Ancient Vertebrate Conserved Noncoding Elements Have Been Evolving Rapidly in Teleost Fishes

Alison P. Lee,† Sze Yen Kerk,† Yue Ying Tan,† Sydney Brenner,† and Byrappa Venkatesh*,†

†Comparative Genomics Laboratory, Institute of Molecular and Cell Biology, A*STAR (Agency for Science, Technology and Research), Biopolis, Singapore

*Corresponding author: E-mail: mcbbv@imcb.a-star.edu.sg.

Associate editor: Naruya Saitou

Abstract

Vertebrate genomes contain thousands of conserved noncoding elements (CNEs) that often function as tissue-specific enhancers. In this study, we have identified CNEs in human, dog, chicken, Xenopus, and four teleost fishes (zebrafish, stickleback, medaka, and fugu) using elephant shark, a cartilaginous vertebrate, as the base genome and investigated the evolution of these ancient vertebrate CNEs (aCNEs) in bony vertebrate lineages. Our analysis shows that aCNEs have been evolving at different rates in different bony vertebrate lineages. Although 78–83% of CNEs have diverged beyond recognition (“lost”) in different teleost fishes, only 24% and 40% have been lost in the chicken and mammalian lineages, respectively. Relative rate tests of substitution rates in CNEs revealed that the teleost fish CNEs have been evolving at a significantly higher rate than those in other bony vertebrates. In the ray-finned fish lineage, 68% of aCNEs were lost before the divergence of the four teleosts. This implicates the "fish-specific" whole-genome duplication in the accelerated evolution and the loss of a large number of both copies of duplicated CNEs in teleost fishes. The aCNEs are rich in tissue-specific enhancers and thus many of them are likely to be evolutionarily constrained cis-regulatory elements. The rapid evolution of aCNEs might have affected the expression patterns driven by them. Transgenic zebrafish assay of some human CNE enhancers that have been lost in teleosts has indicated instances of conservation or changes in trans-acting factors between mammals and fishes.

Key words: conserved noncoding elements, elephant shark, genome duplication, teleost fishes.

Introduction

One of the most surprising discoveries emerging from the comparative analysis of vertebrate genomes is the finding that highly conserved sequences are more often located in noncoding regions than in protein-coding sequences. There is an overrepresentation of these conserved noncoding elements (CNEs) near genes involved in the regulation of transcription and development (Sandelin et al. 2004; Shin et al. 2005; Woolfe et al. 2005; Venkatesh et al. 2006), and functional assays of hundreds of CNEs have revealed that many of them function as cis-regulatory elements (or enhancers) directing tissue-specific expression during early stages of development (Shin et al. 2005; Woolfe et al. 2005; Pennacchio et al. 2006; Navratilova et al. 2009). These findings have led to the hypothesis that CNEs define evolutionarily conserved gene regulatory networks underlying the developmental program of vertebrates (Woolfe et al. 2005; Elgar 2009). However, in contrast to their deep conservation across vertebrates spanning over 450 My of evolution, only a small number of vertebrate CNEs are identifiable in invertebrate chordates (e.g., amphioxus) (Holland et al. 2008; Putnam et al. 2008), and they are extremely rare in nonchordate invertebrates (Woolfe et al. 2005; Venkatesh et al. 2006). In addition, jawless vertebrates such as the sea lamprey (Petromyzon marinus) contain only a subset of the CNEs present in jawed vertebrates, and the lamprey CNEs are typically shorter and less conserved than CNEs in jawed vertebrates (McEwen et al. 2009). Thus, a massive expansion in the number of CNEs seems to have occurred during the evolution of jawed vertebrates from jawless vertebrates. In addition to the CNEs that existed before the divergence of jawed vertebrates, thousands of novel lineage-specific CNEs have been recruited in different jawed vertebrate lineages at various stages of evolution. For example, about 20% of eutherian CNEs were lineage-specific, recruited only after the divergence of eutherian and metatherian lineages (Mikkelsen et al. 2007). More importantly, a significant number of these eutherian-specific CNEs were found to overlap experimentally identified functional elements and were associated with developmental genes, indicating that these lineage-specific CNEs are also likely to be bona fide functional elements (Mikkelsen et al. 2007). Although comparisons of vertebrate genomes have uncovered evolution of lineage-specific CNEs in various bony vertebrate lineages, very little is known about the evolution of ancient CNEs that were present in the common ancestor of jawed vertebrates. We term the CNEs that were present in the common ancestor of jawed vertebrates as “ancient vertebrate CNEs” (aCNEs). Comparisons of CNEs in Hox cluster loci of human, teleost fish (fugu and zebrafish), and cartilaginous fishes (horn shark and elephant shark) have indicated that many aCNEs have diverged beyond recognition in teleost fishes (Chiu et al. 2004; Venkatesh et al. 2006; Ravi et al. 2009). In a genome-wide comparison, we previously discovered that the human genome shares more CNEs with elephant shark (4,782 CNEs), a basal jawed
vertebrate, than with fugu (2,107 CNEs) or zebrafish (2,838 CNEs) (Venkatesh et al. 2006). This indicated that the repertoire of aCNEs is larger than that previously predicted based on comparison of human and teleost fish genomes and that fishes have retained only a minority of aCNEs. However, because the human genome was used as the base genome in the previous study, it was not possible to predict a comprehensive set of aCNEs and analyze their evolutionary pattern in different bony vertebrate lineages. Because aCNEs represent a key set of evolutionarily constrained functional elements, it is important to investigate the lineage-specific evolution of aCNEs in vertebrates.

