LETTERS

Large Number of Ultraconserved Elements Were Already Present in the Jawed Vertebrate Ancestor

Jianli Wang,[1] *Alison P. Lee,*[1] *Rimantas Kodzius, Sydney Brenner, and Byrappa Venkatesh*

Institute of Molecular and Cell Biology, A*STAR, Biopolis, Singapore

Stephen (2008) identified 13,736 ultraconserved elements (UCEs) in placental mammals and investigated their evolution in opossum, chicken, frog, and fugu. They found that there was a massive expansion of UCEs during tetrapod evolution and the substitution rate in UCEs showed a significant decline in tetrapods compared with fugu, suggesting they were exapted in tetrapods. They considered it unlikely that these elements are ancient but evolved at a higher rate in teleost fishes. In this study, we investigated the evolution of UCEs in a cartilaginous fish, the elephant shark and show that nearly half the UCEs were present in the jawed vertebrate ancestor. The substitution rate in UCEs is higher in fugu than in elephant shark, and approximately one-third of ancient UCEs have diverged beyond recognition in teleost fishes. These data indicate that UCEs have evolved at a higher rate in teleost fishes, which may have implications for their vast diversity and evolutionary success.

Mammalian genomes contain thousands of highly conserved sequences that are under evolutionary selection. Surprisingly, a major portion of them are located in the noncoding regions of the genomes. Functional assays of such conserved noncoding sequences have indicated that many of them function as transcriptional regulatory elements (Shin et al. 2005; Woolfe et al. 2005; Pennacchio et al. 2006). Because the amount of noncoding sequences shows a broad correlation with the developmental complexity of eukaryotes, it has been hypothesized that noncoding sequences have played an important role in the evolution of complex developmental programs and phenotypic diversity of organisms (Taft et al. 2007). Therefore, investigations of the evolution of the conserved mammalian sequences in other vertebrate lineages should provide useful insights into the origin and the role of these elements in the evolution of phenotypic diversity of vertebrates. In a recent study, Stephen et al. (2008) identified 13,736 ultraconserved elements (UCEs) in the human genome that are identical over $\geq$100 bp in at least three out of five placental mammals (human, mouse, rat, dog, and cow) and investigated the evolution of these sequences in opossum, chicken, frog, and fugu genomes. They found that about 40% of these eutherian UCEs were present before the speciation of ray-finned fishes; 30% appeared first in the tetrapod ancestor; 18% arose in the amniote ancestor; and 10% evolved in the therian ancestor (fig. 1). In addition, they found that the substitution rate in the UCEs that first appeared in ray-finned fishes reduced significantly in the frog and chicken lineages compared with the fish lineage. It was therefore hypothesized that the UCEs were exapted to perform novel functions in tetrapods. The alternative hypothesis that these elements were present in the ancestral vertebrate and have evolved rapidly in teleost fishes was rejected because it is less parsimonious.

In this study, we investigated the presence of eutherian UCEs in a cartilaginous fish, the elephant shark (*Callorhinchus milii*), for which a 1.4× coverage genome sequence was recently generated (Venkatesh et al. 2007). Cartilaginous fishes are the oldest group of living jawed vertebrates that diverged from bony vertebrates (ray-finned fishes, coelacanths, lungfishes, and tetrapods) ~528 MYA (Hedges and Kumar 2003). Surprisingly, our study showed that a majority of UCEs were in fact present in the common ancestor of jawed vertebrates and that the ancient UCEs have accumulated substitutions at a higher rate in teleost fishes, so much so that one-third of them have diverged beyond recognition.

To identify the orthologs of eutherian UCEs in the elephant shark, we aligned the human genome and five other vertebrate genomes (opossum, chicken, frog, fugu, and elephant shark) using MULTIZ (Blanchette et al. 2004). Altogether orthologs for 5,677 (41.3%) UCEs could be identified in elephant shark (table 1). Because the 1.4× assembly of the elephant shark represents approximately 75% of the genome, we estimate that the whole genome of the elephant shark contains ~7,600 UCEs (55%). Interestingly, although >90% of the UCEs identified in the elephant shark are conserved in frog, chicken, and opossum, ~35% (2,009) of them are absent in fugu. Almost a similar number of UCEs are also not identifiable in the stickleback (1,802) and medaka (2,076) genomes. Thus, about one-third of ancient UCEs have evolved at a significantly higher rate and diverged beyond recognition or are lost in the teleost fish lineage.

