

# Duplicated *Clock* genes with unique polyglutamine domains provide evidence for nonhomologous recombination in Chinook salmon (*Oncorhynchus tshawytscha*)

K. G. O'Malley · M. A. Banks

Received: 30 September 2006 / Accepted: 9 April 2007  
© Springer Science+Business Media B.V. 2007

**Abstract** Circadian rhythms underlie diverse life functions ranging from cellular activities to behavior. Multiple clock genes play a central role in the generation of these rhythms. We partially characterized two copies of the *Clock* gene from Chinook salmon (*Oncorhynchus tshawytscha*), *OtsClock1a* and *OtsClock1b*. The 6,460 bp *OtsClock1a* sequence contains 16 exons, 15 introns and encompasses three highly conserved domains indicating it is a novel member of the bHLH-PAS superfamily of transcription factors. The second copy, *OtsClock1b*, consists of five exons and five introns spanning 1,945 bp. A polyglutamine repeat motif (PolyQ), characteristic of a majority of CLOCK proteins, is present in both OTS-CLOCK1a and OTSCLOCK1b. However, the Chinook PolyQ domains are uniquely positioned inside the gene. Interestingly, a 1,200 bp non-coding segment located downstream of the *OtsClock1a* PolyQ domain is absent from *OtsClock1b*. This insertion/deletion is 91% similar to the *Salmo salar Transferrin* gene. A phylogenetic analysis of 11 CLOCK proteins shows that *OtsClock1a* and *OtsClock1b* are paralogs which likely arose subsequent to the salmonid genome-wide duplication event. Ultimately, the Chinook salmon *Clock* genes are key components to our understanding the genetic mechanisms underlying temporally regulated life history traits in Pacific salmonids.

**Keywords** BAC library · Circadian rhythms · Gene duplication · PAS domains · PolyQ domain · Salmonids

## Introduction

Temporal organization of biological processes is a critical feature of life. Circadian rhythms underlying these processes are recognized as adaptations to cyclically changing environments (Schwassmann 1988). These rhythms are the external expression of an endogenous clock that initiates temporal coordination of physiological and biochemical functions enabling organisms to anticipate and respond to environmental changes during the day-night cycle.

The molecular mechanisms underlying circadian rhythms have been characterized in several model systems including Cyanobacterium, *Neurospora*, *Drosophila*, mouse, and zebrafish. There is remarkable conservation of function in this circadian molecular machinery among these systems. In mammals and fishes, for example, several canonical clock proteins including CLOCK, BMAL, PERIOD (PER), and CRYPTOCHROME (CRY) are critical in maintaining an autoregulatory feedback loop that generates circadian behavior. CLOCK and BMAL heterodimerize to form a transcription factor which activates the transcription of *Period*, *Cryptochrome* and *Rev-Erb $\alpha$*  genes. As the levels of PER proteins increase, they form a complex with CRY and *CKI $\epsilon$ /CKI $\delta$*  proteins and translocate to the nucleus. The CRY-PER-CKI $\epsilon$ /CKI $\delta$  complexes then associate with CLOCK-BMAL heterodimers and shut down transcription, forming the negative feedback loop. For the positive feedback loop, increasing REV-ERB $\alpha$  levels act to repress *Bmal* transcription. CRY-mediated inhibition of CLOCK-BMAL-mediated transcription activates *Bmal* transcription, because REV-ERB $\alpha$ -mediated repression is inhibited (Reppert and Weaver 2002). This autoregulatory feedback loop results in the cyclic expression of clock genes throughout an approximately 24 h period.

K. G. O'Malley (✉) · M. A. Banks  
Coastal Oregon Marine Experiment Station, Hatfield Marine  
Science Center, Department of Fisheries and Wildlife, Oregon  
State University, 2030 SE Marine Science Drive, Newport, OR  
97365, USA  
e-mail: kathleen.omalley@oregonstate.edu

Major gene and genome duplication events have generated diversity in both copy number and function of the core clock genes. In the mammalian system, there are two copies of both *Clock* (*mClock* and *mNpas2*) and *Bmal* (*mBmal1* and *mBmal2*), four *Period* genes (*mPer 1–4*), and two *Cryptochrome* genes (*mCry 1–2*) (reviewed in Looby and Loudon 2005). In contrast, zebrafish have three copies of *Clock* (*zfClock1*, *zfNpas2*, and *zfClock3*), three of *Bmal* (*zfBmal 1–3*), four of *Period* (*zfPer 1–4*) and six of *Cryptochrome* (*zfCry 1–6*) (reviewed in Hirayama et al. 2005).

The duplication of genes and their subsequent functional divergence is a fundamental process of adaptive evolution (Ohno et al. 1968; Kimura and Ohta 1974). Two potential mechanisms that lead to the formation of evolutionarily related but functionally distinct gene families include positive Darwinian selection (Hughes 1994) and subfunctionalization (Force et al. 1999). Subfunctionalization involves the partitioning of the ancestral function between gene duplicates. For example, McNamara et al. (2001) showed that the mammalian ortholog of *Clock*, *mNpas2*, is capable of forming a dimer with *mBmal1* to activate gene expression, however, the tissue distribution of *mNpas2* does not overlap with that of *mClock*. Furthermore, it appears that while *mClock* is light entrainable, *mNpas2* is a stimulus-driven oscillator.