In this study, we have identified aCNEs in the genomes of human, dog, chicken, *Xenopus tropicalis*, and four teleost fishes (zebrafish, stickleback, medaka, and fugu) using elephant shark (*Callorhinus milii*) as the base genome and investigated CNE evolution in bony vertebrate lineages. We found that while 24–55% of aCNEs have diverged beyond recognition (“lost”) in tetrapods, 78–83% of aCNEs have been lost in teleost fishes. Additionally, the aCNEs in teleost fishes were observed to be evolving at a higher rate than their orthologs in tetrapods, providing an explanation for the loss of a large number of aCNEs in teleosts. Teleost fishes contain duplicate copies of many genes that are single-copy in mammals due to a fish-specific whole-genome duplication in the ray-finned fish lineage. Our analysis showed that a similar proportion of aCNEs were lost in both singleton and duplicated genes, suggesting a relaxed constraint on both copies of duplicated CNEs.

**Materials and Methods**

**Identification of Ancient Vertebrate CNEs**

To identify CNEs in various bony vertebrates, contiguous MegaBLAST was run on repeat-masked genome sequences of elephant shark (*Callorhinus milii*) as the base genome and investigated CNE evolution in bony vertebrate lineages. From this set of CNEs, the set of aCNEs that are conserved in all vertebrates (“panvertebrate”) was identified. Functional categories that were statistically overrepresented among human genes associated with the panvertebrate aCNEs (nearest flanking gene in Ensembl release 54 human gene set) were identified using the binomial statistics tool (using Entrez gene identifiers and Bonferroni correction for multiple testing) in the PANTHER classification system (Thomas et al. 2006). Two gene lists were used as references: all genes in the human genome and only genes associated with aCNEs.

**Computation of Substitution Rate of Ancient Vertebrate CNEs**

To determine the substitution rates of aCNEs, multiple global alignments of panvertebrate aCNEs were obtained using DIALIGN (Morgenstern 1999) and concatenated. Gaps were removed from the alignment and a neighbor-joining tree was generated using MEGA4 (Tamura et al. 2007) with 1,000 bootstraps, Kimura 2-parameter model, and uniform rates among sites. The evolutionary rates of panvertebrate aCNEs were compared for every pair of bony vertebrates using Tajima’s relative rate test (Tajima 1993), with elephant shark as the outgroup. The results were plotted as a Hasse diagram, in which every edge between two nodes (species) represents a statistically significant Tajima’s relative rate test ($P < 0.01$), implying a significant difference in nucleotide substitution rates between the two species (Prohaska et al. 2008). Bonferroni correction was applied to the $P$ values to correct for multiple testing.

**Identification of Singleton and Duplicate Teleost Fish Genes**

To identify single-copy (singleton) and duplicate genes in zebrafish, stickleback, medaka, and fugu, we downloaded the orthologs of all protein-coding genes in these teleost fishes, together with their intraspecies paralogs from Ensembl Compara release 54 using BioMart (Vilella et al. 2009). Scripts were written to scan through the data files. Orthologs described as “apparent_ortholog_one2one” were eliminated because these ortholog relationships are possibly the result of lineage-specific duplication events. To obtain singletons present in all four fishes, we eliminated duplicate genes that were the result of fish-specific whole-genome duplication (derived from duplication nodes at “Clupeocephala,” which is the base taxon of all teleost fishes in Ensembl Compara) or lineage-specific duplications thereafter. A total of 6,724 genes were determined to be singletons in all four fishes (supplementary table 1, Supplementary Material online). To obtain duplicated genes in each of the four fishes, we focused on intraspecies fish gene pairs that arose from the fish-specific whole-genome duplication and have not experienced any further duplication following the split from other fishes. A total of 1,978 gene pairs in zebrafish, 2,189 in stickleback,
1,929 in medaka, and 2,011 in fugu fulfilled these criteria (supplementary table 2, Supplementary Material online).

