# Letter

# Partitioning of Tissue Expression Accompanies Multiple Duplications of the Na+/K+ ATPase $\alpha$ Subunit Gene

# Fabrizio C. Serluca,<sup>1</sup> Arend Sidow,<sup>2</sup> John D. Mably,<sup>1</sup> and Mark C. Fishman<sup>1,3</sup>

<sup>1</sup>Cardiovascular Research Center and Developmental Biology Laboratory, Massachusetts General Hospital, and Department of Medicine, Harvard Medical School, Boston, Massachusetts 02119, USA; <sup>2</sup>Departments of Pathology and Genetics, Stanford University, Stanford, California 94305, USA

Vertebrate genomes contain multiple copies of related genes that arose through gene duplication. In the past it has been proposed that these duplicated genes were retained because of acquisition of novel beneficial functions. A more recent model, the duplication-degeneration-complementation hypothesis (DDC), posits that the functions of a single gene may become separately allocated among the duplicated genes, rendering both duplicates essential. Thus far, empirical evidence for this model has been limited to the *engrailed* and *sox* family of developmental regulators, and it has been unclear whether it may also apply to ubiquitously expressed genes with essential functions for cell survival. Here we describe the cloning of three zebrafish  $\alpha$  subunits of the Na(+),K(+)-ATPase and a comprehensive evolutionary analysis of this gene family. The predicted amino acid sequences are extremely well conserved among vertebrates. The evolutionary relationships and the map positions of these genes and of other  $\alpha$ -like sequences indicate that both tandem and ploidy duplications contributed to the expansion of this gene family in the teleost lineage. The duplications are accompanied by acquisition of clear functional specialization, consistent with the DDC model of genome evolution.

[The sequence data described in this paper have been submitted to the GenBank data library under accession nos. AY028628, AY028629, and AY028630]

Vertebrate genomes have been shaped by several episodes of large-scale gene duplications. During a very short time in the evolution of early vertebrates, for example, many gene families expanded from one to several paralogs (Sidow 1992; Suga et al. 1999). Such large-scale duplications are the consequence both of individual gene duplications and ploidy duplications, the latter known to have occurred recently in Xenopus and salmonid fishes. There is strong evidence for an older ploidy duplication during the evolution of bony fishes (Postlethwait et al. 1998). For example, seven Hox clusters are found in zebrafish resulting from a duplication of the four Hox clusters found in most other higher vertebrates (Amores et al., 1998; Prince et al. 1998). As would be predicted from a ploidy duplication, genes linked to the seven Hox clusters are also duplicated with syntenic conservation to the corresponding cluster in mammals (Amores et al. 1998; Woods et al. 2000). Other syntenic regions between human and zebrafish chromosomes lend further support to this model (Postlethwait et al. 2000).

In classical models of genome evolution, the fate of a duplicated gene is dependent upon the nature of the mutations it accumulates. If a new beneficial function is acquired, the duplicate will be retained; if no new function is acquired, it will be lost (Ohno 1970; Sidow 1996). Recently, Force et al.

<sup>3</sup>Corresponding author.

**E-MAIL fishman@cvrc.mgh.harvard.edu; FAX (617) 726-5806.** Article published on-line before print: *Genome Res.*, 10.1101/gr.192001. Article and publication are at http://www.genome.org/cgi/doi/10.1101/ gr.192001. (1999) have proposed an alternate model for certain duplications, one that takes into account the existence of multiple expression domains typically found for products of single genes. Accordingly, the duplication-degenerationcomplementation (DDC) model, partial loss-of-function mutations accumulate in the cis regulatory regions of both paralogs, such that each copy of the gene is expressed in a particular spatio-temporal manner, and both copies of the gene must be maintained to preserve the overall function of the original gene. Thus, paradoxically, the accumulation of degenerative mutations enhances the chances of survival of both paralogs. The first example described for the DDC model involves the two engrailed1 genes in zebrafish in which two sites, the pectoral fin buds and spinal cord neurons, express only one of the paralogs (Force et al. 1999). Members of the sox gene family of developmental regulators show a similar partitioning of expression (de Martino et al. 2000). The engrailed and sox duplicate pairs arose from a chromosomal duplication. However, in principle, the nature of the duplication event giving rise to the two paralogs should be irrelevant.

