Available online at www.sciencedirect.com
SCIENCE DIRECT
GENOMICS
Genomics 88 (2006) 1 – 11
www.elsevier.com/locate/ygeno

Evolution of the vertebrate ABC gene family: Analysis of gene birth and death ☆

Tarmo Annilo [a], Zhang-Qun Chen [b], Sergey Shulenin [a], Julie Costantino [a], Lauren Thomas [a], Hong Lou [a], Stefan Stefanov [a], Michael Dean [a,*]

[a] *Laboratory of Genomic Diversity, Building 560, Room 21-18, NCI–Frederick, Frederick, MD 21702, USA*
[b] *Basic Research Program, Scientific Applications International Corporation, NCI–Frederick, Frederick, MD 21702, USA*

Received 13 October 2005; accepted 2 March 2006
Available online 2 May 2006

Abstract

Vertebrate evolution has been largely driven by the duplication of genes that allow for the acquisition of new functions. The ATP-binding cassette (ABC) proteins constitute a large and functionally diverse family of membrane transporters. The members of this multigene family are found in all cellular organisms, most often engaged in the translocation of a wide variety of substrates across lipid membranes. Because of the diverse function of these genes, their large size, and the large number of orthologs, ABC genes represent an excellent tool to study gene family evolution. We have identified ABC proteins from the sea squirt (*Ciona intestinalis*), zebrafish (*Danio rerio*), and chicken (*Gallus gallus*) and, using phylogenetic analysis, identified those genes with a one-to-one orthologous relationship to human ABC proteins. All ABC protein subfamilies found in *Ciona* and zebrafish correspond to the human subfamilies, with the exception of a single ABCH subfamily gene found only in zebrafish. Multiple gene duplication and deletion events were identified in different lineages, indicating an ongoing process of gene evolution. As many ABC genes are involved in human genetic diseases, and important drug transport phenotypes, the understanding of ABC gene evolution is important to the development of animal models and functional studies.
Published by Elsevier Inc.

Keywords: ATP-binding cassette; Transporter; Evolution; Lipid transport; Drug resistance; Cystic fibrosis; Adrenoleukodystrophy

Gene duplication is recognized as the major force driving evolution [1]. The ATP-binding cassette (ABC)[1] transporters constitute one of the largest protein multigene families, providing an attractive model for the study of gene family evolution. The members of this functionally diverse family are found in all cellular organisms and have been extensively studied in humans. They use the energy from ATP hydrolysis to translocate a wide variety of substrates across extra- and intracellular membranes. ABC genes are responsible for a large fraction of the multidrug resistance of cancer cells and pathogens [2], and mutations in a number of ABC genes are the cause of 18 distinct inherited diseases, such as cystic fibrosis, Stargardt disease, and some disorders of cholesterol metabolism [3,4].

A functional ABC transporter is composed of two nucleotide-binding folds (NBFs) and two transmembrane domains (TMDs). The highly conserved NBF contains three motifs: Walker A and Walker B domains, found in all ATP-binding proteins, and a signature motif (LSGG), a distinctive feature of ABC proteins [2]. The more diverged TMDs recognize and translocate the substrate across the membrane. In humans, the 48 ABC proteins are divided into seven subfamilies, from A to G, based on structural arrangements and phylogenetic analysis [5,6]. These include full transporters that encode a complete protein and half-transporters that form hetero- or homodimers. Although several ABC genes, such as ABCB1/MDR1 or ABCC7/CFTR, have been extensively studied, the functions of many others are still unknown. Through evolutionary analysis, we can identify orthologous genes from other organisms. Inferring that orthologous genes have similar or related

Abbreviations: ABC, ATP-binding cassette; NBF, nucleotide binding fold; TMD, transmembrane domain.

☆ Sequence data from this article have been deposited with the GenBank Data Library under Accession Nos. DQ223858–DQ223861.

* Corresponding author. Fax: +1 301 846 1909.
E-mail address: dean@ncifcrf.gov (M. Dean).

0888-7543/$ - see front matter. Published by Elsevier Inc.
doi:10.1016/j.ygeno.2006.03.001

functions in different organisms helps to generate hypotheses about the possible function of human genes.

The availability of zebrafish genomic sequence will greatly enhance its role as a powerful model organism to study vertebrate development as well as human disease [7,8]. Relatively simple and efficient methods for the analysis of gene function using morpholino antisense oligonucleotides [9] have made zebrafish an attractive model organism. There have been no reports describing ABC transporters from *Ciona* or zebrafish except for the identification of two ABCB subfamily members from zebrafish (TAP2A and TAP2B) [10].