Proportions of CNE Retention and Loss in Singleton and Duplicated Fish Genes

For each of the four fishes, we investigated the relative proportions of CNE retention and loss in singleton and duplicated genes. We searched for the 2,677 aCNEs that were present in at least one of the four teleost fishes and therefore deemed to be present in their last common ancestor.

We associated these aCNEs with their nearest flanking gene in each of the four fishes. For each case of aCNE retained in a particular fish, we determined if the associated gene is singleton or duplicated. For each case of aCNE lost in a particular fish, we postulated the identity of its associated gene and whether it is singleton or duplicated, by a majority vote (minimum 60%) of the genes associated with this aCNE in other fishes. aCNEs that could not be annotated (located on scaffolds without gene annotation), could not be reliably assigned to a unique gene by orthology, or were deemed to be missing from fish genes that have also been lost were eliminated.

Overlap between Experimentally Verified Enhancers and aCNEs

The human noncoding elements tested in transgenic mice at 11.5 days post-coitum (dpc) were downloaded from VISTA Enhancer Browser on 9 February 2010 (Visel et al. 2007). Their coordinates were converted from human genome assembly NCBI35 (hg17) to NCBI36 (hg18) using the “liftOver” tool (http://genome.ucsc.edu/cgi-bin/hgLiftOver). The overlap (minimum 50 bp) of our elephant shark–human CNEs with the tested elements was carried out by comparing the genomic coordinates on the NCBI36 assembly.

Functional Assay of Human CNEs in Transgenic Zebrasfish

Zebrafish (AB strain) were reared following standard methods (Westerfield 2007). Primer sequences to amplify six human enhancer elements tested in our study (elephant shark–human CNEs) (table 1). The tetrapod CNEs are also longer (average length 207–226 bp) than the teleost CNEs (average length 175–179 bp). The average percentage identity of CNEs in all species was 82%. Among tetrapods, the two mammals contain a similar number of CNEs, whereas chicken contains Japan) and photographed using an attached digital microscope camera (DP70; Olympus). Pigmentation that would otherwise obscure GFP expression detection in certain anatomical structures was inhibited by maintaining zebrafish embryos in 0.003% N-phenylthiourea (PTU) (Sigma-Aldrich, Sweden) from 8 hpf onwards. The GFP expression pattern observed in at least 15% of transgenics was considered as specific expression driven by the CNE. Such transient transgenics were reared to maturity and mated with wild-type fish, and the expression of GFP was analyzed in stable transgenic lines (G1s).

Predicted Target Genes of CNEs

In the absence of experimental evidence, it is difficult to establish the target genes of CNEs because enhancers are capable of acting over large distances. In this study, we have designated the genes closest to CNEs as their predicted target genes. Akalin et al. (2009) predicted target genes of CNEs based on the presence of genomic regulatory blocks (GRBs) that show conserved synteny in mammals and teleost fishes. Their list of GRBs includes two of the six human enhancer elements tested in our study (elements #702 and #1,043), and the target genes predicted by them for these two elements are the same as those predicted by us (supplementary table 3, Supplementary Material online).

Table 1. CNEs identified in various bony vertebrates using elephant shark as the base genome.

<table>
<thead>
<tr>
<th>Vertebrate</th>
<th>Number of CNEs</th>
<th>Total Length (kb)</th>
<th>Average Length (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>5,152</td>
<td>1,167</td>
<td>226</td>
</tr>
<tr>
<td>Dog</td>
<td>5,231</td>
<td>1,161</td>
<td>222</td>
</tr>
<tr>
<td>Chicken</td>
<td>6,489</td>
<td>1,439</td>
<td>222</td>
</tr>
<tr>
<td>Xenopus</td>
<td>3,862</td>
<td>800</td>
<td>207</td>
</tr>
<tr>
<td>Zebrasfish</td>
<td>1,936</td>
<td>339</td>
<td>175</td>
</tr>
<tr>
<td>Stickelback</td>
<td>1,500</td>
<td>269</td>
<td>179</td>
</tr>
<tr>
<td>Medaka</td>
<td>1,506</td>
<td>264</td>
<td>175</td>
</tr>
<tr>
<td>Fugu</td>
<td>1,465</td>
<td>262</td>
<td>179</td>
</tr>
</tbody>
</table>

