Basal Vertebrates Clarify the Evolutionary History of Ciliopathy-Associated Genes *Tmem138* and *Tmem216*

Byrappa Venkatesh,*[1] Vydianathan Ravi,[1] Alison P. Lee,[1] Wesley C. Warren,[2] and Sydney Brenner[1]
[1]Comparative Genomics Lab, Institute of Molecular and Cell Biology, A*STAR, Biopolis, Singapore
[2]The Genome Institute, Washington University School of Medicine
***Corresponding author:** E-mail: mcbbv@imcb.a-star.edu.sg.
Associate editor: Billie Swalla

Abstract

Recently, Lee et al. (Lee JH, Silhavy JL, Lee JE, et al. (30 co-authors). 2012. Evolutionarily assembled cis-regulatory module at a human ciliopathy locus. *Science* (335:966–969.) demonstrated that mutation in either of the transmembrane protein encoding genes, *TMEM138* or *TMEM216*, causes phenotypically indistinguishable ciliopathy. Furthermore, on the basis of the observation that their orthologs are linked in a head-to-tail configuration in other mammals and *Anolis*, but present on different scaffolds or chromosomes in *Xenopus tropicalis* and zebrafish, the authors concluded that the two genes were joined by chromosomal rearrangement at the evolutionary amphibian-to-reptile transition to form a functional module. We have sequenced these gene loci in a cartilaginous fish, the elephant shark, and found that the two genes together with a related gene (*Tmem80*) constitute a tandem cluster. This suggests that the two genes were already linked in the vertebrate ancestor and then rearranged independently in *Xenopus* and zebrafish. Analyses of the coelacanth and lamprey genomes support this hypothesis. Our study highlights the importance of basal vertebrates as critical reference genomes.

Key words: *Callorhinchus milii*, coelacanth, teleost fishes, comparative genomics.

Letter

The transmembrane (TM) proteins are an ancient superfamily of proteins conserved from nematodes to mammals. Two of the human genes encoding TMEM proteins, *TMEM138* and *TMEM216*, are linked in a head-to-tail configuration with an intergenic distance of ~23 kb. In a recent study, Lee et al. (2012) demonstrated that mutation in either of these genes causes phenotypically indistinguishable ciliopathy or Joubert syndrome. Furthermore, they showed that the two genes are tightly coexpressed through a shared *cis*-regulatory element in their intergenic region, and their proteins play an interdependent role in ciliary assembly, indicating that the two genes constitute a functional gene cluster. Because the two protein sequences showed very little sequence similarity and did not share any domain other than the TM domains, the authors concluded that the two genes are unlikely to be the result of gene duplication. In addition, the authors compared the genomic organization of these genes in various bony vertebrates and found that although the two genes are linked in mammals and *Anolis*, they are present on different scaffolds and chromosomes in *Xenopus tropicalis* and zebrafish genome assemblies, respectively. These observations led to the core conclusion of the study that the two genes were joined into a functional unit through chromosomal rearrangement at the evolutionary amphibian-to-reptile transition (Lee et al. 2012). However, the authors did not consider the alternative possibility that the genes were linked in the vertebrate ancestor, and the linkage was disrupted in *X. tropicalis* and zebrafish.

To investigate this possibility, we sequenced *Tmem138* and *Tmem216* gene loci in the elephant shark (*Callorhinchus milii*), a cartilaginous fish (Chondrichthyes), representing the most basal phylogenetic group of jawed vertebrates (gnathostomes). Cartilaginous fishes diverged from bony vertebrates ~450 Ma (Inoue et al. 2010). We found an elephant shark bacterial artificial chromosome (BAC) clone (05M17) that contained fragments of both *Tmem138* and *Tmem216* and sequenced it completely (supplementary Material and Methods, Supplementary Material online). Interestingly, the elephant shark *Tmem138* and *Tmem216* genes are closely linked (2.3 kb intergenic region) in the same head-to-tail configuration as in mammals and *Anolis* (fig. 1). In addition, a third *Tmem* gene, *Tmem80*, is present 0.8 kb downstream of *Tmem216* (fig. 1). The human ortholog of this gene is located ~60 Mb upstream of *TMEM138*. A neighbor-joining tree confirmed that the elephant shark *Tmem* genes are the orthologs of human *TMEM* genes (supplementary fig. S1*a*, Supplementary Material online). The tandem array of the three *Tmem* genes in the elephant shark suggests that these genes were linked in the common ancestor of gnathostomes.