In the absence of data from a cartilaginous fish, Stephen et al. had reported that 39.3% of UCEs (5,404) are present in the ray-finned fish, and 30.3% of UCEs subsequently appeared in the tetrapod ancestor. However, with the identification of 41.3% of UCEs in the elephant shark, we found that only 12.6% (1,736) UCEs first appeared in ray-finned fishes, and 18.3% (2,508) first appeared in the tetrapod ancestor. The proportion of UCEs that first appeared in the amniote ancestor is lowered from 18.5% to 16.1% (fig. 1). Thus, the inclusion of the elephant shark in the analysis shows that a major proportion of the UCEs were already present in the jawed vertebrate ancestor, and a lower proportion of UCEs were recruited in the bony vertebrate and tetrapod ancestors. Nevertheless, this finding is consistent with the notion that nearly half of eutherian UCEs were recruited during the evolution of different lineages of tetrapods and that most UCEs have expanded in

[1] These authors contributed equally to this work.

Key words: ultraconserved elements, elephant shark, teleost fish.

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

Mol. Biol. Evol. 26(3):487–490. 2009
doi:10.1093/molbev/msn278
Advance Access publication December 3, 2008

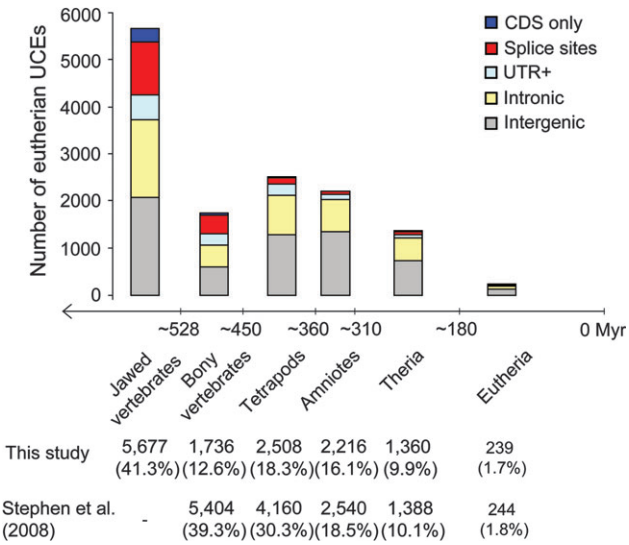


FIG. 1.—Distribution of UCEs reported by Stephen et al. (2008) and this study. No cartilaginous fish was analyzed in the former study, whereas the elephant shark was included in this study. The vertical bars show the genomic distribution of UCEs identified in this study. The estimated divergence times on the *x*-axis are based on molecular data (Hedges and Kumar 2003).

size. This study also shows that majority of the UCEs located within the protein-coding sequences are indeed ancient and were present in the common ancestor of jawed vertebrates, and confirms Stephen et al.'s finding that the new UCEs were added mainly in the noncoding (intergenic and intronic) sequences (fig. 1).

The substitution rate in UCEs in fugu is unusually higher compared with other vertebrates as shown by the extended branch length for fugu in a Neighbor-Joining tree built on ancient UCEs in all six species (fig. 2*A* and *C*). However, the higher evolutionary rate of UCEs in fugu seems to be the unique property of the UCEs and not a general property of the fugu genome, because the protein-coding sequences in fugu (CDSs; 271 genes, 198 kb aligned sequence) are evolving at a similar rate as in frog, chicken, and opossum (fig. 2*B* and *C*). Interestingly, the substitution rate (substitutions per site per 10^9 years) in the CDSs in the elephant shark (0.74) is lower than in all other vertebrates (fugu, 0.93; frog, 0.95; and chicken, 0.87; and opossum, 1.13). Previous studies of evolutionary rate of protein-coding sequences in cartilaginous fishes based on a few mitochondrial and nuclear genes had indicated that the rate of evolution of protein sequences in cartilaginous fishes is

Table 1
Eutherian UCEs (EU100+) Conserved in Vertebrates

	Eutherian UCEs		Total Length	Average Identity
	Number	%	(kb)	(%)
Human	13,736	100	2,131	NA
Opossum	13,184	96.0	2,035	94.7
Chicken	11,470	83.5	1,778	92.0
Frog	9,137	66.5	1,418	84.2
Fugu	5,404	39.3	773	74.1
Elephant shark[a]	5,677	41.3	837	79.8

[a] Elephant shark UCEs were identified in this study. The rest of the data are from Stephen et al. (2008).