Results

Ancient Vertebrate CNEs in Jawed Vertebrates

We used elephant shark as the base genome and identified aCNEs in key tetrapod (human, dog, chicken, and Xenopus tropicalis) and teleost fish (zebrasfish, stickelback, medaka, and fugu) genomes. Because the current elephant shark sequence assembly covers only 75% of its genome (Venkatesh et al. 2007), these CNEs should represent approximately 75% of the CNEs shared between the whole-genome sequences of elephant shark and these bony vertebrates. Overall, the four tetrapods contain double or more CNEs (3,862–6,489 CNEs) found in teleost fishes (1,465–1,936 CNEs) (table 1). The tetrapod CNEs are also longer (average length 207–226 bp) than the teleost CNEs (average length 175–179 bp). The average percentage identity of CNEs in all species was 82%. Among tetrapods, the two mammals contain a similar number of CNEs, whereas chicken contains...
25% more and *Xenopus* contains 25% fewer CNEs than mammals, indicating a distinct evolutionary pattern of CNEs in these three major lineages of tetrapods. Among teleost fishes, zebrafish (an ostariophysan) contains ~30% more CNEs than the three acanthopterygians (stickleback, medaka, and fugu). It is interesting that chicken, whose genome (1,100 Mb) is smaller than that of human (3,100 Mb), dog (2,500 Mb), *Xenopus* (1,500 Mb), and zebrafish (1,600 Mb) and slightly larger than that of medaka (870 Mb), contains more aCNEs than other bony vertebrates. This reflects the relatively low rate of divergence of CNEs in chicken compared with other vertebrates. To verify if the slow evolutionary rate of CNEs is unique to chicken or common to birds, we identified CNEs in the recently completed genome of zebra finch and found that it also contains a high number of aCNEs (6,501 CNEs; total length 1,417 kb; average length 218 bp). Bird genomes are known to have experienced a slower rate of rearrangements than mammals, and the karyotype of the ancestral amniote has been predicted to be more similar to that of birds (International Chicken Genome Sequencing Consortium 2004; Bourque et al. 2005; Ellegren 2010). Phylogenetic analysis of protein-coding sequences in vertebrates has indicated that the nucleotide substitution rate is lower in chicken than in other tetrapods and teleost fishes (Wang et al. 2009). Our analysis of aCNEs has now indicated that highly constrained noncoding sequences have also been evolving slowly in birds. Among the tetrapods, aCNEs have been diverging at the highest rate in the *Xenopus* lineage. A previous analysis of CNEs in the *HoxA* cluster has shown that CNEs in *Xenopus* have been experiencing a higher rate of nucleotide substitutions than those in mammals (Wagner et al. 2004). However, the reason(s) for this increased rate of nucleotide substitutions in *Xenopus* is unknown.

To estimate the number of aCNEs in the common ancestor of these jawed vertebrates, we formed a union of CNEs (see Materials and Methods) found in all these vertebrates. A total of 8,458 CNEs (combined length of 1.82 Mb) were obtained. The evolutionary distribution of these aCNEs (fig. 1) indicates that the CNEs recruited in the common ancestor of jawed vertebrates have not remained static in any of the vertebrate lineages and that they have been diverging in a lineage-specific manner. Consequently, tetrapods have retained 45–76% of aCNEs, whereas the four teleost fishes have retained only 17–22% of aCNEs. Thus, a majority of aCNEs have diverged beyond recognition (“lost”) in teleost fishes. Indeed, 68% of the aCNEs (5,781/8,458) were lost before the diversification of the four teleost fish lineages. In contrast, only 7% (566/8,458) of aCNEs had been lost in the tetrapod ancestor. The aCNEs have continued to diverge at a relatively high rate in the four teleost lineages. Of the 2,677 CNEs that were present in their common ancestor, 30% (795/2,677) have diverged beyond recognition in zebrafish, whereas ~46% (1,218–1,264/2,677) have disappeared in stickleback, medaka, and fugu during the last ~150 My of evolution. In the tetrapod lineage, of the 7,892 aCNEs retained in the common ancestor, ~35% (2,747–2,799/7,892) have diverged in the mammalian lineage in the last 330 My of evolution, whereas only 18% (1,455/7,892) have diverged in the chicken lineage. In contrast to the mammalian and chicken lineages, the CNEs in the *Xenopus* lineage have experienced an accelerated rate of evolution during this period with nearly 52% (4,088/7,892) diverging beyond recognition.