Polyploidy is the primary mechanism for generating genomic redundancy, as no other process can produce a comparable increase of genetic material on which selection may act. Considerable evidence indicates that all extant salmonid fish species evolved from a single autotetraploidization event approximately 25–100 MYA (Allendorf and Thorgaard 1984). Approximately 50% of protein loci examined in salmonid fishes continue to exhibit duplicate gene expression (Bailey et al. 1978; Allendorf and Thorgaard 1984). Thus, considering both their genomic complexity and diverse life histories, salmonid fishes provide a compelling system to study the evolution of clock genes.

The *Clock* gene codes for one of the most essential proteins of the circadian system (Antoch et al. 1997). *Clock* has been characterized in several organisms, including mammals (Antoch et al. 1997; Avivi et al. 2001; King et al. 1997; Steeves et al. 1999), insects (Allada et al. 1998; Chang et al. 2003; Darlington et al. 1998), birds (Yoshimura et al. 2000), and fishes (Mazurais et al. 2000; Whitmore et al. 1998). As a member of the basic-helix-loop-helix (bHLH)/PER-ARNT-SIM (PAS) superfamily of transcription factors, CLOCK has a basic DNA binding domain (bHLH) and two protein dimerization domains (PAS-A and PAS-B). All three domains are highly conserved across a wide range of taxa. The majority of CLOCK proteins also have a carboxyl-terminal polyglutamine repeat motif (PolyQ) that corresponds to the

transactivation domain (Darlington et al. 1998). Here, we report our findings from the isolation and characterization of two Chinook salmon (*Oncorhynchus tshawytscha*) *Clock* genes, *OtsClock1a* and *OtsClock1b*, demonstrating that gene organization is different. Each copy contains a polyglutamine repeat motif uniquely located inside the gene. Furthermore, this single amino acid repeat, which is considerably longer in *OtsClock1b* compared to *OtsClock1a*, is positioned directly upstream of a nonhomologous recombination event. Lastly, we conducted a phylogenetic analysis to illustrate the evolutionary relationship among OTSCLOCK1a, OTSCLOCK1b and 9 CLOCK proteins from four other species.

## Material and methods

### Identification of the *Oncorhynchus tshawytscha* *Clock* gene, *OtsClock1a*

We amplified DNA sequence from a Chinook salmon *Clock* gene, *OtsClock1a*, using PCR primers designed from highly conserved regions of *Mus musculus* (GenBank accession no. AF146793), *Danio rerio* (GenBank accession no. AF133306) and *Oncorhynchus mykiss* (GenBank accession no. AF266745) *Clock* sequences. The primer sequences were 5'-TGTGCACTGTTGAAGAACCCAATGAAGAATT-3' as the sense, and 5'-GTGGCATTGTCGAAAGTCTCCAGGTCATC-3' as the antisense primer.

Genomic DNA was extracted from Chinook salmon liver using DNeasy Tissue Kit (Qiagen) and amplified in 40  $\mu$ l reactions using the following touchdown PCR profile: one initial denaturing cycle of 3 min at 94°C, followed by one cycle of 1 min at 94°C, 1 min at 62°C annealing temperature, and 1 min 30 s at 72°C. In subsequent cycles, the annealing temperature was decreased by 2°C until 56°C was reached, followed by 29 more cycles of 1 min 94°C, 1 min at 56°C, 1 min 30 s at 72°C, and a final extension of 10 min at 72°C. PCR products were excised from 1.5% agarose gels, purified using the QIAquick Gel Extraction Kit (Qiagen) and cloned into pCR4-TOPO using the TOPO TA Cloning Kit for Sequencing (Invitrogen). Plasmid DNA was isolated using Wizard<sup>®</sup> Plus SV Minipreps (Promega) and sequenced using Big Dye<sup>™</sup> Terminator v3.1 Cycle Sequencing Ready Reaction. All sequences were generated on an Applied Biosystems 3730xl DNA Analyzer and the results were viewed using BioEdit Sequence Alignment Editor (Hall 1997). To obtain additional genomic sequence both upstream and downstream of the conserved region, genome “walking” was performed multiple times using the Universal GenomeWalker<sup>™</sup> Kit (BD Biosciences) according to the manufacturer’s instructions.

## Screening a Chinook salmon BAC library

We screened a Chinook salmon bacterial artificial chromosome (BAC) library to (1) validate sequencing data obtained for *OtsClock1a* by PCR and genome “walking” and (2) isolate gene duplicates of *Clock*. The BAC library was constructed by BACPAC Resources, Children’s Hospital Oakland Research Institute (CHORI) Center with funds provided by Dr. Bob Devlin of the Department of Fisheries and Oceans, British Columbia, Canada (<http://bacpac.chori.org>). Five Hybond-XL filters, representing seven times coverage, were hybridized with a 130 bp probe designed complementary to exon 16 of *OtsClock1a*. This probe was PCR amplified using the sense primer 5′-CCA ATGGTACAGTTCTCCACC-3′ and 5′-CTCCTCCTGGA TCTGCCTG-3′ as the antisense primer. PCR product was purified and labeled with [ $\alpha$ 32P]dCTP using a rediprime II labeling system (Amersham Biosciences). Filter hybridization was performed according to the Hybond-XL protocol (Amersham Biosciences). Positive clones were identified and confirmed using PCR. BAC-clone minipreps were used as templates for PCR while high quality DNA required for direct sequencing of BAC clones was extracted using Qiagen’s Large Construct kit. BAC templates were sequenced using Big Dye<sup>TM</sup> Terminator v3.1 Cycle Sequencing Ready Reaction and an Applied Biosystems 3730xl DNA Analyzer.