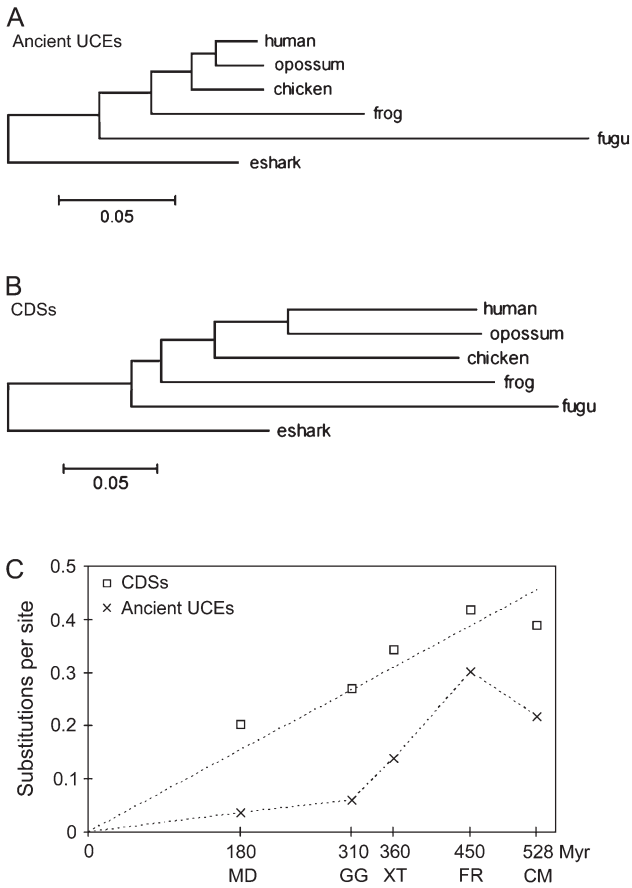


FIG. 2.—Substitution rates in UCEs and CDSs. The phylogenetic trees were generated using the Neighbor-Joining method and by specifying elephant shark as the outgroup. The standard errors for all branch lengths are less than 10^{-3}. (*A*) Phylogenetic tree built from 3,297 ancient UCEs present in the elephant shark and other vertebrates (443 kb ungapped alignment). (*B*) Phylogenetic trees built from 271 elephant shark–coding sequences ranging in length from 228 bp to 2.1 kb and their orthologs in other species (198 kb ungapped alignment). eshark, elephant shark. Linearized versions of these trees show an incorrect placement of fugu as the most basal species (see supplementary fig. 1, Supplementary Material online). (*C*) Plot of divergence times (Hedges and Kumar 2003) versus substitution distance to human as derived from the trees in (*A*) and (*B*). MD, opossum; GG, chicken; XT, frog; FR, fugu; CM, elephant shark.

significantly lower than in mammals (Martin et al. 1992; Martin 1999; Kumazawa et al. 2000). Our analysis of a much larger set of protein sequences has confirmed that protein sequences in cartilaginous fishes have indeed been evolving at a slower rate than in other vertebrates.

The distribution of conservation across UCEs shows that the fugu UCEs are less conserved than the elephant shark UCEs across their entire length, with the level of divergence being relatively higher at the periphery (fig. 3). This indicates that the divergence of fugu UCEs has occurred mainly by accumulating a higher level of substitutions at the edges. It is likely that an acceleration of the decay at the edges might have led to the inability to recognize many ancient UCEs in fugu.

The investigation of the UCEs in the elephant shark has clearly shown that a substantial number of UCEs arose before the divergence of jawed vertebrates, and at least

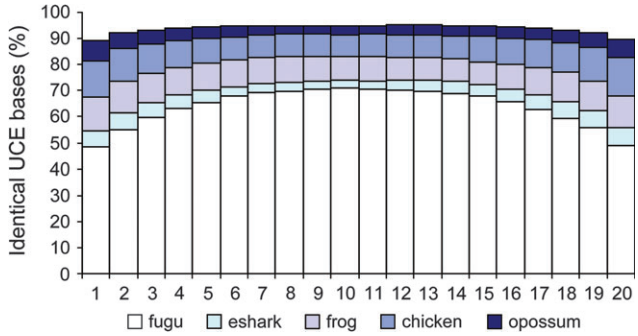


FIG. 3.—Distribution of conservation across UCEs. The eutherian UCEs conserved in each species (given in table 1) was divided into 20 equal sections and the percentage of identical bases was calculated for each section. *x*-axis, sections of UCEs; *y*-axis, percentage of identical bases.

