

The Circadian Clock of Teleost Fish: A Comparative Analysis Reveals Distinct Fates for Duplicated Genes

Jessica Toloza-Villalobos · José Ignacio Arroyo ·
Juan C. Opazo

Received: 9 May 2014 / Accepted: 26 November 2014 / Published online: 9 December 2014
© Springer Science+Business Media New York 2014

Abstract The circadian clock is a central oscillator that coordinates endogenous rhythms. Members of six gene families underlie the metabolic machinery of this system. Although this machinery appears to correspond to a highly conserved genetic system in metazoans, it has been recognized that vertebrates possess a more diverse gene inventory than that of non-vertebrates. This difference could have originated in the two successive rounds of whole-genome duplications that took place in the common ancestor of the group. Teleost fish underwent an extra event of whole-genome duplication, which is thought to have provided an abundance of raw genetic material for the biological innovations that facilitated the radiation of the group. In this study, we assessed the relative contributions of whole-genome duplication and small-scale gene duplication to generate the repertoire of genes associated with

the circadian clock of teleost fish. To achieve this goal, we annotated genes from six gene families associated with the circadian clock in eight teleost fish species, and we reconstructed their evolutionary history by inferring phylogenetic relationships. Our comparative analysis indicated that teleost species possess a variable repertoire of genes related to the circadian clock gene families and that the actual diversity of these genes has been shaped by a variety of phenomena, such as the complete deletion of ohnologs, the differential retention of genes, and lineage-specific gene duplications. From a functional perspective, the sub-functionalization of two ohnolog genes (PER1a and PER1b) in zebrafish highlights the power of whole-genome duplications to generate biological diversity.

Keywords Whole-genome duplication · Gene duplication · Differential retention · Gene family evolution

Electronic supplementary material The online version of this article (doi:10.1007/s00239-014-9660-x) contains supplementary material, which is available to authorized users.

J. Toloza-Villalobos (✉) · J. I. Arroyo · J. C. Opazo (✉)
Instituto de Ciencias Ambientales y Evolutivas, Facultad de
Ciencias, Universidad Austral de Chile, Valdivia, Chile
e-mail: tolozavillalobos.dc@gmail.com

J. C. Opazo
e-mail: jopazo@gmail.com

J. Toloza-Villalobos · J. I. Arroyo
Programa de Doctorado en Ciencias mención Ecología y
Evolución, Facultad de Ciencias, Universidad Austral de Chile,
Valdivia, Chile

Present Address:

J. I. Arroyo
Departamento de Ecología, Facultad de Ciencias Biológicas,
Pontificia Universidad Católica de Chile, Alameda 340,
Santiago, Chile

Introduction

Light represents the most important signal able to entrain endogenous rhythms, and central oscillators play a key role in controlling daily biochemical, physiological, and behavioral rhythms (Reppert and Weaver 2001). Mammalian species possess a central oscillator, known as the circadian clock, which is located in the suprachiasmatic nuclei of the anterior hypothalamus (Vatine et al. 2009). In contrast, experimental evidence has failed to identify a central control mechanism in non-mammalian vertebrates, and it has instead been indicated that they possess more distributed local oscillators. In this group, the pineal gland appears to play a central role in the endogenous response to light and melatonin secretion as early as 19 h post-fertilization (Danilova et al. 2004).

The circadian clock system plays fundamental roles in anticipating the day/night cycle and coordinating the molecular oscillator that generates an endogenous rhythm that is close to 24 h. Thus, the photoperiod regulates gene expression through the central clock, generating feedback and ultimately producing an endogenous rate (Devlin 2002). Genes involved in circadian rhythms have been identified through genetic screening in a variety of organisms and have been classified into six groups. The period circadian clock gene family (PER) is composed of three genes (PER1, PER2, and PER3), which are located on three separate chromosomes and perform critical functions in regulating the circadian rhythms of locomotor activity, metabolism, and behavior. The cryptochrome circadian clock genes (CRY1 and CRY2) are also located on separate chromosomes; these genes encode a flavin adenine dinucleotide-binding protein that belongs to the circadian core oscillator complex. The CLOCK/NPAS2 genes encode transcription factors that bind to the upstream regions of PER and CRY genes, activating their transcription. The ARNTL/ARNTL2 genes are also located on separate chromosomes and encode a basic helix-loop-helix transcription factor that forms a heterodimer with CLOCK and then binds enhancer elements upstream of the PER and CRY gene systems. The casein kinase 1 epsilon (CSNK1E) and delta (CSNK1D) genes encode a serine/threonine protein kinase involved in the regulation of circadian rhythms through the phosphorylation of members of the PER gene family. Finally, the Timeless gene is another key component of circadian rhythms, which interacts with the PER genes and others to downregulate the activation of PER1 by CLOCK/ARNTL.