## RNA extraction and reverse transcriptase PCR

To locate intron/exon splice sites, *OtsClock1a* was aligned manually to the *D. rerio Clock* gene (*zfClock1*) (GenBank accession no. AF133306). All Chinook salmon *Clock* exons are numbered based on their homology to *zfClock1* exons. Oligonucleotide primers were then designed complementary to predicted coding regions of *OtsClock1a*: 5′-GACGCCAGCAGTATCTTTGATGGGT-3′ as the sense primer and 5′-GTAGGGACCATGACCGCACCACA-3′ as the antisense primer. To amplify *OtsClock1b* mRNA transcript, we used the sense primer 5′-GGATGCCAGTCC GTCTCTGCTATTG-3′ and 5′-GGAGCGACCTGACATG CTACCTG-3′ as the antisense primer.

To obtain Chinook *Clock* mRNA, we sampled brain tissue from four juvenile Chinook salmon at the Rock Creek Fish Hatchery on the Umpqua River, Oregon. Brain tissue was preserved in RNAlater<sup>®</sup> (Ambion) and total RNA was isolated using Trizol/TriReagent V 2.2. Reverse-transcription-polymerase chain reaction (RT-PCR) was carried out using SuperScript<sup>TM</sup> One-Step RT-PCR System with Platinum<sup>®</sup> TaqDNA Polymerase (Invitrogen). We amplified the *Clock* mRNA transcripts using a modified PCR procedure consisting of a 30 min hot start at 55°C and 2 min denaturing at 94°C, one cycle of 15 s at 94°C, 30 s

at 62°C, 1 min at 68°C; after which the annealing temperature was decreased by 2°C until 58°C was reached, followed by 37 cycles of 15 s at 94°C, 30 s at 58°C, 1 min at 68°C, with a final extension of 5 min at 68°C. Products were excised from a 1.5% agarose gel, purified, cloned and sequenced (as mentioned above).

## Gene annotation and sequence comparison

Sequence similarity searches were conducted using four BLAST programs: Blastn, Blastx, tBlastx, and Blastp (Altschul et al. 1990). The Simple Modular Architecture Research Tool (SMART) was used to identify conserved domains of the predicted proteins (Schultz et al. 1998). Multiple sequence alignments were created manually using BioEdit Sequence Alignment Editor (Hall 1997) and generated automatically using ClustalW (Thompson et al. 1994).

To examine the phylogenetic relationship of CLOCK proteins, we constructed a gene tree using the ProtDist Neighbor-Joining method in the phylogeny inference package, PHYLIP (Felsenstein 1991). Nine CLOCK protein sequences were retrieved from NCBI using tBlastx and Blastp (Altschul et al. 1990). A multiple sequence alignment based on amino acid sequence from exons 15–19 was generated using ClustalW (Thompson et al. 1994). The tree was bootstrapped 1,000 times to calculate support for each node on the tree.

## Nucleotide sequence accession number

Sequence data has been deposited in GenBank under the following accession numbers: *OtsClock1a* genomic DNA (DQ780892), *OtsClock1a* mRNA (DQ780893), *OtsClock1b* genomic DNA (DQ780894), *OtsClock1b* mRNA (DQ780895).

## Results

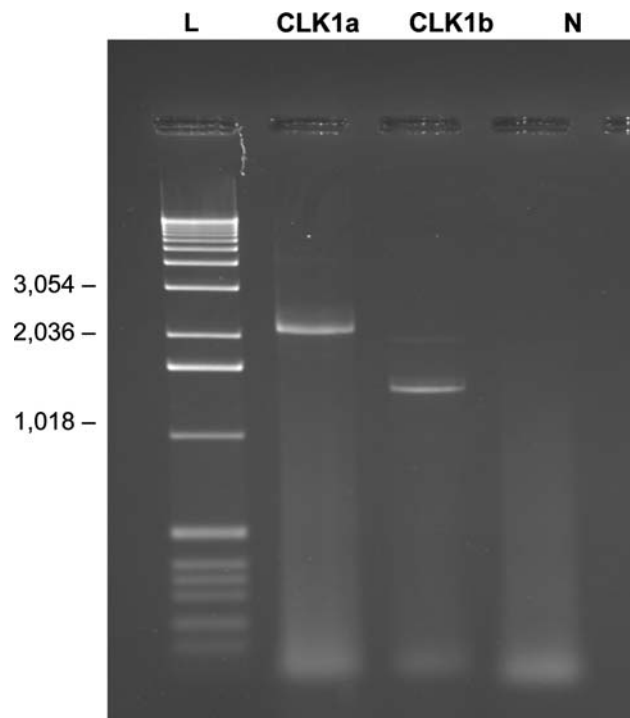
### Cloning and characterization of *OtsClock1a*

A two-step process involving PCR amplification and genome walking was employed to obtain 6,460 bp of the Chinook salmon *Clock* gene, *OtsClock1a*. The partial *OtsClock1a* sequence consists of 16 exons and 15 introns. Using reverse-transcriptase PCR, we also amplified a 2,241 bp open reading frame product that encodes a clock protein of 746 amino acids (OTSCLOCK1a) (Fig. 1). Three conserved domains, the basic Helix-Loop-Helix (bHLH) and the two Per-Arnt-Sim domains (PAS-A and PAS-B), are located in exon 3, exons 5–6, and exons 9–11, respectively, indicating that *OtsClock1a* is a member of the

bHLH-PAS superfamily of transcription factors. Additionally, we identified a putative polyglutamine domain, encoded by a stretch of 20 glutamine residues, located in exon 15.