The Na(+),K(+)-ATPase or sodium pump is responsible for maintaining proper intracellular and extracellular concentrations of sodium and potassium ions, and is essential to membrane potential generation (Glynn 1993; Lingrel and Kuntzweiler 1994). The protein consists of a large multi-pass  $\alpha$  subunit and an associated glycosylated single-pass  $\beta$  subunit, and has been studied extensively at the biochemical level. The highly conserved  $\alpha$  family of genes that encode the catalytic subunits of the pump contain the binding sites for Na(+), K(+), ATP and the digitalis glycosides (Lingrel and Kuntzweiler

# Serluca et al.

1994). The mature protein can be associated with a third, smaller  $\gamma$  subunit (Mercer et al. 1993). Invertebrate genomes possess one  $\alpha$  subunit of the Na(+),K(+)-ATPase (Emery et al. 1995). Four  $\alpha$  subunits have been reported in mammalian systems (Kawakami et al. 1986; Shull et al. 1986; Martin-Vasallo et al. 1989; Shamraj and Lingrel 1994; Malik et al. 1996). In mammals, the expression patterns of the different genes show strong tissue preferences (Sweadner 1989; Blanco and Mercer 1998). The  $\alpha 1$  gene (*atp1a1*) is preferentially expressed in the kidney, gut, and heart, and ubiquitously at lower levels. The  $\alpha 2$  (atp1a2) gene is expressed in muscle, adipocytes, heart, and brain. The  $\alpha 3$  (atp1a3) gene is expressed throughout the nervous system. A less conserved fourth isoform,  $\alpha 4$ , is expressed in the testes (Shamraj and Lingrel 1994). The  $\alpha 1$  isoform is essential to cell survival. Homozygous disruption of this gene in mice causes embryonic lethality (James et al. 1999).

We evaluated these genes in the zebrafish because of evidence that it has undergone an additional round of genome duplication over 100 million years ago (Amores et al., 1998). Through cloning and database searches, we discovered that zebrafish have at least eight  $\alpha$  subunits. Five are in the  $\alpha 1$ class, two in the  $\alpha 3$  class, and one in  $\alpha 2$  class. Sequence, mapping, and expression analyses all suggest that even in the relatively short period of time since gene duplication, they have acquired specialized functions, supporting the subfunctionalization model of genome evolution.

# RESULTS

#### Homology to Other $\alpha$ Subunits

We cloned three  $\alpha$  subunits of the Na(+),K(+)-ATPase. Two of the genes belong to the  $\alpha 1$  family and the third is in the  $\alpha 2$ group. From the predicted amino acid sequences, the two  $\alpha 1$ genes share 91% identity between each other and 83% identity with the  $\alpha 2$  homolog (Fig. 1). This amino acid identity is preserved when comparing the zebrafish subunits with their human counterparts. *atp* $\alpha 1A1$  and *atp* $\alpha 1B1$  show 88% and 89% identity, respectively, to the human  $\alpha 1$  subunit and *atp* $\alpha 2$  shares 86% identity to the human  $\alpha 2$  protein.

A number of motifs essential for the sodium pump's function are noted in Figure 1. The conserved cytoplasmic TGES/A motif, critical to the catalytic cycle of all P-type ATPases, is present in all three of the zebrafish  $\alpha$  subunits. The catalytically phosphorylated aspartyl residue is also conserved and found within the consensus DKTGTLTQ sequence. Protein kinases can modulate the catalytic activity of the sodium pump in a complex fashion (Therien and Blostein 2000). Phosphorylation by PKA is thought to occur at Ser 943 of the human and rat sequences. This serine is also found in all three of the zebrafish isoforms within the context of a PKA consensus site (XRRXSX). The Na(+),K(+)-ATPase can also bind to ankyrin repeats (Zhang et al. 1998), and this binding site is conserved in all three fish isoforms. The presence of a leucine residue in the  $\alpha 1$  genes instead of a methionine is unusual, although it is found in the related P-type pump, the gastric H(+),K(+)-ATPase.

Ouabain resistance in the rat  $\alpha 1$  isoform has been mapped to the extracellular loop between the first and second membrane-spanning segments. Mutation of two amino acid residues within this loop to those of rat  $\alpha 1$  can confer resistance to ouabain-sensitive isoforms (Price et al. 1990; Jewell and Lingrel 1992). The zebrafish genes have sequences predicted to be ouabain sensitive in that they contain a glutamine or leucine rather than an arginine residue found at the first position of the resistant isoform, and an asparigine instead of an aspartate at the second.