Although the genetic regulation of circadian rhythms appears to be similar in fish, insects, and mammals (Dunlap 1999), the numbers of gene copies involved in this process vary among these groups. For example, rodents possess a more complex repertoire that is characterized by three PER genes, two CRY genes, two CLOCK genes, two ARNTL genes, two CSNK1 genes, and one timeless gene, whereas species of the genus *Drosophila* possess only one gene copy per family (Reppert and Weaver 2001). Furthermore, zebrafish (*Danio rerio*) exhibits an increased number of genes (Fig. 1) compared with that of rodents. These differences have been linked to the genome duplication that occurred in the last common ancestor of teleost fish between 320 and 400 mya (Amores et al. 1998, 2011; Postlethwait et al. 2000; Taylor et al. 2001, 2003; Van de Peer et al. 2003; Hoegg et al. 2004; Jaillon et al. 2004; Meyer and Van de Peer 2005; Kasahara et al. 2007; Sato and Nishida 2010).

This teleost-specific genome duplication (TGD) is thought to have provided abundant raw genetic material to allow biological innovations that facilitated the radiation of

the group (Amores et al. 2004, 2011; Postlethwait et al. 2000; Taylor et al. 2001, 2003; Hoegg et al. 2004; Jaillon et al. 2004; Kasahara et al. 2007; Sato and Nishida 2010; Opazo et al. 2013) into a wide variety of lineages inhabiting both marine and freshwater environments, making teleosts the most diverse vertebrate group (Meyer and Van de Peer 2005; Nelson 2006). This conclusion is supported by the many studies that have shown a link between new genes that originated during the TGD and the evolution of new functions in a variety of teleost lineages (Lister et al. 2001; Mulley et al. 2006; Hashiguchi and Nishida 2007; Hoegg and Meyer 2007; Sato and Nishida 2007; Siegel et al. 2007; Yu et al. 2007; Douard et al. 2008; Arnegard et al. 2010; Opazo et al. 2013).

Most of the studies on the genetic structure of the circadian clock in teleost fish have targeted individual genes (Wang 2008a, b, 2009), although none has taken a more holistic, multi-gene approach to understand the evolution of the complete genetic machinery associated with the regulation of endogenous rhythms. Accordingly, the goal of this study was to assess the relative contributions of whole-genome duplication and small-scale gene duplication in generating the diversity within the genetic machinery associated with the regulation of endogenous rhythms. To achieve this goal, we annotated genes from six gene families associated with the circadian clock in eight teleost fish species with available genomes and reconstructed their evolutionary history by inferring phylogenetic relationships. The results of our comparative analysis indicated that teleost species possess a variable repertoire of genes associated with the circadian clock gene families and that the actual diversity of these genes has been shaped by a variety of phenomena, such as a complete deletion of ohnologs, differential retention of genes, and lineage-specific gene duplications. From a functional perspective, the subfunctionalization of two ohnolog genes (PER1a and PER1b) in zebrafish highlights the power of whole-genome duplication in the generation of biological diversity.

Materials and Methods

DNA Sequence Data

We used bioinformatics tools to manually annotate the complement of genes in the genomes of eight teleost fish species available from the Ensembl database (Supplementary Table 1) (zebrafish, *Danio rerio*; Atlantic cod, *Gadus morhua*; fugu, *Takifugu rubripes*; stickleback, *Gasterosteus aculeatus*; medaka, *Oryzias latipes*; platyfish, *Xiphophorus maculatus*; Nile tilapia, *Oreochromis niloticus*; and Tetraodon, *Tetraodon nigroviridis*). When available, sequences from the spotted gar (*Lepisosteus oculatus*), a non-

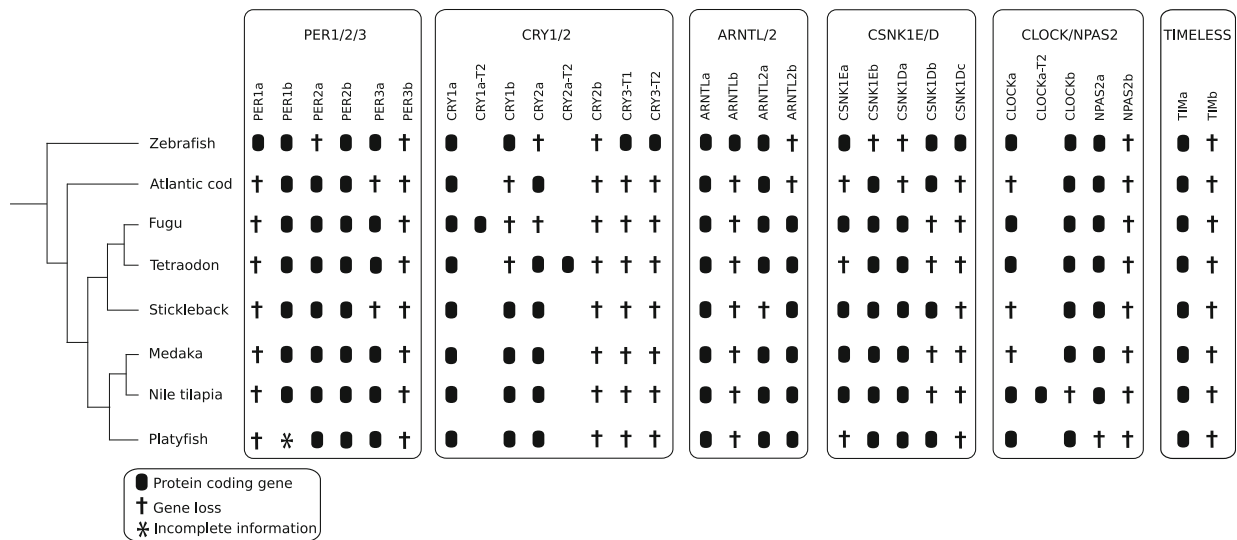


Fig. 1 Genomic structure of the gene families associated with the circadian clock of teleost fish