**Panvertebrate aCNEs**

It is interesting that despite this continued trend of divergence of aCNEs across vertebrate lineages, a set of 557 aCNEs have remained highly conserved in all vertebrates examined in this study (supplementary table 4, Supplementary Material online). These elements represent a core set of CNEs that is likely to be conserved in most jawed vertebrates (panvertebrate) and is expected to be fundamental to the biology of these vertebrates. In the human genome, the panvertebrate aCNEs are associated with 246 genes (supplementary table 5, Supplementary Material online). Statistical analysis of their PANTHER gene ontology functional categories (Thomas et al. 2006) compared with a reference set of all genes in the human genome revealed that these genes are predominantly involved in transcription and transcriptional regulation (supplementary fig. 1, Supplementary Material online). These categories remained significantly overrepresented when the reference
set consisted of only CNE-associated genes (fig. 2). This suggests that the genes associated with panvertebrate aCNEs are a distinctively important group of genes regulating gene transcription in vertebrates.

Nucleotide Substitution Rates in Ancient Vertebrate CNEs

To gain insights into the evolutionary rate of aCNEs in bony vertebrate lineages, we estimated the nucleotide substitution rates in aCNEs using elephant shark as the outgroup. For this analysis, we chose the panvertebrate aCNEs, concatenated their sequences in each jawed vertebrate and generated a neighbor-joining tree. As can be seen by their longer branch lengths (fig. 3), the aCNEs in teleost fishes have been collectively experiencing a higher evolutionary rate. The average number of substitutions per site in teleost fishes (0.12 substitutions per site) is twice that of the tetrapods (0.06 substitutions per site). Among tetrapods, the substitution rate is the highest for *Xenopus*, which is consistent with the loss of the highest number of aCNEs in *Xenopus* compared with other tetrapods. These data support the notion that the loss of aCNEs in bony vertebrates is largely due to the rapid divergence of their sequences rather than due to other potential causes, such as deletions, rearrangements, and microinversions.

To obtain a more quantitative measure of the differences in nucleotide substitution rates in the aCNEs, we carried out Tajima’s relative rate tests of CNE substitution rates between all pairs of bony vertebrates using elephant shark as the outgroup. The summary of the relative rates is presented in figure 4 as a Hasse diagram (detailed results in supplementary table 6, Supplementary Material online). The results confirm that CNEs in tetrapods have significantly lower substitution rates than those in teleost fishes, with chicken CNEs experiencing the lowest rate of substitutions among all the bony vertebrates. The substitution rate of CNEs in mammals is intermediary to the rates in chicken and *Xenopus*, which is in close agreement with the number of aCNEs lost in these tetrapods. Among teleost fishes, zebrasfish CNEs have experienced the highest rate of substitution, whereas stickleback CNEs have experienced the lowest rate of substitution (fig. 4). The substitution rates of CNEs in fugu and medaka are intermediary to those in zebrafish and stickleback. It is paradoxical that although zebrasfish has retained the highest number of aCNEs among teleost fishes, its CNEs have been evolving at a higher rate than those in other teleosts.

Divergence of Ancient Vertebrate CNEs in Single-Copy and Duplicate Fish Genes

Our analysis shows that 68% of aCNEs had disappeared in the ray-finned fish lineage before the divergence of zebrafish (subdivision Otocephala) and the three acanthopterygians (stickleback, medaka, and fugu; subdivision Euteleostei). These fishes belong to very distantly related taxonomic groups of teleosts, and their subdivisions together include ~95% of living teleosts (Nelson 2006). The disappearance of 68% of aCNEs in the common ancestor of these fishes implies that the majority of the aCNEs had diverged before the diversification of teleost fishes. There is now convincing evidence that a whole-genome duplication occurred in the ray-finned fish lineage before the diversification of teleost fishes (Christoffels et al. 2004; Hoegg et al. 2004; Jaillon et al. 2004; Vandepoele et al. 2004; Crow et al. 2006). Although our data are unable to indicate if the loss of the large number of aCNEs occurred before or after this whole-genome duplication, it seems likely that the whole-genome duplication allowed a relaxed constraint on the duplicated aCNEs leading to the rapid divergence of their sequences.

To investigate whether aCNEs were preferentially lost from genes retained in duplicate because of their redundancy, we analyzed the pattern of CNE loss in duplicate and singleton genes. We focused on CNEs that were present in the common ancestor of the four teleost fishes (2,677 CNEs) and assigned them to their nearest flanking genes. We then identified a confident set of duplicate and singleton genes (see Materials and Methods) and determined the proportion of CNEs that were lost in these
genes. With regard to singleton genes, we restricted the analysis to genes that are singleton in all the four teleost fishes. By parsimony, such genes are likely to have become singletons before the divergence of the four species and hence are ideal candidates for tracing the divergence of aCNEs associated with singleton genes. The lists of singleton and duplicate genes identified in the four fish genomes are given in supplementary tables 1 and 2, respectively (Supplementary Material online). Our analyses show that a similar percentage of aCNEs have diverged beyond recognition in singleton (38–41%) and duplicate genes (39–42%) in stickleback, medaka, and fugu. However, in zebrafish, duplicate genes have lost a higher percentage (34%) of aCNEs than singleton genes (27%) (table 2 and fig. 5). Nevertheless, the overall trend is that a similar proportion of aCNEs have been lost in duplicate and singleton genes of teleost fishes. This implies that aCNEs in both duplicate and singleton genes have been diverging at the same rate.