Sequence similarity searches using Blastn (Altschul et al. 1990) revealed that *OtsClock1a* exhibits high identity to *Clock* genes from a number of organisms including a best match to the zebrafish *Clock1* gene (*zfClock1*) (GenBank accession no. AF133306). *OtsClock1a* and *zfClock1*, which is comprised of 20 exons, share an 84% nucleotide sequence identity in exon 3 through 18. Comparison of amino acid sequence revealed that OTSCLOCK1a is 95% similar to ZFCLOCK1 across the three conserved domains. However, the two proteins differ in both length and position of the polyglutamine repeat motif with the zebrafish PolyQ domain consisting of 51 glutamine repeats and localized to the C-terminal end.

One interesting feature of the *OtsClock1a* gene is an uncharacteristically long intron located between exon 15 and exon 16 (intron15/16). This 1,500 bp intron, which is three times the length of the next longest intron, is 91% similar to approximately 1,200 bp of the *Salmo salar Transferrin* gene partial 5' sequence (GenBank accession no. AF304198) (Fig. 2). This *Transferrin* sequence is a part of the 5' untranslated region (UTR) extending from -1,347 to -3,035 bp (Gately 2000).



**Fig. 1** Expression analysis of *OtsClock1a* and *OtsClock1b* mRNA transcripts in brain tissue by reverse-transcriptase PCR. L: 1 Kb DNA ladder; CLK1a: *OtsClock1a*; CLK1b: *OtsClock1b*; N: negative control

#### Identification of *OtsClock1b* by screening the Chinook salmon BAC library

When amplifying any locus from a polyploidy organism or members of multigene families from diploid organisms, the process of PCR-mediated recombination can occur and lead to the formation of in vitro chimera (Cronn et al. 2002). Therefore, to validate nucleotide sequence obtained for *OtsClock1a* and identify gene duplicates, we screened a Chinook salmon BAC library (CHORI-217). We identified a BAC clone containing the *OtsClock1a* gene and direct sequenced 6,460 bp of the insert. Our results validated the sequence data obtained for *OtsClock1a* via genome walking.

In addition, we identified a second copy of *Clock* within the Chinook salmon genome, *OtsClock1b*, and partially characterized the gene by direct sequencing 1,947 bp. Sequence similarity searches revealed that *OtsClock1b* is homologous to exon 15 through exon 19 of both *OtsClock1a* and *zfClock1*. It is important to note that not all BAC clones which hybridized with *Clock* probe were sequenced and therefore additional *Clock* homologs may be present in the Chinook salmon genome. Using reverse-transcriptase PCR, we amplified a 1,205 bp open reading frame product that encodes a 401 amino acid protein, OTSCLOCK1b (Fig. 1). Similar to OTSCLOCK1a, OTSCLOCK1b also contains a putative PolyQ domain located in exon 15, however, this glutamine-rich domain is twice the length containing 41 glutamine residues (Fig. 3a).

#### Pairwise sequence comparison of *OtsClock1a* and *OtsClock1b*

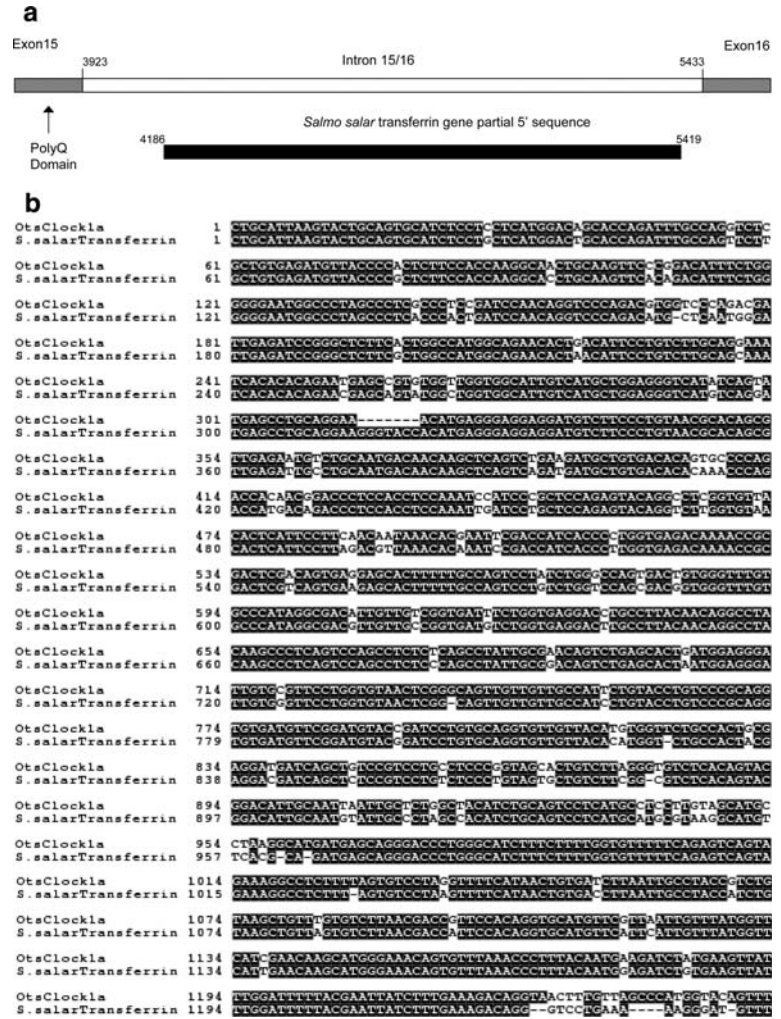
*OtsClock1a* and *OtsClock1b* share four regions of high sequence identity. The first region spans approximately 235 bp of exon 14 and exon 15 where *OtsClock1a* and *OtsClock1b* exhibit 86% sequence identity. The second, third and fourth regions span 1,050 bp of exon 16 through exon 18. The length and percent sequence identity for these three conserved sequence blocks are: 194 bp (91% identity), 570 bp (88% identity), and 92 bp (97% identity), respectively.