teleost ray-finned fish, were included as well. We also included sequences from five tetrapod species as an out-group for phylogenetic analyses (western clawed frog, *Xenopus tropicalis*; anole lizard, *Anolis carolinensis*; chicken, *Gallus gallus*; mouse, *Mus musculus*; and human, *Homo sapiens*).

We compared the annotated exon sequences obtained from public databases (e.g., GenBank) to unannotated genomic fragments (Supplementary Table 1) obtained from ENSEMBL using the program Blast2seq (Tatusova and Madden 1999), which is available from NCBI (<http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi>). An intact open reading frame with the canonical structure typical of vertebrate genes characterized the putative functional genes, whereas pseudogenes were identifiable based on their high sequence similarity to functional orthologs and the presence of inactivating mutations and/or a lack of exons. To distinguish among tandemly arrayed gene copies, we indexed each gene copy with the symbol T, followed by a number corresponding to the linkage order in the 5 to 3' orientation; thus, the first gene in the cluster is labeled T1, the second T2, and so forth.

Phylogenetic Inference

We estimated phylogenetic relationships separately for each gene family in the teleost fish. We performed maximum likelihood and Bayesian analyses, as implemented in the programs Treefinder, March 2011 version (Jobb et al. 2004), and Mr. Bayes, v3.1.2 (Ronquist and Huelsenbeck 2003), respectively. We also reconstructed the phylogenetic relationships using maximum likelihood as implemented in the program CodonPhyML (Gil et al. 2013).

This approach employs a more realistic description of the evolutionary process at the protein-coding sequence level by incorporating the structure of the genetic code into the model. Sequence alignments were carried out using the G-INS-i strategy in Mafft v6.8 (Katoh et al. 2009). The best fitting models for each codon position were estimated separately using the proposed model routine from the program Treefinder, March 2011 version (Jobb et al. 2004). In the maximum likelihood analysis, we estimated the best tree under the selected models and assessed support for each of the nodes with 1,000 bootstrap pseudoreplicates. In the Bayesian analysis, two simultaneous independent runs were performed for 30×10^6 iterations of a Markov Chain Monte Carlo algorithm, with six simultaneous chains sampling trees every 1,000 generations. Support for the nodes and parameter estimates was derived from a majority rule consensus of the last 15,000 trees sampled after convergence. The average standard deviation of split frequencies remained at 0.01 after the burn-in threshold. For the maximum likelihood approach implemented in the program CodonPhyML (Gil et al. 2013), the model described by Goldman and Yang (1994) involving a subtree pruning and regrafting (SPR) heuristic search with 5 random starting trees was employed to reconstruct the phylogenetic relationships. Support for the nodes was assessed using the aBayes method (Anisimova et al. 2011).

Results

For each species, we identified and characterized all of the structural genes belonging to the circadian clock gene families (Fig. 1). We observed that most of the putative

functional genes fell into separate clades that did not deviate significantly from the expected organismal phylogenies. In some cases, we observed differential retention of genes that originated in the last common ancestor of the group, whereas in others, lineage-specific duplications and the deletion of complete sets of ohnologs were observed (Fig. 1).

Phylogenetic reconstruction of the PER gene family recovered three main groups (Fig. 2). Within each group, the monophyly of the teleost ohnologs was well supported, as were the sister group relationships among them (Fig. 2). When sequences from the spotted gar were included (e.g., PER1 and PER3), they were recovered as a sister group to the clade containing teleost ohnologs (Fig. 2). The clade that included the tetrapod sequences was recovered as a sister group to the teleost fish clade (Fig. 2). We note that the PER3b gene was lost in all of the examined species, and differential retention of PER1a was observed in zebrafish (Fig. 2).

For the genes associated with the CRY gene family, our tree recovered the sister group relationship between the CRY1a and CRY1b clades with strong support (Fig. 2). The latter gene was independently lost in Atlantic cod and

in the last common ancestor of fugu and Tetraodon (Fig. 1). Additionally, the CRY1a gene was independently duplicated in fugu (Figs. 1, 2). The monophyly of the CRY2a gene was strongly supported (Fig. 2); this gene was independently lost in fugu and zebrafish (Fig. 1) and was duplicated in Tetraodon (Figs. 1, 2). The ohnolog CRY2b was not found in any of the teleost fish species analyzed, and we assume that it was lost in the last common ancestor of the group (Fig. 1). In zebrafish, we found two additional genes, CRY3-T1 and CRY3-T2, that were recovered as a sister group to the CRY2 clade (which included tetrapods). We believe that this lineage represents an old gene that originated in the common ancestor of tetrapods and fish and was only retained in zebrafish, where it was later duplicated (Figs. 1, 2).