Biological Relevance of Ancient Vertebrate CNEs

To investigate the biological relevance of aCNEs, we determined the overlap of aCNEs in human with experimentally verified human enhancers (cis-regulatory elements). The VISTA Enhancer Browser Project (Visel et al. 2007) has tested a total of 1,255 human noncoding elements in transgenic mice at embryonic stage 11.5 dpc. Eighty-eight percent of these elements (1,101/1,255) are conserved in other tetrapods (mouse, chicken, and Xenopus) and/or in teleost fishes (fugu and zebrafish). Of the 1,255 elements assayed, 48% of the elements (605/1,255) were found to function as tissue-specific enhancers in transgenic mice. We found that among the 5,093 aCNEs in human, 642 overlap (>50 bp) of these CNEs are part of 307 tissue-specific enhancers (supplementary table 7, Supplementary Material online). These overlaps between aCNEs and functional enhancers indicate that aCNEs are rich in cis-regulatory elements that function as tissue-specific developmental enhancers.

Interestingly, of the 358 ancient human CNEs that overlapped experimentally verified enhancers, 62 aCNEs (located within 50 enhancers) have diverged beyond recognition in the four teleost fishes analyzed in our study (category I elements in supplementary table 7, Supplementary Material online). This indicates that even aCNEs that have diverged beyond recognition in teleost fishes are potential cis-regulatory elements in human. If a significant number of aCNEs are indeed functional cis-regulatory elements, what is the consequence of the divergence of their sequences in teleost fishes? Are the transcription factors in teleost fishes still able to recognize such ancient cis-regulatory elements? To address the latter question, we selected six experimentally verified human enhancers that are conserved in tetrapods but have diverged beyond recognition in teleost fishes and tested their function in transgenic zebrafish.

Functional Assay of Ancient Vertebrate CNEs in Zebrafish

We generated transgenic zebrafish for each of the enhancers, evaluated the G0 embryos for GFP expression, and generated stable transgenic lines. In zebrafish embryos, formation of all somites occurs within 25 hpf, whereas in mouse embryos, somite formation is completed by 14 dpc. In addition, other embryonic developments such as formation of neural plate, neuromeres corresponding to telencephalon and diencephalon, optic primordium, otic placode, and regionalization of the heart also occur within 24 hpf of zebrafish embryos. These developments take place within 9.5 dpc of mouse embryos. Thus, the 72 hpf stage zebrafish embryos should cover the 11.5 dpc stage of mouse embryos. We therefore evaluated GFP expression in zebrafish embryos up to 72 hpf. Of the six enhancers tested, two enhancers (elements #901 and #1,114) were able to reproduce their mouse expression patterns completely in transgenic zebrafish (fig. 6B and C; supplementary table 3, Supplementary Material online). Element #901 was able to recapitulate its mouse midbrain and hindbrain expression in zebrafish (fig. 6B). Element #1,114 that drove expression in the midbrain, hindbrain, and ventral neural tube of mouse embryos was able to drive expression in the zebrafish midbrain, hindbrain, and neural tube (fig. 6C). The expression pattern of this element in zebrafish partially overlaps with the expression domains of its predicted target gene in mouse and zebrafish (supplementary
These results suggest that although these enhancers have diverged beyond recognition in zebrafish, zebrafish trans-acting factors are able to recognize them and drive expression patterns similar to the mouse trans-acting factors. Two of the human enhancers tested (elements #567 and #1,358) were able to recapitulate their mouse expression patterns only partially in zebrafish (fig. 6A and D; supplementary table 3, Supplementary Material online). Element #567 that drove expression in the mouse midbrain and ventral neural tube could drive expression only in the floor plate (ventral-most part of the neural tube) of zebrafish (fig. 6A), whereas element #1,358, which showed expression in a wider region of the mouse forebrain, showed a restricted expression in the zebrafish telencephalon (fig. 6D). The expression pattern of element #1,358 partially recapitulated the expression domains of its predicted target gene in mouse and zebrafish (supplementary table 3, Supplementary Material online). The remaining two human enhancers (elements #702 and #1,043) were inactive in zebrafish until 72 hpf (data not shown). The partial or total failure of the human enhancers to reproduce their mouse expression patterns in zebrafish reflects changes in the upstream regulatory transcription factors and/or other trans-acting factors in zebrafish.