One notable difference between the Chinook salmon *Clock* gene sequences is a 1,200 bp insertion/deletion located in intron 15/16. As mentioned previously, this nucleotide segment shows high sequence identity with *Salmo salar Transferrin* gene partial 5' sequence and is present in *OtsClock1a* yet absent in *OtsClock1b*. Interestingly, the two PolyQ domains are located directly upstream of this indel (Fig. 3b).

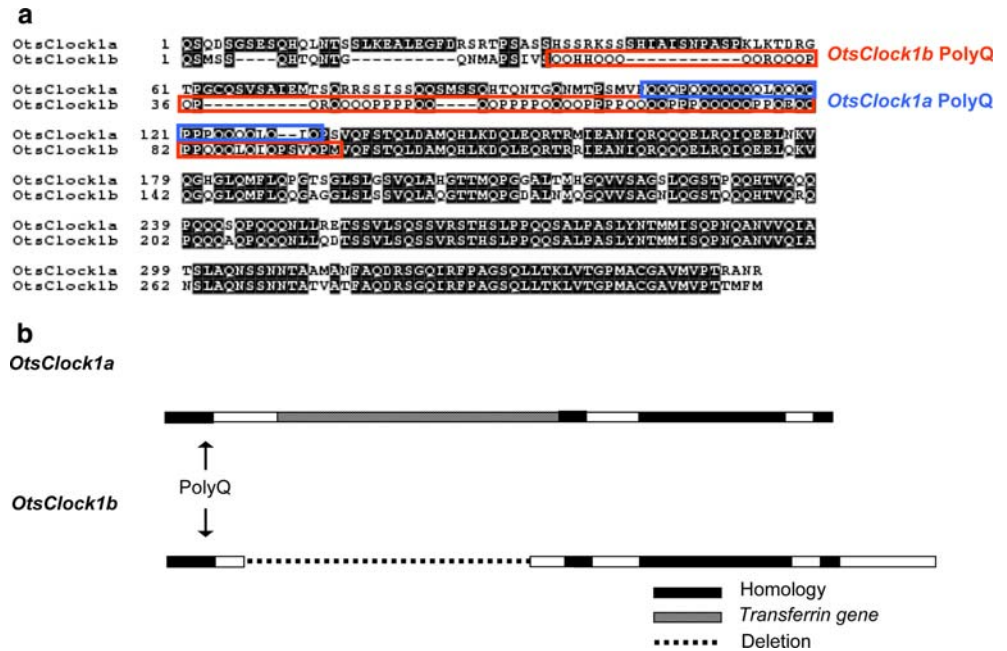
#### Inter-specific comparison of *Clock* genes

Phylogenetic analysis of the two Chinook salmon CLOCK proteins and nine CLOCK proteins from four other

**Fig. 2** (a) Schematic diagram of *OtsClock1a* gene illustrating the location of the region of high sequence identity between *OtsClock1a* and the *Salmo salar* *Transferrin* gene partial 5' sequence. Numbers represent base pair (b) Nucleotide sequence alignment of the *OtsClock1a* and *S. salar* *Transferrin* gene 5' partial sequence