#### Phylogenetic Analysis

To generate an evolutionary comparison, we identified additional vertebrate  $\alpha$  genes in GenBank. Between gene cloning and database analysis, we have identified a total of eight  $\alpha$ genes. Five are  $\alpha 1$  paralogs, two are  $\alpha 3$  paralogs, and one is an  $\alpha 2$  gene. (The  $\alpha 4$  gene in mammals is far less conserved and sequence information from multiple species is lacking, so we did not include it in the analysis.) Protein maximum likelihood analysis predicts the evolutionary tree shown in Figure 2.

Invertebrate genomes contain only one  $\alpha$  subunit (Emery et al. 1995), which we used to root the tree. The first gene duplication created the  $\alpha$ 3 branch and an ancestral  $\alpha$ 1 $\alpha$ 2 gene, which then duplicated to give rise to the  $\alpha$ 1 and  $\alpha$ 2 paralogs. Nodes that correspond to these duplications are shown in black in Figure 2, and are likely due to the large-scale gene duplications that occurred in the ancestral lineage of vertebrates. All additional gene duplications appear only in the teleosts. The duplication giving rise to the two  $\alpha$ 3 paralogs (*atp* $\alpha$ 3*A* and *atp* $\alpha$ 3*B*) appears to have preceded the divergence of the *Tilapia* and zebrafish lineages (ProtML *BootP* value = 83%) and is likely to have occurred as part of the known large-scale gene duplication in the teleost lineage.

The  $\alpha 1$  gene appears to have undergone additional rounds of duplication. The first  $\alpha 1$  gene duplication event appears to have occurred in an early bony fish ancestor prior to the divergence of the lineage leading to present day eels (Fig. 2, blue node). Placement of the eel lineage before the gene duplication is rejected at a ProtML BootP value of 0.001%, confirming that this duplication was not due to the known large-scale teleost duplication. The two resulting  $\alpha 1$ homologs then further duplicated, generating at least five  $\alpha 1$ genes in zebrafish. Duplication of the  $atp\alpha 1A$  group (gray node) occurred after divergence from the Tilapia lineage (supported at BootP = 100%). A subsequent duplication (vellow node) generated the other two known  $atp\alpha 1A$  homologs  $(atp\alpha 1A2a \text{ and } atp\alpha 1A2b)$ . The one gene duplication in the subtree containing  $atp\alpha 1B1$  (red node) occurred before the divergence of the lineage leading to the sucker fish (Catostomus commersoni), suggesting that this duplication may have been due to the known large-scale teleost genome duplication.

# Map Positions of the Zebrafish Catalytic Subunits

To examine the origin of the duplications, we mapped all of these genes by somatic cell radiation hybrid (RH) mapping (Geisler et al. 1999). As shown in Figure 2,  $atp\alpha 2$  is on linkage group 2 (LG 2), whereas  $atp\alpha 1A1$  and  $atp\alpha 1B1$  both map to LG 1 in the z9394–z9382 interval. The  $atp\alpha 3A$  gene maps to LG 16 close to the *Hoxab* cluster, and  $atp\alpha 3B$  maps to LG 19 near the *Hoxaa* cluster. Both  $atp\alpha 1A2a$  and  $atp\alpha 1B1$ , whereas  $atp\alpha 1B2$  maps to LG 1 in the same interval as  $atp\alpha 1A1$  and  $atp\alpha 1B1$ , whereas  $atp\alpha 1B2$  maps to LG 9 between markers z9112 and z20031 (Shimoda et al. 1999).

# Expression of Zebrafish αl Isoforms

If the DDC model of subfunctionalization of genes after duplication applies, expression patterns should have been partitioned among the zebrafish ATPases relative to, for example,