Discussion

In this study, we have used a cartilaginous fish, elephant shark, as base genome and identified CNEs in human, dog, chicken, *Xenopus*, and four teleost fish genomes. Based on the number of CNEs shared by elephant shark and these bony vertebrates, we predict that their last common ancestor contained at least ~8,500 CNEs (total length of 1.8 Mb).
Investigation of the evolutionary patterns of these CNEs in various bony vertebrate lineages revealed their numbers have not been static and that they have been diverging and evolving at different rates in different bony vertebrate lineages. The most striking finding is that about 80% of them have diverged beyond recognition (“lost”) in teleost fishes, whereas 24–55% have been lost in various tetrapods. Estimation of the substitution rates in aCNEs showed that the CNEs in teleost fishes have been evolving at a significantly higher rate than those in tetrapods. Previous studies have shown that another category of conserved elements, the mammalian ultraconserved elements (UCEs; identical over 100 bp or longer in at least three placental mammals), have also been evolving at a higher rate in teleost fishes (Stephen et al. 2008; Wang et al. 2009). However, it should be noted that the mammalian UCEs were recruited at various stages of jawed vertebrate evolution and were subjected to constraint only during recent evolutionary period. In contrast, the aCNEs are evolutionarily constrained elements that existed in the common ancestor of jawed vertebrates and have been evolving at different rates in different bony vertebrate lineages.

Our analysis indicated that in the ray-finned fish lineage, a majority of aCNEs (~70%) were in fact lost before the diversification of teleost fishes. With about 27,000 species, teleost fishes are the largest and most diverse group of vertebrates. In contrast, the basal nonteleost ray-finned fishes are represented by only about 50 living species (Nelson 2006). The fish-specific whole-genome duplication that occurred in the ancestor of teleost fishes has been proposed to be responsible for the diversification of teleost fishes (Hoegg et al. 2004; Meyer and Van de Peer 2005; Crow et al. 2006; Santini et al. 2009). A whole-genome duplication event allows relaxed constraint on one or both copies of duplicated genes resulting in loss of a large number of duplicated genes and an asymmetric rate of evolution of genes retained in duplicate (Lynch and Conery 2000; Otto 2007; Semon and Wolfe 2007). Consistent with this prediction, analysis of the evolutionary rate of protein-coding genes has indicated that both singleton and duplicate genes in teleost fishes have been evolving at a faster rate than their orthologs in mammals (Jaillon et al. 2004; Brunet et al. 2006; Steinke et al. 2006). It is therefore likely that the fish-specific genome duplication might have triggered an accelerated rate of nucleotide substitution in teleosts resulting in rapid divergence of protein-coding sequences and aCNEs.

Previous studies of CNEs predicted based on alignment of human and teleost fish genomes have shown that many teleost genes retained in duplicates have lost one copy of their duplicate CNEs in a manner whereby the retained single-copy CNEs are partitioned between the duplicate genes (Santini et al. 2003; Postlethwait et al. 2004; Woolfe and Elgar 2007; Jovelin et al. 2010; Navratilova et al. 2010). This pattern of loss of duplicate CNEs is consistent with the duplication–degeneration–complementation model (Force et al. 1999) and provides evidence for subfunction-alization of CNEs between the duplicated genes. However, our analysis of duplicated aCNEs has indicated that both copies of a large number of duplicated aCNEs have been lost in teleost fishes. Thus, this is an unusual instance of duplication followed by large-scale degeneration of both copies of duplicated elements.

The higher substitution rate in some aCNEs could be indicative of positive selection acting on these CNEs. In the case of zebrafish, it is of particular interest that although it has retained more aCNEs than other fishes, the nucleotide substitution rate is significantly higher than that in other fishes. It is possible that some of these highly divergent CNEs in zebrafish are the result of adaptive evolution. Comparisons of evolutionary rate of CNEs conserved among mammals have indicated that nearly one-third of the CNEs show accelerated or decelerated rates of substitutions in some lineages and that a subset of them have even evolved significantly faster than the local neutral rate in their neighborhood, providing strong evidence for adaptive evolution (Kim and Pritchard 2007). In another study, a subset of human CNEs was found to have accumulated a higher number of nucleotide substitutions than in other mammals and that a few of them were the result of positive selection acting in the human lineage (Pollard et al. 2006). Another possibility for the increased nucleotide substitution rate in aCNEs is gene conversion during meiotic recombination that results in biased propagation of GC alleles compared with AT alleles (Galtier and Duret 2007). However, our comparisons of distantly related genomes are less susceptible to the influence of biased gene conversion because recombination hotspots tend to be short-lived (Ratnakumar et al. 2010). Hence, the effect of biased gene conversion on aCNEs is considered to be minimal. The highly divergent CNEs in zebrafish, and other teleost fishes are potential candidates for functional studies of adaptive evolution.