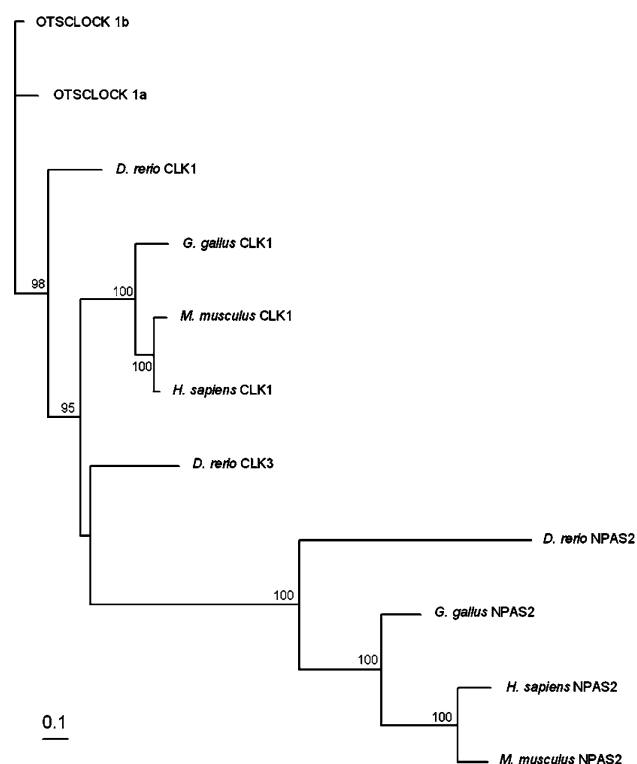
**Fig. 3** (a) Alignment of OTSCLOCK1a and OTSCLOCK1b demonstrating length variation between the two polyglutamine repeat motifs (boxed). (b) Comparison of *OtsClock1a* and *OtsClock1b* genes illustrating four regions of high sequence similarity (solid bar). *OtsClock1a* contains an approximately 1,200 bp non-coding nucleotide segment that is absent from *OtsClock1b* (dotted line). This region is homologous to the *Salmo salar* *Transferrin* gene partial 5' sequence (gray bar). A putative PolyQ domain is located upstream of this insertion/deletion in both *Clock* genes



organisms shows that OTSCLOCK1a and OTSCLOCK1b are more similar to each other than to other CLOCK orthologs. In the four other organisms, however, CLOCK orthologs are generally more closely related to each other than to their respective paralogs. For example, human CLOCK1 is more similar to *M. musculus* CLOCK1 than to its respective paralog, human NPAS2. Similarly, *D. rerio* NPAS2 clusters with *G. gallus*, *M. musculus* and human NPAS2 proteins and not with its respective paralog *D. rerio* CLK1 (Fig. 4).

## Discussion

Genome duplication events such as the one preceding the evolution of salmonid fishes, have equipped circadian clock systems with multiple versions of the core oscillator components. Here, we report the isolation and partial characterization of two Chinook salmon *Clock* genes,



**Fig. 4** Phylogenetic analysis of 11 CLOCK proteins from five species. Multiple sequence alignment of exons 15–19 was generated using ClustalW (Thompson et al. 1994). The tree was constructed using ProDist and Neighbor-Joining programs in Phylip (Felsenstein 1991). Values at the nodes represent percent support from bootstrapping the tree 1,000 times. GenBank accession numbers for published sequences are as follows: *D. rerio* CLK1: NP571032; *D. rerio* NPAS2: NP840084; *D. rerio* CLK3: NP840080; *G. gallus* CLK1: AAD32860; *G. gallus* NPAS2: AAU93340; *H. sapiens* CLK1: AAH41878; *H. sapiens* NPAS2: NP002509; *M. musculus* CLK1: AAD30565; and *M. musculus* NPAS2: AAI09167

*OtsClock1a* and *OtsClock1b*, each containing a distinct polyglutamine repeat motif (PolyQ). This is the first study to identify multiple copies of a circadian rhythm gene in a salmonid fish and the first discovery a glutamine-rich region located inside a CLOCK protein.

Comparative genomics has led to major advances in our understanding of the mechanisms and complexity of clock systems. Detailed description of oscillator components identified in *Drosophila*, for example, has facilitated the discovery of most mammalian orthologs. Since a majority of genes identified in non-model organisms will never be studied experimentally, transfer of functional information among orthologs is a primary means to characterize molecular networks. Based on the phylogenetic analysis, we conclude that OTSCLOCK1a and OTSCLOCK1b are gene duplicates which likely evolved following the *Salmonidae* autotetraploidization event. Expression of these gene duplicates in Chinook salmon brain tissue indicates that neither copy has evolved into a nonfunctional pseudogene. While the evolutionary processes underlying maintenance of *Clock1* gene duplicates are unknown, functional divergence between duplicates is usually required for their long-term retention in the genome (He and Zhang 2006). One potential mechanism is the partitioning of the ancestral function between the two copies by the process of subfunctionalization (Force et al. 1999). Additional research is necessary to elucidate the evolutionary processes leading to the preservation of *Clock* gene duplicates in the Chinook salmon genome.

Phylogenetic comparison of CLOCK proteins revealed that orthologs with the same known or inferred function are more closely related to each other than they are to their respective paralogs. This pattern suggests that CLOCK paralogs appear to have undergone functional diversification and specialization whereas the fundamental role of orthologs has been retained through major evolutionary transitions. Both OTSCLOCK1a and OTSCLOCK1b are most similar to ZFCLOCK1, the primary copy driving the autoregulatory feedback loop (Whitmore et al. 1998), suggesting that these proteins may also play an essential role in regulating the circadian clock network of Chinook salmon.