Evolution of Na+/K+ ATPase  $\alpha$  Subunit Genes

atpα1A1	1	MGLGTGNDKWKLAATSEDEAKAPRKGKKKQKDMDELKKEVDLDDHKLTLDELHRKYGTDLTRGLSSSRAKEVL
atpα1B1	1	MGRGEGREQYELAATSEQGGKKS <mark>K</mark> SKGKNEKDKDMDELKKEVDLDDHKLSLDELTRKY <mark>N</mark> TDLTPGLSGTRAKEIL
atpα2	1	MGRGYGHESSPEAAPTGGKRKKKDKDLDELKKEVSLDDHKLTLDELSTRYG <mark>V</mark> DL <mark>A</mark> RGLTHKRA <mark>M</mark> EIL
atpα1A1	74	ARDGPNALTPPPTTPEWVKFCKQLFGGFSTLLWIGAILCFLAYGIQAASEDDP <mark>T</mark> NDNLYLGIVLAGVVIITGCFS
atpα1B1	76	ARDGPNALTPPPTTPEWVKFCKQLFGGFSTLLWIGAILCFLAYGI <mark>L</mark> AASE <mark>EEPA</mark> NDNLYLGIVLSAVVMITGCFS
atpα2	68	ARDGPNALTPPPTTPEWVKFCKQLFGGFS <mark>I</mark> LLWIGAILCFLAY <mark>SIQAATEDEPV</mark> NDNLYLGVVLSAVVIITGCFS
atpα1A1	149	YYQEAKSSKIMESFKNLVPQQALVVRDGEKK <mark>S</mark> INAEEVVAGDLVEVKGGDRIPADLRIISAHGCKVDNSSLTGES
atpα1B1	151	YYQEAKSSKIMDSFKNLVPQQALVIRDGEKKNINAEEV <mark>A</mark> VGDLVEVKGGDRIPADLRIISAHGCKVDNSSLTGES
atpα2	143	YYQEAKSSRIMDSFKNMVPQQALVIRDGEK <mark>LQINAEEVVQ</mark> GDLVEIKGGDRVPADLRIIS <mark>SS</mark> GCKVDNSSLTGES
atpα1A1	224	EPQTRTPDFSNDNPLETRNIAFFSTNCVEGTARGIVINTGDPTVMGRIATLAS <mark>S</mark> LE <mark>G</mark> SQTPI <mark>AK</mark> EIEHFIHIITG
atpα1B1	226	EPQTRTPDFSNDNPLETRNIAFFSTNCVEGTARGIVINTGDRTVMGRIATLASGLEVG <mark>R</mark> TPISTEIEHFIHIITG
atpα2	218	EPQTRSPEFTHENPLETRNISFFSTNCVEGTAHGIVI <mark>R</mark> TGDHTVMGRIATLASGLEVGQTPINVEIEHFIHIITG
atpα1A1	299	VAVFLGVSFFILSLILGY <mark>G</mark> WLEAVIFLIGIIVANVPEGLLATVTVCLTLTAKPMAKKNCLVKNLEAVETLGSTST
atpα1B1	301	VAVFLGVSFFILSLILGYSWLEAVIFLIGIIVANVPEGLLATVTVCLTLTAKPMAKKNCLVKNLEAVETLGSTST
atpα2	293	VAIFLGMSFFILSTILGYTWLEAVIFLIGIIVANVPEGLLATVTVCLTLTAKPMARKNCLVKNLEAVETLGSTST
atpα1A1 atpα1B1 atpα2	374 376 368	$\label{eq:construction} ICSDKTGTLTQNRMTVAHMWFDSQ1HEADTTENQSGTTFDRSSPTWSALARVAGLCNRAVFLADQRNVPILKRDTICSDKTGTLTQNRMTVAHMWFDNQIHEADTTENQSGTSFDRSSATWAALARVAGLCNRAVFLABQENIPILKRDTICSDKTGTLTQNRMTVAHMWFDNQIHEADTTEDQSGCDFDKSSPTWFSLSRVGGLCNRAVFKAGQEEIPIRTRDT$
atpα1A1	449	AGDASESALLKCIELCCGSV <mark>NEMPEKYT</mark> KIAEIPFNSTNKYQLSIH <mark>K</mark> NPNS <mark>SB</mark> PKHLLVMKGAPERILDRCST
atpα1B1	451	AGDASESALLKCIELCCGSV <mark>KEMPEKYN</mark> KISEIPFNSTNKYQLSIH <mark>O</mark> NPNS <mark>NNTB</mark> SKHLLVMKGAPERILDRCSS
atpα2	443	AGDASESALLKCVEILS <mark>SNVETLRENNRKVAEIPFNSTNKYQLSIHELEDSPTG</mark> HLLVMKGAPERILDRCST