The overlap of a significant number of functionally verified human enhancers with aCNEs indicate that these CNEs are rich in cis-regulatory elements that direct tissue-specific expression during embryonic development. Thus, it would be interesting to examine the potential consequences of their loss or divergence in different bony vertebrate lineages. Cis-regulatory elements typically comprise clusters of transcription factor binding sites (TFBS), often with multiple sites for the same transcription factor (Lifanov et al. 2003; Gotea et al. 2010). Even subtle changes that affect the TFBS (Anand et al. 2003; Tumpel et al. 2006) and/or the spacing and phasing of TFBS (Saráfova and Siu 2000) can alter or abrogate the element’s expression pattern. In fact, duplicated CNEs in zebrafish that have undergone divergent evolution (but still can be recognized as CNEs) have been shown to drive different patterns of expression in transgenic zebrafish (Navratilova et al. 2010). Therefore, it is possible that the cis-regulatory CNEs that have diverged beyond recognition or have accumulated a high number of substitutions in teleost fishes and other vertebrates have either become inactive or acquired an altered expression pattern. In such cases, the expression patterns of their target genes may also be altered. However, we cannot rule out the possibility of creation of alternative
cis-regulatory elements to compensate for the lost CNEs, similar to the stabilizing selection observed in the stripe-2 enhancer of even-skipped gene in Drosophila (Ludwig et al. 2000). Teleost fish gene loci that have lost a large number of aCNEs are good candidates for verifying this possibility.

Another interesting aspect to the divergent CNEs in teleost fishes is whether the transcription factors and other trans-acting factors in teleost fishes are still capable of recognizing the cis-regulatory elements contained in the CNEs. To investigate this, we selected six human enhancers that overlap aCNEs conserved in human and other tetrapods but diverged beyond recognition in teleost fishes and tested them in transgenic zebrafish. The expression patterns driven by these enhancers provided some interesting insights into the trans-acting factors regulating gene transcription in teleost fishes. Of the six enhancers, two (elements #901 and #1,114) were able to recapitulate their mouse expression patterns completely in transgenic zebrafish. This indicates that the trans-acting factors responsible for the activity of these enhancers are totally conserved between zebrafish and mouse. Two other enhancers (elements #567 and #1,358) were able to reproduce only part of their mouse expression patterns, whereas two others (elements #702 and #1,043) were inactive in zebrafish. Although we cannot exclude the possibility of different promoters used in the mouse and zebrafish reporter constructs contributing to some differences in expression, we believe that the two promoters used act very similarly based on our observations for elements #901 and #1,114. The other possibility is that there are some changes in trans-acting factors between mouse and zebrafish. The changes could have occurred in the transcription factors, either in protein sequence affecting DNA-binding affinity or protein–protein interactions, or expression pattern affecting the time, domain, or level of expression. The changes could also have occurred in the cellular environment or chromatin state in zebrafish. Such changes have been previously proposed to be responsible for differences in transcriptional regulation between species (Schmidt et al. 2010). Nevertheless, the loss of expression in transgenic zebrafish is unlikely to be due to the chromatin state as transgene integration is generally random and no expression was detected in the large number of G0 transgenic zebrafish examined by us. A recent comparison of enhancer activities of 13 zebrafish CNEs and their human orthologs in transgenic zebrafish and mouse found that different expression patterns of 39% of CNEs in zebrafish or mouse were due to trans-changes (Ritter et al. 2010). These results indicate that even regulatory sequences that are conserved in mammals and fish are not interpreted in the same way in mouse and zebrafish and that trans-changes are far more common between mammals and fish than previously thought.

Supplementary Material

Supplementary tables 1–8 and figures 1–5 are available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the Biomedical Research Council of A*STAR, Singapore. We thank Axel Visel and Len Pennacchio for testing the human enhancer elements and allowing the use of mouse embryo images from VISTA Enhancer Browser. We also thank Shawn Hoon, Eddie Loh, Jeremy Parsons, and Vyedianathan Ravi for useful suggestions on the manuscript. Finally, we thank the zebrafish and stickleback genome communities for sequencing and making available the zebrafish and stickleback genome sequences in the public domain.

References