one-third of ancient UCEs have diverged beyond recognition in teleost fishes. Furthermore, the ancient UCEs retained in teleost fishes have accumulated substitutions at an unusually higher rate compared with other vertebrates. Together, these data indicate that the UCEs have been evolving rapidly in the teleost fish lineage. The absence of a large number of ancient UCEs in teleost fishes is consistent with our previous observation that a significant number (~65%) of ancient conserved noncoding elements (>70% identical across >100 bp) present in the elephant shark and human have diverged beyond recognition in teleost fishes (Venkatesh et al. 2006). The rapid evolution of UCEs in teleost fishes could be the result of the whole-genome duplication in the ray-finned fish lineage (Christoffels et al. 2004; Jaillon et al. 2004). The genome duplication, which occurred in the teleost ancestor, might have relaxed the selective constraint on UCEs allowing them to diverge freely in teleosts. This hypothesis is supported by the observations that many duplicated genes in teleost fishes are evolving asymmetrically (Brunet et al. 2006; Steinke et al. 2006), and that the noncoding sequences in the duplicated HoxA clusters in zebrafish are less conserved compared with the noncoding sequences in the unduplicated HoxA cluster in the horn shark and human (Chiu et al. 2002). The consequences of the rapid divergence of ancient UCEs on the phenotypic evolution of teleost fishes are unclear. With about 27,000 living species, teleost fishes are the largest and most diverse group of vertebrates. The rapid evolution of UCEs might have influenced the vast divergence and speciation of teleost fishes.

Methods

Pairwise alignments of human genome (hg18 assembly) with opossum (monDom4), chicken (galGal3), frog (xenTro2), fugu (fr2), stickleback (gasAcu1), and medaka (oryLat1) genomes were obtained from the UCSC Genome Browser (Karolchik et al. 2008). Pairwise alignment of human genome and elephant shark genome (version 1; Venkatesh et al. 2007) was generated using BlastZ (Schwartz et al. 2003). To identify UCEs present in elephant shark ("eshark"), multiple alignment of human and five other vertebrate genomes was generated using MULTIZ/autoMZ (Blanchette et al. 2004) following the tree topology (((((human opossum) chicken) frog) fugu) eshark). Orthologous UCEs in elephant shark were identified using the same criteria as Stephen et al. (at least 20 alignable nucleotides). UCEs were annotated using the UCSC RefSeq gene track (dated August 20, 2008).

To build a phylogenetic tree that measures UCE substitution rates across six vertebrates, multiple alignments of the ancient UCEs were concatenated, and gapped columns were discarded. The multiple alignment was used to construct a Neighbor-Joining tree in MEGA4 (Tamura et al. 2007) with 1,000 bootstraps, Kimura 2-parameter model, and uniform rates among sites. To build a similar phylogenetic tree for coding sequences, protein sequences were first predicted from ~3,000 full-length elephant shark cDNA sequences using OrfPredictor (Min et al. 2005). These proteins were searched against the proteomes of human and other vertebrates by reciprocal BlastP ($E < 10^{-20}$) and 271 proteins (unpublished; sequences available at http://blast.fugu-sg.org/, see Downloads) with orthologs in all six species were identified. Multiple alignments of the proteins from the six species were generated using ClustalW and converted to ungapped codon-based alignments by PAL2NAL (Suyama et al. 2006). Finally, the alignments were concatenated into a single alignment, and a Neighbor-Joining tree was constructed using similar settings as for UCEs.

Supplementary Material

Supplementary figure 1 is available at *Molecular Biology and Evolution* (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We are thankful to Stuart Stephen for providing human genome coordinates for the UCEs. This work was supported by the Biomedical Research Council of the A*STAR, Singapore.

Literature Cited

Blanchette M, Kent WJ, Riemer C, et al. (12 co-authors). 2004. Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res. 14:708–715.

Brunet FG, Crollius HR, Paris M, Aury JM, 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.

Chiu CH, Amemiya C, Dewar K, Kim CB, Ruddle FH, Wagner GP. 2002. Molecular evolution of the HoxA cluster in the three major gnathostome lineages. Proc Natl Acad Sci USA. 99:5492–5497.

Christoffels A, Koh EG, Chia JM, 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.

Hedges SB, Kumar S. 2003. Genomic clocks and evolutionary timescales. Trends Genet. 19:200–206.

Jaillon O, Aury JM, Brunet F, et al. (61 co-authors). 2004. Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. Nature. 431:946–957.

Karolchik D, Kuhn RM, Baertsch R, et al. (25 co-authors). 2008. The UCSC genome browser database: 2008 update. Nucleic Acids Res. 36:D773–D779.