atpα1A1	522	ILIQGKQQPLDEEMKDAFQNAYVELGGLGEPVLGFCHB <mark>C</mark> LPDDQFPEGB <mark>A</mark> FDTEEVNFPTENLCFVGLMSMIDPP
atpα1B1	526	ILIQGKEQPLDDEMKDAFQNAYLELGGLGERVLGFCHB <mark>N</mark> LPDBQFPB <mark>D</mark> FQFDTEEVNFPTENLCFIGLMSMIDPP
atpα2	515	IMING <mark>EBF</mark> PIDDDWMDAFQ <mark>G</mark> AYMELGGLGERVLGFCH <mark>LFLSPS</mark> QFP <mark>R</mark> GF <mark>B</mark> FDCDDVNFPVNQLCFIGLISMIDPP
atpα1A1	597	RAAVPDAVGKCRSAGIKVIMVTGDHPITAKAIAKGVGIISEGNETVEDIAARLNIPV <mark>GEVNPRDAKACVVHGGEL</mark>
atpα1B1	601	RAAVPDAVGKCRSAGIKVIMVTGDHPITAKAIAKGVGIISEGNETVEDIAARLNIPI <mark>NE</mark> VNPRDAKACVIHGGDL
atpα2	590	RAAVPDAVGKCRSAGIKVIMVTGDHPITAKAIAKGVGIISEGNETVEDIA <mark>B</mark> RLNIPLSOVNPRBAKACVVHG <mark>S</mark> DL
atpα1A1	672	K <mark>S</mark> MS <mark>EBE</mark> LDDILKHHTEIVFARTSPOOKLIIVEGCQROGAIVAVTGDGVNDSPALKKADIGVAMGIAGSDVSKQA
atpα1B1	676	KDIS <mark>PEQ</mark> LDDVLKHHTEIVFARTSPOOKLIIVEGCQRQGAIVAVTGDGVNDSPALKKADIGVAMGIAGSDVSKQA
atpα2	665	KDMS <mark>AB</mark> YLDDILFNHTEIVFARTSPOOKLIIVEGCQRQGAIVAVTGDGVNDSPALKKADIGVAMGI <mark>T</mark> GSDVSKQA
atpα1A1	747	ADMILLDDNFASIVTGVEEGRLIFDNLKKSIAYTLTSNIPEISPFLLFIIANIPLPLGTVTILCIDLGTDMVPAI
atpα1B1	751	ADMILLDDNFASIVTGVEEGRLIFDNLKKSIAYTLTSNIPEITPFLLFIIANIPLPLGTVTILCIDLGTDMVPAI
atpα2	740	ADMILLDDNFASIVTGVEEGRLIFDNLKKSIAYTLTSNIPEITPFLLFIIA <mark>S</mark> VPLPLGTVTILCIDLGTDMVPAI
atpα1A1	822	SLAYESAESDIMKROPPN <mark>A</mark> KTDKLVNERLISMAYGQIGMMQAVAGFFSYFVILAENGFLPS <mark>D</mark> LVGIRVNWDDKYV
atpα1B1	826	SLAYEAAESDIMKROPPNPKTDKLVNERLISTAYGQIGMIQALAGFFTYFVILAENGFLPSSLLGIRVFWDDKYV
atpα2	815	SLAYESAESDIMKROPPNPKTDKLVNERLISMAYGQIGMIQALGGFFTYFVIMAENGFLP <mark>O</mark> FLLGIRIDWDDRTV
atpα1A1 atpα1B1 atpα2	897 901 890	NDLEDSYGQQWTYE <mark>R</mark> RKIVEFTCHTAFF <mark>A</mark> SIVIVQWADLIICKTRRNSTVQQGMRNKILIFGLFEETALAAFLSY NDLEDSYGQQWTYEQRKIVEFTCHTAFFTSIVIVQWADLIICKTRRNSVFQQGMKNKILIFGLF <mark>A</mark> ETALAAFLSY NDLED <mark>G</mark> YGQQWTYEQRKITEFTCHTSFFVSIVVVQWADLIICKTRRNSVFQQGMKNKILIFGLF <mark>A</mark> ETALAAFLSY *
atpα1A1	972	CPGMDVALRMYPLKE <mark>C</mark> WWFCAPPYSLLIFYYD <mark>EA</mark> RRYILERSPGGWVE <mark>O</mark> ETYY
atpα1B1	976	CPGMDVALRMYPLKENWWFCAPPYSLLIFTYDEIRR <mark>L</mark> ITRSPGGWVE <mark>R</mark> ETYY
atpα2	965	CPGMDVALRMYPLK <mark>IM</mark> WWFCAPPYSLLIFYYDEIRRFILRRNPGGWVE <mark>I</mark> ETYY