Polyglutamine repeats, encoded by the trinucleotide repeat CAG, are known to function in the transcriptional activation of proteins (Darlington et al. 1998; Mitchell and Tijan 1989). Similar to what has been described for zebrafish, the Chinook salmon putative PolyQ domains are long and interrupted by non-glutamine residues. *OtsClock1a* contains a stretch of 20 glutamine residues, whereas 41 glutamine repeats constitute the polyglutamine repeat of *OtsClock1b*. Comparison among a number of other organisms revealed that the *Clock* PolyQ domain is highly variable in length. For instance, in human and

baboon the glutamine stretch is short and fixed (Q)<sub>6</sub> whereas in *Drosophila melanogaster* the stretch is considerably longer and shows intraspecific length variation: (Q)<sub>25</sub>, (Q)<sub>29</sub>, and (Q)<sub>33</sub> (Saleem et al. 2001). Studies have demonstrated that expansion or contraction in glutamine repeats within a single copy of *Clock* directly affects the corresponding gene product and may even alter the phenotype. For example, a mouse *Clock* mutant allele where 51 amino acids have been deleted from the Q-rich region resulted in a lengthening of the circadian period by 1 h in heterozygotes and by 4 h in homozygotes (Vitaterna et al. 1994; King et al. 1997; Gekakis et al. 1998). Taking into account both the functional importance and degree of inter- and intra-specific length polymorphism, the polyglutamine repeat region makes for an interesting subject to study allelic variation in the *Clock* gene.

While the Chinook *Clock* transcripts appear to share high sequence identity, they are differentiated by a 1,200 bp non-coding segment homologous to the *S. salar Transferrin* gene. Interestingly, the PolyQ domain is located directly upstream of this indel. It has been reported that the trinucleotide repeat, (CAG)<sub>N</sub>, forms non-B DNA conformations that can play a major role in the involvement of double-strand breaks and thus facilitate chromosomal rearrangements (Bacolla and Wells 2004). In a study investigating the evolution of hemoglobin genes in Antarctic icefishes, Near et al. (2006) proposed that repeat sequences facilitated intrachromosomal DNA rearrangements via recombination-repair activities at non-B DNA conformations resulting in the diversification and loss of hemoglobin genes. Similarly, other studies have demonstrated that trinucleotide repeats can act as recombination hot spots by enhancing the rate of recombination relative to the genome average (Jeffreys et al. 1998; Richard and Paques 2000). These intragenic repeats trigger frequent recombination events and the subsequent variation in gene size can lead to quantitative alterations in phenotypes (Verstrepen et al. 2005). Thus, the Chinook salmon PolyQ domain may be a hotspot for nonhomologous recombination and/or gene conversion resulting in the introgression of the *Transferrin* gene and subsequent diversification of duplicated *Clock* genes.

In conclusion, this is the first study to characterize multiple copies of the *Clock* gene in a non-model fish species and the first report of a putative polyglutamine domain located inside a CLOCK protein. Deciphering the molecular basis of circadian rhythms is critical to understanding the temporal coordination of a wide range of biological processes from cellular activities to behavior. Furthermore, circadian rhythm genes indirectly influence a number of physiological and biochemical process by regulating a suite of “clock-controlled” genes, including cell cycle regulator genes and gonadotropin-releasing

hormone (Matsuo et al. 2003; Chappell et al. 2003). Ultimately, the molecular characterization of *OtsClock1a* and *OtsClock1b* establishes a foundation for future studies of clock gene evolution in salmonids and provides a valuable resource to examine the role of *Clock* genes in regulating temporally controlled life history traits, such as time of migration and spawning, in natural populations of Chinook salmon.

**Acknowledgements** This research was supported by the California Department of Water Resources and the Mamie M. Markham Research Grant. We are grateful to M. Camara and G. Moyer for comments that considerably improved the manuscript and I. Meusnier for help with mRNA sampling and screening the BAC library. We thank B. Devlin of the Department of Fisheries and Oceans, British Columbia, Canada for providing reagents, laboratory equipment, and technical support to probe the Chinook salmon BAC library.

## References

- Allada R, White NE, So WV, Hall JC, Robash M (1998) A mutant *Drosophila* homolog of mammalian *Clock* disrupts circadian rhythms and transcription of *period* and *timeless*. *Cell* 93:791–804
- Allendorf FW, Thorgaard GH (1984) Tetraploidy and the evolution of salmonid fishes. In: Turner B (ed) *Evolutionary genetics of fishes*. Plenum Press, New York, pp 1–53
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Antoch MP, Song EJ, Chang AM, Vitaterna MH, Zhao Y, Wilsbacher LD, Sangoram AM, King DP, Pinto LH, Takahashi JS (1997) Functional identification of the mouse circadian *Clock* gene by transgenic BAC rescue. *Cell* 89:655–667
- Avivi A, Albrecht U, Oster H, Joel A, Beiles A, Nevo E (2001) Biological clock in total darkness: the *Clock/MOP3* circadian system of the blind subterranean mole rat. *Proc Natl Acad Sci U S A* 98:13751–13756
- Bacolla A, Wells RD (2004) Non-B DNA conformations, genomic rearrangements, and human diseases. *J Biol Chem* 279:47411–47414
- Bailey GS, Poulter RT, Stockwell PA (1978) Gene duplication in tetraploid fish: model for gene silencing at unlinked duplicated loci. *Proc Natl Acad Sci U S A* 78:5575–5579
- Chang DC, McWatters HG, Williams JA, Gotter AL, Levine JD, Reppert SM (2003) Constructing a feedback loop with circadian clock molecules from the silkworm, *Antheraea pernyi*. *J Biol Chem* 278:38149–38158
- Chappell PE, White RS, Mellon PL (2003) Circadian gene expression regulates pulsatile gonadotropin-releasing hormone (GnRH) secretory patterns in the hypothalamic GnRH-secreting GT1–7 cell line. *J Neuro* 23:112022–11213
- Children’s Hospital Oakland Research Institute [<http://bacpac.chori.org/highdensity.htm>]
- Cronn R, Cedroni M, Haselkorn T, Grover C, Wendel JF (2002) PCR-mediated recombination in amplification products derived from polyploidy cotton. *Theor Appl Genet* 104:482–489
- Darlington TK, Wager-Smith K, Ceriani MF, Staknis D, Gekakis N, Steeves TDL, Weitz CJ, Takahashi JS, Kay SA (1998) Closing the circadian loop: *Clock*-induced transcription of its own inhibitors *per* and *tim*. *Science* 280:1599–1603
- Felsenstein J (1991) PHYLIP: phylogeny inference package, version 3.4. Department of Genetics, SK-50, University of Washington, Seattle

- Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545
- Gately KA (2000) Regulation of the Atlantic salmon (*Salmo salar*) *transferrin* gene and liver gene expression. Thesis (Ph.D.): National University of Ireland, Galway pp 197–224
- Gekakis N, Staknis D, Nguyen HB, Davis FC, Wilsbacher LD, King DP, Takahashi JS, Weitz CJ (1998) Role of the CLOCK protein in the mammalian circadian mechanism. *Science* 280:1564–1569
- Hall T (1997) BioEdit Sequence Alignment Editor. Ibis Therapeutics, Carlsbad, Ca
- He X, Zhang J (2006) Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* 169:1157–1164
- Hirayama J, Kaneko M, Cardone L, Cahill G, Sassone-Corsi P (2005) Analysis of circadian rhythms in zebrafish. *Methods Enzymol* 393:186–204
- Hughes A (1994) The evolution of functionally novel proteins after gene duplication. *Proc R Soc Lon B* 256:119–124
- Jeffreys AJ, Neil DL, Neumann R (1998) Repeat instability at human minisatellites arising from meiotic recombination. *EMBO J* 17:4147–4157
- Kimura A, Ohta T (1974) On some principles governing molecular evolution. *Proc Nat Acad Sci U S A* 71:2848–2852
- King DP, Zhao YL, Sangoram AM, Wilsbacher LD, Tanaka M, Antoch MP, Steeves TDL, Vitaterna MH, Kornhauser JM, Lowrey PL, Turek FW, Takahashi JS (1997) Positional cloning of the mouse circadian CLOCK gene. *Cell* 89:641–653
- Looby P, Loudon ASI (2005) Gene duplication and complex circadian clocks in mammals. *Trends Genet* 21:46–53
- Matsuo T, Yamaguchi S, Mitsui S, Emi A, Shimoda F, Okamura H (2003) Control mechanism for the circadian clock for timing of cell division in vivo. *Science* 202:255–259
- Mazurais D, Le Drean G, Brierley I, Anglade I, Bromage N, Williams LM, Kah O (2000) Expression of clock gene in the brain of rainbow trout: comparison with the distribution of melatonin receptors. *J Comp Neurol* 422:612–620
- McNamara P et al (2001) Regulation of CLOCK and MOP4 by nuclear hormone receptors in the vasculature: a humoral mechanism to reset a peripheral clock. *Cell* 105:877–889
- Mitchell PJ, Tijian R (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* 245:371–378
- Near TJ, Parker SK, Detrich HW (2006) A genomic fossil reveals the key steps in hemoglobin loss by the Antarctic icefishes. *Mol Biol Evol* 23:2008–2016
- Ohno S, Wolf U, Atkin NB (1968) Evolution from fish to mammals by gene duplication. *Hereditas* 59:169–187
- Reppert SM, Weaver DJ (2002) Coordination of circadian timing in mammals. *Nature* 418:935–941
- Richard G-F, Paques E (2000) Myotonic dystrophy: RNA pathogenesis comes into focus. *Am J Hum Genet* 74:793–804
- Saleem Q, Anand A, Jain S, Brahmachari SK (2001) The polyglutamine motif is highly conserved at the *Clock* locus in various organisms and is not polymorphic in humans. *Hum Genet* 109:136–142
- Schultz J, Milpetz F, Bork P, Ponting CP (1998) SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* 95:5857–5864
- Schwassmann HO (1988) Biological rhythms. In: Hoar WS, Randall DJ (eds) *Fish Physiology*. Academic Press, New York, pp 371–416
- Steeves TDL, King DP, Zhao Y, Sangoram AM, Du F, Bowcock AM, Moore RY, Takahashi JS (1999) Molecular cloning and characterization of the human CLOCK gene: expression in the superchiasmatic nuclei. *Genomics* 57:189–200
- Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
- Verstrepen K, Jansen A, Lewitter F, Fink G (2005) Intragenic tandem repeats generate functional variability. *Nat Genet* 37:986–990
- Vitaterna MH, King DP, Chang AM, Kornhauser JM, Lowrey PL, McDonald JD, Dove WF, Pinto LH, Turek FW, Takahashi JS (1994) Mutagenesis and mapping of a mouse gene, *Clock*, essential for circadian behavior. *Science* 264:719–725
- Whitmore D, Foulkes NS, Strahle U, Sassone-Corsi P (1998) Zebrafish *Clock* rhythmic expression reveals independent peripheral circadian oscillators. *Nature Neurosci* 1:701–707
- Yoshimura T, Suzuki Y, Makino E, Suzuki T, Kuroiwa A, Matsuda Y, Namikawa T, Ebihara S (2000) Molecular analysis of avian clock genes. *Mol Brain Res* 78:207–215