Kumazawa Y, Yamaguchi M, Nishida M. 2000. Mitochondrial molecular clocks and the origin of euteleostean biodiversity: Familial radiation of perciforms may have predated the Cretaceous/tertiary boundary. In: Kato M, editor. The biology of biodiversity. New York: Springer-Verlag.p. 35–52

Martin AP. 1999. Substitution rates of organelle and nuclear genes in sharks: implicating metabolic rate (again). Mol Biol Evol. 16:996–1002.

Martin AP, Naylor GJ, Palumbi SR. 1992. Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. Nature. 357:153–155.

Min XJ, Butler G, Storms R, Tsang A. 2005. OrfPredictor: predicting protein-coding regions in EST-derived sequences. Nucleic Acids Res. 33:W677–W680.

Pennacchio LA, Ahituv N, Moses AM, et al. (19 co-authors). 2006. In vivo enhancer analysis of human conserved non-coding sequences. Nature. 444:499–502.

Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, Haussler D, Miller W. 2003. Human–mouse alignments with BLASTZ. Genome Res. 13:103–107.

Shin JT, Priest JR, Ovcharenko I, Ronco A, Moore RK, Burns CG, MacRae CA. 2005. Human–zebrafish non-coding conserved elements act in vivo to regulate transcription. Nucleic Acids Res. 33:5437–5445.

Steinke D, Salzburger W, Braasch I, Meyer A. 2006. Many genes in fish have species-specific asymmetric rates of molecular evolution. BMC Genomics. 7:20.

Stephen S, Pheasant M, Makunin IV, Mattick JS. 2008. Large-scale appearance of ultraconserved elements in tetrapod genomes and slowdown of the molecular clock. Mol Biol Evol. 25:402–408.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34:W609–612.

Taft RJ, Pheasant M, Mattick JS. 2007. The relationship between non-protein-coding DNA and eukaryotic complexity. Bioessays. 29:288–299.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol. 24:1596–1599.

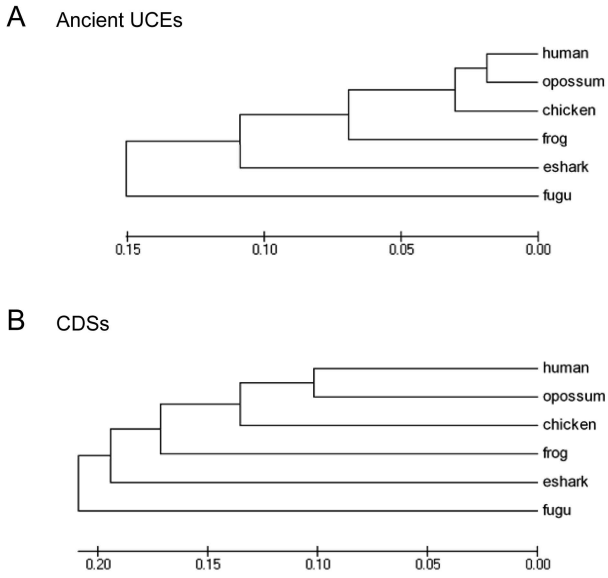
Venkatesh B, Kirkness EF, Loh YH, et al. (12 co-authors). 2006. Ancient noncoding elements conserved in the human genome. Science. 314:1892.

Venkatesh B, Kirkness EF, Loh YH, et al. (12 co-authors). 2007. Survey sequencing and comparative analysis of the elephant shark (*Callorhinchus milii*) genome. PLoS Biol. 5:e101.

Woolfe A, Goodson M, Goode DK, et al. (16 co-authors). 2005. Highly conserved non-coding sequences are associated with vertebrate development. PLoS Biol. 3:e7.

Naruya Saitou, Associate Editor

Accepted November 23, 2008



Supplementary Fig. 1: Substitution rates in UCEs and CDSs. The phylogenetic trees were generated using Neighbor-Joining method and without specifying an outgroup. The distance in substitutions per site is shown below each tree. The standard errors for all branch lengths are less than 10^{-3}. eshark, elephant shark. (A) Linearized phylogenetic tree built from 3,297 ancient UCEs present in the elephant shark and other vertebrates (443 kb ungapped alignment). (B) Linearized phylogenetic tree built from 271 elephant shark coding sequences ranging in length from 228 bp to 2.1 kb and their orthologs in other species (198 kb ungapped alignment). Note that because of the high substitution rates in UCEs in fugu, the phylogenetic tree for UCEs (A) is inconsistent with the species tree of these vertebrates; fugu is incorrectly placed as the most basal lineage. In the CDSs tree, largely because of the lower level of substitutions in the CDSs in the elephant shark, the elephant shark is incorrectly placed as a sister group of tetrapods (B).