For the ARNTL/ARNTL2 genes, we recovered a clade containing the ARNTLa genes of all teleost species as a sister to the ARNTLb gene of zebrafish (Fig. 2), which was the only species that retained this gene (Fig. 1). The single-copy ARNTL gene of the spotted gar was recovered as a sister to this clade (Fig. 2). We also recovered the sister group relationship of the ARNTL2a and ARNTL2b clades (Fig. 2). The first gene was lost in the stickleback, whereas

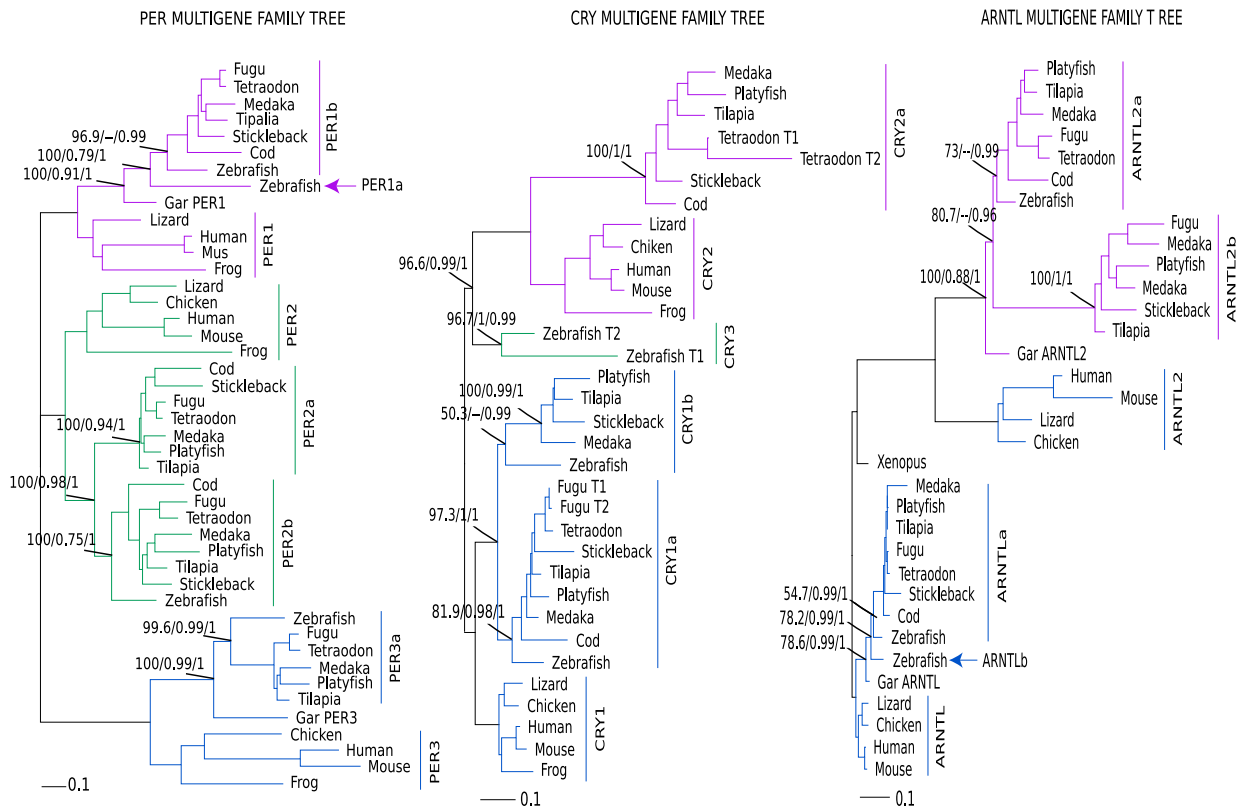


Fig. 2 Maximum likelihood phylogenetic trees depicting the relationships of the period circadian clock gene family (PER; *left*), the cryptochrome circadian clock gene family (CRY; *center*), and the aryl hydrocarbon receptor nuclear translocator-like gene family (ARNTL; *right*). Values on the relevant nodes denote the bootstrap support values, Bayesian posterior probabilities, and aBayes posterior probabilities

the second was independently lost in zebrafish and Atlantic cod (Fig. 1).

For the casein kinase 1 epsilon (CSNK1E) and delta (CSNK1D) genes, our phylogenetic analysis indicated a sister group relationship between the CSNK1Ea and CSNK1Eb clades, which was in turn recovered as a sister to the CSNK1E clade of tetrapods (Fig. 3). A similar pattern was found for the CSNK1Da and CSNK1Db genes (Fig. 3). In zebrafish, we identified an additional gene, CSNK1Dc, that was recovered as a sister group of the CSNK1D clade of teleost fish. It can be assumed that CSNK1Dc is an ancient gene that originated in the common ancestor of teleost fish and was only retained in zebrafish (Figs. 1, 3). We also observed examples of the differential retention of genes (Fig. 3), where genes that originated in the last common ancestor of teleost fish were later lost in certain species (Fig. 3). For example, CSNK1Db was independently lost in the last common ancestor of fugu and Tetraodon and in the common ancestor of medaka and Nile tilapia (Fig. 1).

