution of teleost fish *clock* genes but also provides insights into the evolution of the teleost fish circadian clock mechanisms that could have played a role in the remarkable teleost radiation.

2. Materials and methods

2.1. Phylogenetic analysis

The 12 teleost fish *clock* genes (Table 1) were obtained from Ensembl (Flicek et al., 2008) (http://www.ensembl.org/index.html) as of January, 2008. Multiple sequence alignments of CLOCK peptides were generated using ClustalX (Thompson et al., 1997). The phylogenetic tree of CLOCK proteins (Fig. 1) was constructed using the neighbor-joining method with MEGA4 (Tamura et al., 2007).

2.2. Splice site analysis

Exon boundaries of the coding regions of these 12 teleost fish *clock* genes as well as human, mouse, chicken, and fly *Clock* genes were determined according to Ensembl. The numbers of nucleotides (nt) for each exon as well as the phase of each slicing site also were determined and shown in Figs. 2 and 3, respectively.

2.3. Genome neighborhood analysis

Human *CLOCK1* (4q12, 55.99 mb) and *CLOCK2* (*NPAS2*, 2q311.2, 100.8 mb) (NCBI Build 36. 2) were used as anchor sites. The orthologous comparisons of the genes in the regions of approximately 10 to 30 mb flanking the human *CLOCK* loci with the mouse genome (NCBI Build 37.1), the chicken genome (WASHUC2), the zebrafish genome (Ensembl Zv7), the fugu genome (FUGU 4.0), the tetraodon genome (TETRAODON 7), the medaka genome (MEDAKA1) or the stickleback genome (BROAD S1) were done with the BioMart mode (Kasprzyk et al., 2004) in Ensembl,

Table 1
Clock genes in *Danio rerio*, *Takifugu rubripes*, *Tetraodon nigroviridis*, *Oryzias latipes* and *Gasterosteus aculeatus*[a]

Species	Gene name	Ensembl Gene ID	Number of exons	Transcript length (bps)	Peptide length (aa)	Genome location
Danio rerio	*clock1a*	ENSDARG00000011703	20	2688	895	Chromosome 20: 20,828,006-20,872,003
	clock1b	ENSDARG00000003631	22	2547	848	Chromosome 1: 15,317,208-15,347,590
	clock2	ENSDARG00000016536	21	2538	845	Chromosome 3: 6,544,543-6,583,353
Takifugu rubripes	*clock1a*	SINFRUG00000126258	21	2523	841	Scaffold_13: 1,587,632-1,594,017
	clock1b	SINFRUG00000142142	19	2433	811	Scaffold_563: 29,065-36,831
	clock2	SINFRUG00000141792	14	1497	499	Scaffold_6: 1,324,869-1,331,068
Tetraodon nigroviridis	clock1a	GSTENG00032758001	19	2658	885	Chromosome 1_random: 2,133,062-2,138,682
	clock1b	GSTENG00006765001	17	2499	832	Chromosome 18: 3,385,333-3,395,579
Oryzias latipes	clock1b	ENSORLG00000004495	20	2646	882	Chromosome 1: 14,996,851-15,007,644
	clock2	ENSORLG00000009426	23	2562	853	Chromosome 14: 17,135,493-17,147,943
Gasterosteus aculeatus	*clock1b*	ENSGACG00000015939	25	2490	830	GroupIXcontig_8223: 488,493-499,446
	clock2	ENSGACG00000020338	22	2469	823	GroupVIIcontig_581: 17,207,286-17,217,881

[a]Note: the information of these 12 teleost fish *clock* genes is from Ensemble (http://www.ensembl.org/index.html) as of January, 2008.