The elephant shark Tmem138 and Tmem216 proteins show 66% and 56% identity to human TMEM138 and TMEM216, respectively. The human TMEM138 protein is predicted to contain a signal peptide and three TM domains, whereas TMEM216 lacks a signal peptide but contains four TM domains. Interestingly, both elephant shark proteins are composed of a signal peptide and three TM domains (predicted by both SMART and InterProScan) (supplementary fig. S1*b* and *c*, Supplementary Material online). The similar domains shared by the elephant shark Tmem138 and Tmem216 proteins suggest that the two proteins are related. Furthermore, the tandem linkage of their genes suggests they are likely to be the result of gene duplication. It appears that

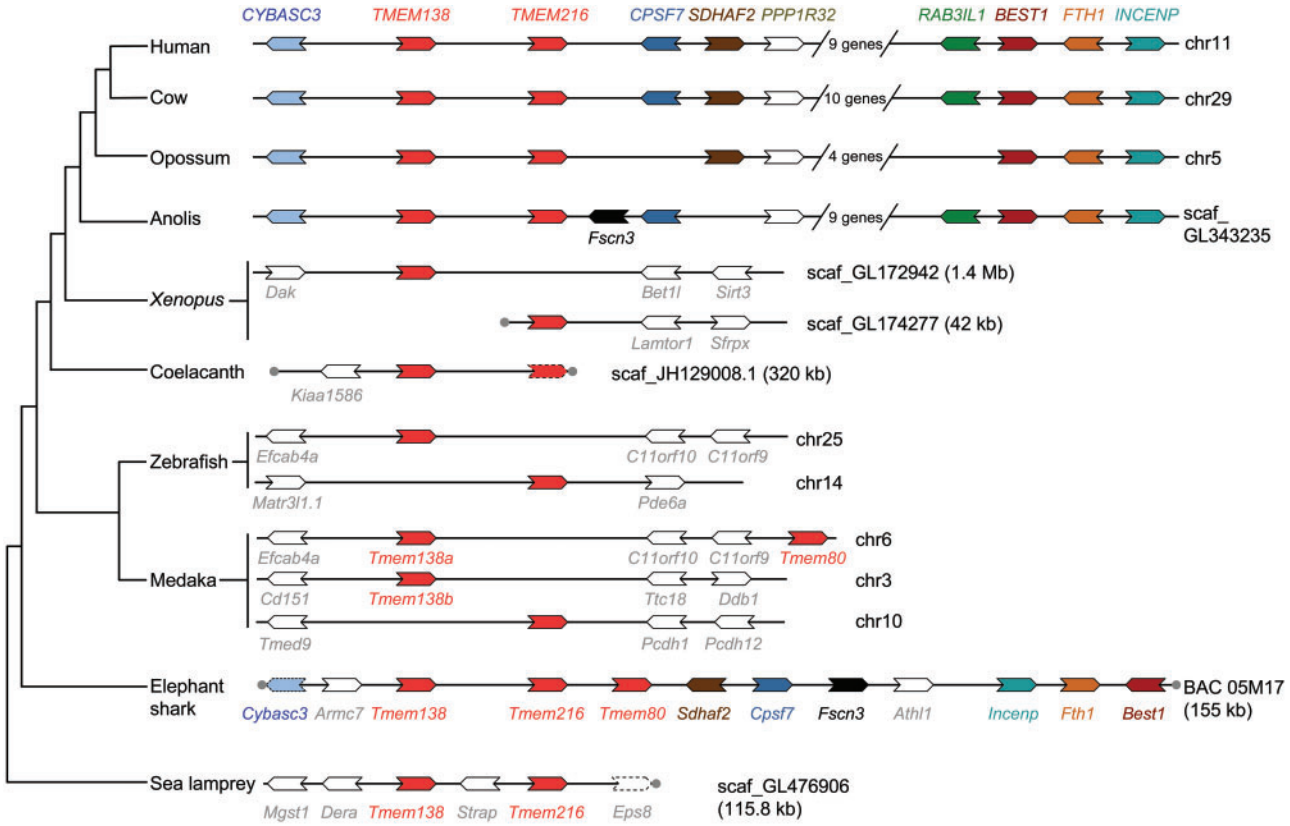

Fig. 1. *Tmem138* and *Tmem216* gene loci in various vertebrates. Genes are shown as block arrows. Genes with conserved synteny in elephant shark and other vertebrates are colored. Incomplete genes are shown as dotted arrows. Scaffold ends are marked by gray dots. The gene order for bony vertebrates is from Ensembl (http://asia.ensembl.org, last accessed April 20, 2012) or the University of California, Santa Cruz (UCSC) Genome Browser (http://genome.ucsc.edu, last accessed April 20, 2012).

after the divergence of cartilaginous fishes and bony vertebrates, the coding sequence of *Tmem138* has remained relatively conserved among various vertebrates, whereas that of *Tmem216* has diverged considerably in the bony vertebrate lineage.