In the case of the CLOCK/NPAS2 gene family, we recovered the sister group relationship between the CLOCKa and CLOCKb clades of teleost fish (Fig. 3). The

single-copy CLOCK gene of the spotted gar was recovered as a sister to this clade (Fig. 3). This copy of the CLOCKa gene was independently lost in Atlantic cod, stickleback, and medaka (Fig. 1), whereas the CLOCKb gene was only lost in platyfish. Among the NPAS2 genes, we only identified orthologous genes corresponding to the NPAS2a gene, and these genes were found in all species of teleost fish except for the Nile tilapia (Fig. 1).

Finally, the evolutionary pattern of the timeless gene has been simple, as we only identified a single-copy gene, timeless-a, in all teleost species examined. All of these sequences fell within a well-supported clade (Fig. 3), which was in turn recovered as a sister to the single-copy gene identified in the spotted gar (Fig. 3).

Discussion

Our findings regarding the circadian clock gene families of teleost species are consistent with the TGD that occurred in the stem lineage of this group (Amores et al. 1998, 2011; Postlethwait et al. 2000; Taylor et al. 2001, 2003; Van de Peer et al. 2003; Hoegg et al. 2004; Jaillon et al. 2004;

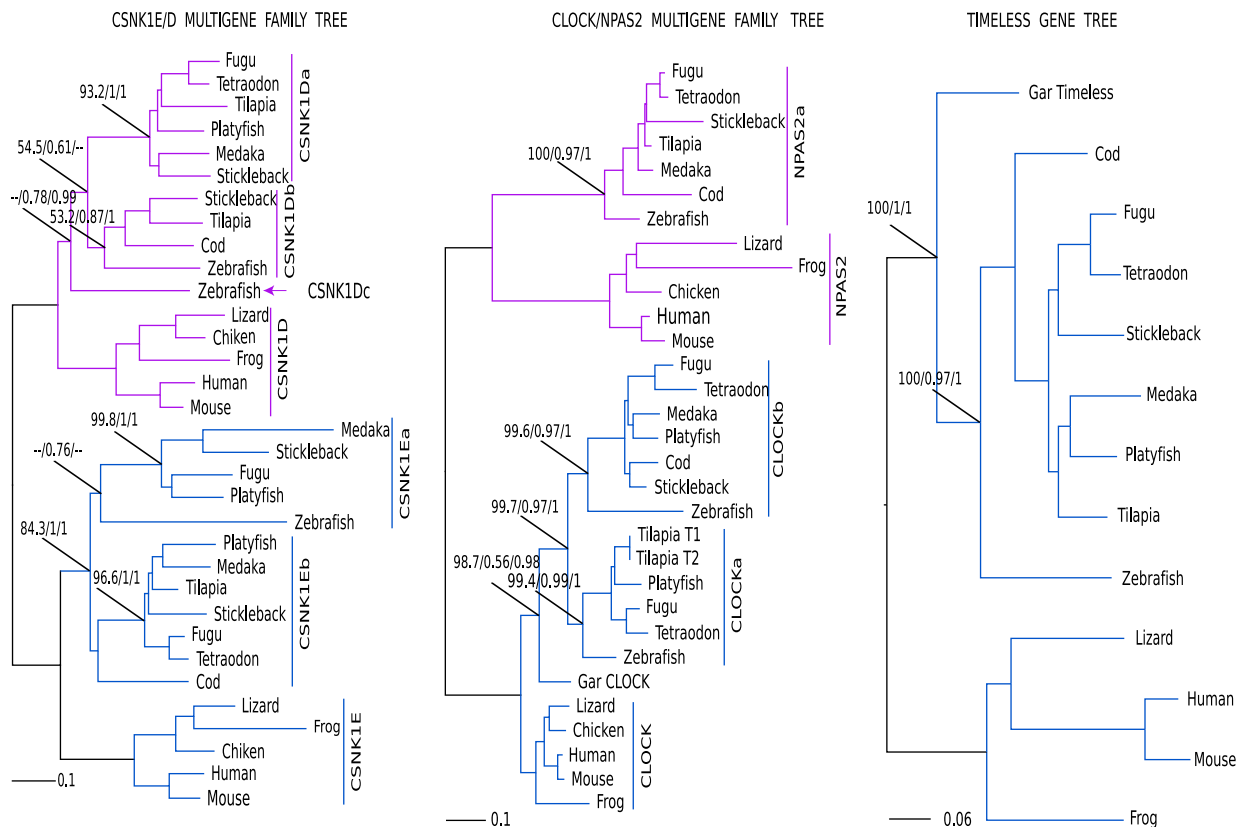


Fig. 3 Maximum likelihood phylogenetic trees depicting the relationships of the casein kinase 1 genes (CSNK1; left), clock circadian regulator/neuronal PAS domain protein 2 genes (CLOCK/NPAS2;

center), and timeless circadian clock genes (right). Values on the relevant nodes denote the bootstrap support values, Bayesian posterior probabilities, and aBayes posterior probabilities