Comparison of the genomes of vertebrates and the sea squirt *Ciona intestinalis*, a nonvertebrate chordate, allows for the study of vertebrate genome evolution and the identification of genes involved in vertebrate developmental innovations [11]. On the other hand, the radiation of the teleost fishes has been shown to involve a whole-genome duplication followed by silencing or loss of copies of many duplicated genes [12]. The draft genome sequence of the 159-Mb *Ciona* genome [11] revealed approximately 16,000 genes, compared to the estimated 25,000 genes in mammals [13] and 38,000 genes in *Fugu* [14].

Recent reviews have provided overviews of ABC proteins in bacteria [15] and archae [16]. The evolution of paralogous genes, derived from gene duplication, can been studied by comparing the genomes of related organisms. In eukaryotes, paralogous genes often are located in segmental duplications or gene clusters. Here we provide the evolutionary analysis of chordate ABC proteins, comparing the proteins from *Ciona* (representative of ancestral chordate), zebrafish (a species with an ancient whole-genome duplication), chicken (a bird), and dog and human (mammals). In addition, we have determined the cDNA sequence of several ABC genes from zebrafish and analyzed the ABC subfamily in the light of whole-genome duplication [12] in teleost lineage. We have determined a number of one-to-one orthologous relationships between mammalian and fish genes, indicating significant gene loss after the ancient whole-genome duplication. These studies provide insight into the evolution of ABC genes and give further knowledge of the evolution of gene families.

Results

Predicted gene models based on ENSEMBL genome assemblies were analyzed using TblastN to characterize the ABC genes from zebrafish, chicken, and *Ciona.* Alignments were performed to the human homologs and since the analysis was performed against the preassembly state of the chicken and zebrafish genomes, manual correction of the sequence was performed in many cases in which the fragments of a single gene were located on different contigs or were rearranged. For many zebrafish genes primers were designed and cDNA was cloned and sequenced to either confirm the sequence or fill in gaps. The resulting predicted ABC genes were assigned to subfamilies based on BLAST similarity score and structural organization and through phylogenetic analysis. The NBFs of ABC proteins (which are about 200 aa long) are highly conserved and allow for phylogenetic analysis between the proteins from distantly related organisms or from different subfamilies. To achieve better resolution of proteins within a subfamily the full protein sequences were analyzed for the individual subfamilies (except for E and F as well as G and H, which are closely related and were analyzed together). In total, 33 ABC genes from *Ciona,* 52 from zebrafish, 41 from chicken, and 51 from dog were identified (Table 1 and Supplementary Tables). In addition, a few gene fragments that could not be further evaluated were identified. Only sequences that cover at least 60% of the related human protein were included in phylogenetic analysis.

From this analysis of the vertebrate ABC genes a large number of duplication events are apparent as well as several gene deaths. Fig. 1A displays each of the ancestral vertebrate ABC gene and the gene duplication events (green, i.e., *Abca14–16* in rodents and dog), gene deaths (red, i.e., *ABCA14–16* in human), and duplications specific to fish, rodents, or primates (in blue, i.e., *Abca1a* in fish). In total, nine ABC genes (Abca10, −14,−15,−16, and −17; Abcb1,−4, and −5; and Abcc11) are the result of duplications that have persisted in most mammalian genomes that have been sequenced to date, and there are four examples of gene loss. In addition there are 17 restricted gene duplications, 13 of which occurred in exclusively in fish, 3 in rodents, and 1 in primates. These data are supported by analysis of EST sequences confirming that Abcg3 is not found outside rodents (data not shown).

To analyze further the gene duplications, the locations of the duplicated genes were examined and adjacent duplications identified (Fig. 1B). A total of seven ABC genes have participated in adjacent duplications. The most complex locus is that of the *ABCA5*-related genes, which has undergone an expansion to five genes in mammals. For example zebrafish contains a single *Abca5*-like gene, whereas chicken has three and all mammals have four (cow) or five genes in this cluster (Supplementary Figs. 7–12). The *ABCB1* gene has undergone one duplication to create *ABCB4* and a separate duplication to generate the *Abcb1b* gene in rodents and opossum (Supplementary Figs. 1 and 13–15). Table 2 displays the number of duplicated and unduplicated genes by subfamily. The genes expressed in most cells (ABCE and ABCF, ABCB half-transporters (HT)) and the ABCC genes have undergone few duplications, whereas the ABCA, ABCB full-transporters (FT) and ABCG genes have undergone frequent duplication. These analyses show an active process that includes adjacent duplications, presumably by unequal crossovers; nonadjacent duplications believed to occur by chromosome or genome duplication; and the loss of genes through deletions and sequence degeneration. There are undoubtedly more gene death events that cannot be observed because the sequences have undergone extensive mutation.

The ABCA subfamily

In humans, the ABCA subfamily contains 12 members (Fig. 2A), all of them full-transporters, including the unusually large ABCA12 and ABCA13 proteins (2595 and 5058 amino acids,