Lee et al. (2012) have noted that human proteins TMEM216 and TMEM80 (and their related protein TMEM17) contain Transmemb_17 domain, whereas TMEM138 lacks this domain. The elephant shark Tmem138 also lacks Transmemb_17 domain, whereas Tmem216 and Tmem80 contain this domain. The Transmemb_17 domain is an approximately 100 amino acid region of a family of proteins (Transmemb_17 superfamily) conserved from nematodes to humans (Marchler-Bauer et al. 2011). It is possible that the ancestral TMEM protein contained Transmemb_17 domain and following the gene duplication that gave rise to *Tmem138* and *Tmem17/216/80*, the Transmemb_17 domain of *Tmem138* accumulated substitutions resulting in its divergence beyond recognition.

Recently, draft genome sequences of a jawless vertebrate, the sea lamprey (*Petromyzon marinus*) (http://www.ensembl.org/Petromyzon_marinus/Info/Index, last accessed August 18, 2012), and a lobe-finned fish, the coelacanth (*Latimeria chalumnae*) (http://www.broadinstitute.org, last accessed April 20, 2012) have been generated. Jawless vertebrates reside on the most basal branch of living vertebrates, whereas the coelacanths diverged early from the tetrapod lineage after the divergence of ray-finned fishes. A search of the lamprey and coelacanth genome assemblies (http://www.ensembl.org/, last accessed April 20, 2012) showed that *Tmem138* and *Tmem216* are linked in a head-to-tail configuration in the coelacanth (intergenic distance ~68 kb), whereas they are separated by an unrelated gene (*Strap*) in the lamprey (fig. 1). Such an arrangement of *Tmem138* and *Tmem216* genes in coelacanth and lamprey provides strong support that they were linked in the common ancestor of vertebrates, and their linkage has been disrupted independently in *X. tropicalis* and zebrafish.

To verify whether the disruption of the linkage of *Tmem138* and *Tmem216* genes is unique to zebrafish or shared by other teleosts, we analyzed the genomic organization of these genes in medaka, stickleback, fugu, and *Tetraodon*. Our analysis showed that although these genes are single copy in zebrafish, other teleosts such as medaka, stickleback, fugu, and *Tetraodon* contain two unlinked copies of *Tmem138* (*Tmem138a* and *Tmem138b*) besides a single *Tmem216*. Although some syntenic genes in the human *TMEM138–TMEM216* locus are conserved in the *Tmem138a* locus, the synteny is thoroughly disrupted in the *Tmem138b* locus (fig. 2). The disruption of synteny in the duplicated *Tmem138* loci could be explained by the secondary loss of duplicated genes and/or chromosomal rearrangements. The presence of the single copy *Tmem216* on a chromosome different from those of *Tmem138a* and