Meyer and Van de Peer 2005; Kasahara et al. 2007; Sato and Nishida 2010). The presence of ohnologous clades and their phylogenetic arrangement as sister groups in all of the examined species indicate the distinctive features of genomes that have experienced an extra round of genomic duplication (Christoffels et al. 2004; Brunet et al. 2006; Amores et al. 2011). Furthermore, the phylogenetic position of the spotted gar, a non-teleost ray-finned fish that was not affected by the TGD (Figs. 2, 3), as a sister to the clade containing a pair of ohnologous clades, is also consistent with the TGD. However, given that the extra round of whole-genome duplication occurred between 320 and 400 mya, lineage-specific dynamics following the TGD would have been responsible for the gene repertoire found in extant species (Fig. 1). In teleost fish species, the actual gene diversity would have been shaped by a variety of phenomena, such as the complete deletion of ohnologs (e.g., PER3b, NPAS2b, and timeless-b), the differential retention of genes (e.g., CRY1b, ARNTL2b, and CSNK1Ea), and lineage-specific gene duplications (e.g., CRY1a-T2, CLOCKa-T2).

To date, few studies have provided insight regarding the evolutionary history of genes related to the circadian clock of teleost fish, and those that have were conducted by analyzing individual groups of genes (Wang 2008a, b, 2009). From a comparative and functional perspective, the studies conducted by Wang (2008a, b, 2009) have made important progress regarding evolution of the ARNTL, PER, and CLOCK/NPAS2 gene families. From a phylogenetic perspective, our results regarding the ARNTL gene family are similar to those reported by Wang (2009) but with some discrepancies. For example, in Wang's phylogenetic reconstruction, he recovered a sister group relationship between the zebrafish ARNTLa and ARNTLb genes, which is a topology that could be attributable to a gene conversion event. In the present study, we were able to resolve the orthology among the ARNTLa genes, as we recovered a clade containing all ARNTLa genes from teleost fish that did not deviate significantly from the expected organismal phylogeny (Fig. 2).

The sister group relationship between the zebrafish ARNTLb gene, the unique teleost species that retained this gene (Fig. 1), and the ARNTLa clade indicates a homology relationship derived from a whole-genome duplication event, i.e., ohnology. Although our results regarding the PER gene family are essentially the same as those of Wang (2008b), it is worth highlighting that zebrafish was the only species that retained the PER1a gene, that the PER3b gene was lost in the last common ancestor of all teleost fish, and that PER3a was independently lost in two teleost species (Fig. 1; Wang 2008b). Similar comparative results were obtained for the CLOCK/NPAS2 gene family, except for

that we found an extra copy of the CLOCKa gene in the Nile tilapia, which we refer to as CLOCKa-T2 (Figs. 1, 3).

The pattern observed for the CRY1/2 gene family corresponds to a combination of various evolutionary models (Fig. 2). This gene family is characterized by a complete deletion of the CRY2b gene (Fig. 1) and by differential retention of the CRY1b and CRY2a, CRY3-T1, and CRY3-T2 genes. Additionally, lineage-specific duplications of the CRY1a and CRY2a genes in the lineages leading to fugu and Tetraodon have also been found (Fig. 2). It is worth noting that the CRY3 gene represents an old gene lineage that originated in the last common ancestor of tetrapods and fish but was only retained in zebrafish (Fig. 2). Although the functional diversity of these genes is not fully understood, it has been demonstrated that they oscillate in a circadian manner in zebrafish, showing various patterns and playing a variety of roles in the molecular clock (Kobayashi et al. 2000). A process of differential retention at each locus characterizes the mode of evolution of the CSNK1 gene family. Finally, the simplest scenario was found for genes belonging to the timeless gene family, in which one ohnolog, timeless-b, was lost in the last common ancestor of teleost fish, and the other ohnolog, timeless-a, was retained as a single-copy gene in all of the examined species.

Although it is important to understand the evolutionary fate of the genetic machinery underlying the molecular clock, linking this machinery to its functional consequences is one of the major goals of evolutionary biology. In this regard, Wang (2008b) determined that the duplicated copies of the PER1 gene found in zebrafish show distinct patterns of temporal and spatial expression. According to his functional evidence, when the fish are in a dark environment, the PER1a gene is expressed in the retina and in both the telencephalon and the diencephalon of the forebrain, and there is no detectable expression of PER1b. However, when the fish are exposed to light, the expression of PER1a is significantly upregulated, and the PER1b gene is constitutively expressed throughout the head region, including the brain and retina (Wang et al. 2008b). This pattern represents a clear example of sub-functionalization, in which each daughter copy adopts part of the function of the parental gene (Force et al. 1999).

Finally, we wish to offer a note of caution regarding variation in the quality of the genomes of the surveyed species. In the literature, it has been reported that genome assembly has an impact on gene prediction (Florea et al. 2011). Hence, future improvements in fish genome assemblies will help us to better understand gene copy number variation in the gene families associated with the circadian clock of teleost fish.

Concluding Remarks

Our study revealed that following TGD, not all ohnologous genes were equally retained in the genomes of the species surveyed. In most gene families, at least one ohnolog was lost in the last common ancestor of teleost fish. In cases in which genes were retained, lineage-specific dynamics were responsible for the gene repertoire in present-day species. This mode of evolution highlights the role of the differential retention of relatively old genes in shaping the actual gene complement in teleost fish. Additionally, by connecting our comparative results with those of functional studies, we have provided examples of subfunctionalization, in which daughter copies adopt a portion of the function of their parental gene, as was observed for the *PER1* duplicates. Taking all of these findings together, this study provides further evidence that teleost-specific genome duplication fueled key innovations in the genes related to the circadian clock, highlighting that whole-genome duplications have played a fundamental role in the origin of evolutionary novelties.

Acknowledgments This work was supported by Grants from the Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT 1120032) to JCO and from the Programa de Mejoramiento de la Calidad y la Equidad de la Educación Superior (MECESUP AUS1203). JTV holds a MECESUP (AUS0703) doctoral fellowship. JIA holds a CONICYT doctoral fellowship. We also thank Espiritu del Elqui for helpful comments and edits.

References

- Amores A, Force A, Yan YL, Joly L, Amemiya C, Fritz A, Ho K, Langeland J, Prince V, Wang YL, Westerfield M, Ekker M, Postlethwait JH (1998) Zebrafish hox clusters and vertebrate genome evolution. *Science* 282:1711–1714
- Amores A, Suzuki T, Yan Y, Pomeroy J, Singer A, Amemiya C, Postlethwait J (2004) Developmental roles of Pufferfish Hox clusters and genome evolution in ray-fin fish. *Genome Res* 14:1–10
- Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH (2011) Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. *Genetics* 188:199–808
- Anisimova M, Gil M, Dufayard JF, Dessimoz C, Gascuel O (2011) Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst Biol* 60:685–699
- Arnégaré ME, Zwickl DJ, Lu Y, Zakon HH (2010) Old gene duplication facilitates origin and diversification of an innovative communication system—twice. *Proc Natl Acad Sci* 107:22172–22177
- Brunet F, Crollius H, Paris M, Aury J, Gibert P, Jaillon O, Laudet V, Robinson-Rechavi M (2006) Gene loss and evolutionary rates following whole-genome duplication in teleost fishes. *Mol Biol Evol* 23:1808–1816
- Christoffels A, Koh EGL, Chia J, Brenner S, Aparicio S, Venkatesh B (2004) Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes. *Mol Biol Evol* 21:1146–1151
- Danilova D, Krupnik VE, Sugden D, Zhdanova I (2004) Melatonin stimulate cell proliferation in Zebrafish embryo and accelerate its development. *FASEB J* 18:751–753
- Devlin PF (2002) Sings of the tie: environmental input to the circadian clock. *J Exp Bot* 53:1535–1550
- Douard V, Brunet F, Boussau B, Ahrens-Fath I, Vlaeminck-Guillem V, Haendler B, Laudet V, Guiguen Y (2008) The fate of the duplicated androgen receptor in fishes: a late neofunctionalization event? *BMC Evol Biol*. doi:10.1186/1471-2148-8-336
- Dunlap JC (1999) Molecular bases for circadian clock. *Cell* 96:271–290
- Florea L, Souvorov A, Kalbfleisch TS, Salzberg SL (2011) Genome assembly has a major impact on gene content: a comparison annotation in two *Bos taurus* assemblies. *PLoS ONE* 6:e21400
- Force A, Lynch M, Pickett FB, Amores A, Yan Y, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151:1531–1545
- Gil M, Serrano Zanetti M, Zoller S, Anisimova M (2013) Codon-hyml: fast maximum likelihood phylogeny estimation under codon substitution model. *Mol Biol Evol* 30:1270–1280
- Goldman N, Yang Z (1994) A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol* 11:725–736
- Hashiguchi Y, Nishida M (2007) Evolution of trace amine associated receptor (TAAR) gene family in vertebrates: lineage-specific expansions and degradations of a second class of vertebrate chemosensory receptors expressed in the olfactory epithelium. *Mol Biol Evol* 24:2099–2107
- Hoegg S, Meyer A (2007) Phylogenomic analyses of KCNA clusters in vertebrates: why do some clusters stay intact? *BMC Evol Biol* 7:139
- Hoegg S, Brinkmann H, Taylor JS, Meyer A (2004) Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. *J Mol Evol* 59:190–203
- Jaillon O, Aury J, Brunet F, Petit JL, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, Nicaud S, Jaffe D, Fisher S, Lutfalla G, Dossat C, Segurens B, Dasilva C, Salanoubat M, Levy M, Boudet N, Castellano S, Anthouard V, Jubin C, Castelli V, Katinka M, Vacherie B, Biémont C, Skalli Z, Cattolico L, Poulain J, De Berardinis V, Cruaud C, Duprat S, Brottier P, Coutanceau JP, Gouzy G, Parra G, Lardier G, Chapple C, McKernan KJ, McEwan P, Bosak S, Kellis M, Volff JN, Guigo R, Zody MC, Mesirov J, Lindblad-Toh K, Birren B, Nusbaum C, Kahn D, Robinson-Rechavi M, Laudet V, Schachter V, Quétiér F, Saurin W, Scarpelli C, Wincker P, Lander ES, Weissbach J, Roest Crollius H (2004) Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 431:946–957
- Jobb G, von Haeseler A, Strimmer K (2004) Treefinder: a powerful graphical analysis environment for molecular phylogenetics. *BMC Evol Biol* 4:18
- Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y, Jindo T, Kobayashi D, Shimada A, Toyoda A, Kuroki Y, Fujiyama A, Sasaki T, Shimizu A, Asakawa S, Shimizu N, Hashimoto S, Yang J, Lee Y, Matsushima K, Sugano S, Sakaizumi M, Narita T, Ohishi K, Haga S, Ohta F, Nomoto H, Nogata K, Morishita T, Endo T, Shin-I T, Takeda H, Morishita S, Kohara Y (2007) The Medaka draft genome and insights into vertebrate genome evolution. *Nature* 447:714–719
- Katoh K, Asimenos G, Toh H (2009) Multiple alignment of DNA sequences with MAFFT. *Methods Mol Biol* 537:39–64
- Kobayashi Y, Ishikawa T, Hirayama J, Daiyasu H, Kanai S, Toh H, Fukuda I, Tsujimura T, Terada N, Kamei Y, Yuba S, Iwai A, Todo T (2000) Molecular analysis of Zebrafish photolyase/cryptochrome family: two types of cryptochromes present in Zebrafish. *Genes Cell* 5:725–738