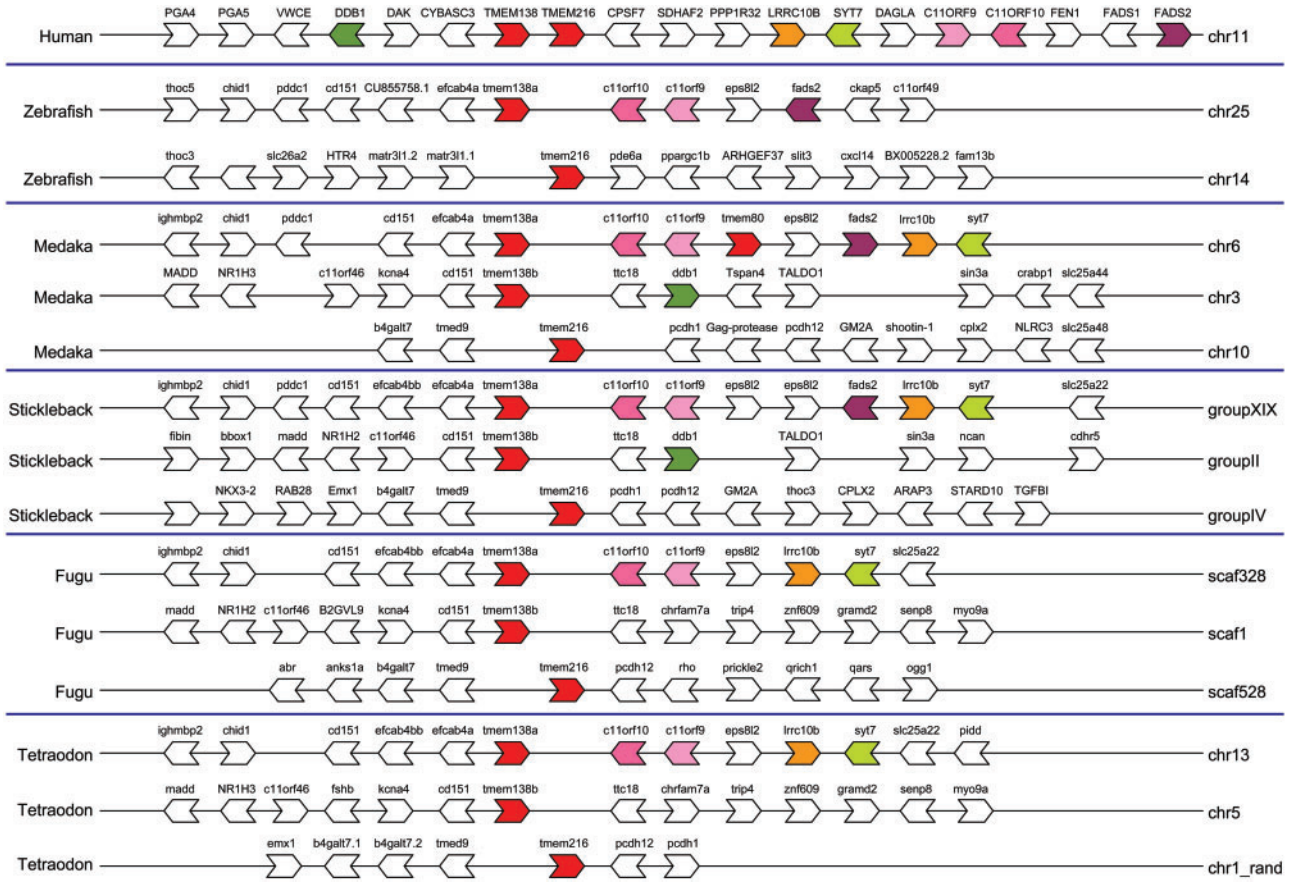

Fig. 2. Genome organization of *Tmem138* and *Tmem216* gene loci in human and various teleost fishes. Genes are represented by block arrows. Genes with conserved synteny in human and teleost fishes are colored. The gene order is from Ensembl (http://asia.ensembl.org, last accessed April 20, 2012) or the University of California, Santa Cruz (UCSC) Genome Browser (http://genome.ucsc.edu, last accessed April 20, 2012).

Tmem138b suggests that this gene was translocated to a new locus before the divergence of the five teleosts. Thus, the teleost *Tmem138–Tmem216* locus has experienced multiple genomic changes. Our analysis provides further support to the previous observations that teleost genomes have experienced a higher rate of chromosomal rearrangements compared with mammals (Jaillon et al. 2004; Nakatani et al. 2007), elephant shark (Venkatesh et al. 2007), and basal ray-finned fishes such as the spotted gar (Amores et al. 2011). Thus, one needs to exercise caution when using teleost genomes as reference in comparative studies. Unlike teleosts, *X. tropicalis* has not experienced an additional whole-genome duplication, and thus, the delinking of *Tmem138* and *Tmem216* genes in *X. tropicalis* cannot be explained by rearrangements due to genome duplication.

Of the 12 genes present in the elephant shark BAC no. 05M17, orthologs for two or more of 10 genes (*Cybasc3*, *Tmem138*, *Tmem216*, *Tmem80*, *Sdhaf2*, *Cpsf7*, *Fscn3*, *Incenp*, *Fth1*, and *Best1*) are present in the *Tmem138–Tmem216* locus of at least one of the bony vertebrates (fig. 1). This implies that these 10 genes were present as a syntenic block in the ancestral gnathostome *Tmem138–Tmem216* locus. Thus, the elephant shark locus appears to have experienced very little rearrangement after it diverged from the gnathostome ancestor. This is noteworthy considering that the elephant shark lineage has also been evolving independently since its divergence from the bony vertebrate ancestor ~450 Ma. Thus, our study underscores the importance of the elephant shark as a valuable basal gnathostome reference genome.

Supplementary Material

Supplementary Material and Methods and figure S1 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Boon-Hui Tay for technical assistance in sequencing the BAC clone, and the Lamprey Genome Consortium and the Broad Institute for sequencing the genomes of sea lamprey and coelacanth, respectively, and making the data available in the public domain. This work was supported by the Biomedical Research Council of A*STAR, Singapore.

References

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* 188:799–808.

Inoue JG, Miya M, Lam K, Tay BH, Danks JA, Bell J, Walker TI, Venkatesh B. 2010. Evolutionary origin and phylogeny of the modern holocephalans (Chondrichthyes: Chimaeriformes): a mitogenomic perspective. *Mol Biol Evol.* 27:2576–2586.

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.

Lee JH, Silhavy JL, Lee JE, et al. (30 co-authors). 2012. Evolutionarily assembled cis-regulatory module at a human ciliopathy locus. *Science* 335:966–969.

Marchler-Bauer A, Lu S, Anderson JB, et al. (27 co-authors). 2011. CDD: a Conserved Domain Database for the functional annotation of proteins. *Nucleic Acids Res.* 39:D225–D229.

Nakatani Y, Takeda H, Kohara Y, Morishita S. 2007. Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. *Genome Res.* 17: 1254–1265.

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.