- Lister JA, Close J, Raible DW (2001) Duplicate mitf genes in zebrafish: complementary expression and conservation of melanogenic potential. *Dev Biol* 237:333–344
- Meyer A, Van de Peer Y (2005) From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *BioEssays* 27:937–945
- Mulley JF, Chiu CH, Holland PW (2006) Breakup of a homeobox cluster after genome duplication in teleost. *Proc Natl Acad Sci USA* 103:10369–10372
- Nelson JS (2006) *Fishes of the world*. Hoboken, New Jersey
- Opazo JC, Butts GT, Nery MF, Storz JF, Hoffmann FG (2013) The role of whole-genome duplication in promoting the functional diversification of teleost fish hemoglobins. *Mol Biol Evol* 30:140–153
- Postlethwait JH, Woods IG, Ngo-Hazelett P, Yan YL, Kelly PD, Chu F, Huang H, Hill-Force A, Talbot WS (2000) Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res* 10:1890–1902
- Reppert S, Weaver DR (2001) Molecular analysis of mammalian circadian rhythms. *Annu Rev Physiol* 63:647–676
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574
- Sato Y, Nishida M (2007) Post-duplication charge evolution of phospho-glucose isomerases in teleost fishes through weak selection on many amino acid sites. *BMC Evol Biol* 7:204
- Sato Y, Nishida M (2010) Teleost fish with specific genome duplication as unique models of vertebrate evolution. *Environ Biol Fishes* 88:169–188
- Siegel N, Hoegg S, Salzburger W, Braasch I, Meyer A (2007) Comparative genomics of Parahox cluster of teleost fishes: gene cluster breakup and the retention of genes sets following whole genome duplication. *BMC Genome* 8:312
- Tatusova TA, Madden TL (1999) BLAST2 Sequences a new tool for comparing protein and nucleotide sequence. *FEMS Microbiol Lett* 177:187–188
- Taylor JS, Van de Peer Y, Braasch I, Meyer A (2001) Comparative genomics provides evidence for an ancient genome duplication event in fish. *Phil Trans R Soc B* 356:1661–1679
- Taylor JS, Braasch I, Frickey T, Meyer A, Van de Peer Y (2003) Genome duplication, a trait shared by 22,000 species of ray-finned fish. *Genome Res* 13:382–390
- Van de Peer Y, Taylor JS, Meyer A (2003) Are all fishes ancient polyploids? *J Struct Funct Genomics* 3:65–73
- Vatine G, Vallone D, Appelbaum L, Mracek P, Ben-Moshe Z, Lahiri K, Gothilf Y, Foulkes NS (2009) Light directs zebrafish period 2 Expression via conserved D and E boxes. *PLoS Biol* 7:10
- Wang H (2008a) Comparative analysis of teleost fish genomes reveals preservation of different ancient clock duplicates in different fishes. *Mar Genomics* 1:69–78
- Wang H (2008b) Comparative analyses of period genes in teleost fish genomes. *J Mol Evol* 67:29–40
- Wang H (2009) Comparative genomic analysis of teleost fish bmal genes. *Genetica* 136:149–161
- Yu WP, Yew K, Rajasegaran V, Venkatesh B (2007) Sequencing and comparative analysis of fugu protocadherin clusters reveal diversity of protocadherin genes among teleost. *BMC Evol Biol* 7:49