**Figure 1** Predicted amino acid alignment of the three zebrafish sodium pump  $\alpha$  subunits. Identities are shown in black boxes, similarities in gray. The TGES motif is underlined and the phosphorylated aspartate is shown by a triangle. The conserved PKA phosphorylation site is marked by an asterisk. The broken line indicates the ankyrin repeat binding site.

the single mammalian  $\alpha 1$  gene. Therefore, we examined the expression of  $atp\alpha 1A1$  and  $atp\alpha 1B1$ .

The expression patterns of  $atp\alpha 1A1$  and  $atp\alpha 1B1$  in early development are quite distinct. At the 17 somite stage,  $atp\alpha 1A1$  is expressed in the intermediate mesoderm and, at lower levels, in the otic placode (Fig. 3A,D), whereas the  $atp\alpha 1B1$  gene is expressed in the optic cup, the developing nervous system, the otic placode, and, weakly, in the intermediate mesoderm (Fig. 3B,E). The  $atp\alpha 2$  gene is expressed in the developing somites (Fig. 3C,F). Thus, at this early stage,  $atp\alpha 1A1$  is the dominant kidney isoform, whereas  $atp\alpha 1B1$  is expressed in a variety of other tissues.

By 24 h postfertilization (hpf),  $atp\alpha 1A1$  is still expressed in the nephric duct (Fig. 3G) and, to a lesser extent, in the otic vesicle. We also detect weak expression of  $atp\alpha 1B1$  in the kidney, but it remains expressed in many tissues (Fig. 3H). Expression of  $atp\alpha 2$  remains exclusive to the myogenic lineage (Fig. 3I). At 48 hpf, expression of  $atp\alpha 1A1$  in the ear has increased and is comparable with the levels seen in the nephric duct (Fig. 3J). The dominant isoform in the nervous system is



**Figure 2** Phylogenetic analysis of vertebrate sodium pump catalytic subunits. Invertebrate sequences are used to root the tree. Gene duplication events are indicated by colored rhomboids. The tree is drawn to scale with the horizontal distances joining two nodes proportional to the number of amino acids substitutions. Scale bar, 10% substitution. The GenPept protein identification numbers for each gene are given. Genes in boldface represent those described in the present study. Partial sequences were not used in the analysis. For clarity,  $\alpha$ 1 sequences for mammals are grouped into one branch because the number of amino acid substitutions are small. The mammalian sequences included *Equus caballus* (89128), *Canis familiaris* (1703466), *Ovis aries* (67958), *Rattus norvegicus* (92517), *Homo sapiens* (88220), and *Sus scrofa* (89248). Chromosome assignments for the zebrafish subunits were obtained by radiation hybrid mapping and are indicated at *right*.

 $atp\alpha 1B1$  (Fig. 3K). It is also transiently expressed in the pectoral fin buds (Fig. 3K) and the heart, in which the ventricular myocardium expresses much higher levels of  $atp\alpha 1B1$ , than does the atrium (Fig. 3L). The  $atp\alpha 2$  gene remains in the somites and is also seen in the fin buds. (Fig. 3M).

At 96 hpf,  $atp\alpha 1B1$  is highly expressed in the brain (Fig. 3O). Low levels of  $atp\alpha 1A1$  can also be detected in this tissue (Fig. 3N). Both genes are expressed in the pronephric tubule and gut to a similar level (Fig. 3N,O). However, only the  $atp\alpha 1B1$  paralog is detected in the developing liver (Fig. 3, cf. N and O).

# DISCUSSION

# Tandem and Ploidy Gene Duplications

Invertebrate genomes contain one sodium pump catalytic

subunit (Emery et al. 1995). We show here that the three vertebrate paralogs originated by gene duplications that preceded the last common ancestor of jawed vertebrates (which is represented by the node in the  $\alpha$ 1 subtree from which the lineage to *Torpedo californica* originated; Fig. 2). We find that the zebrafish genome encodes at least eight subunits that fall into the  $\alpha$ 1,  $\alpha$ 2, or  $\alpha$ 3 subfamilies. Our phylogenetic analysis indicates that a gene duplication of the  $\alpha$ 1 gene took place in the bony fish lineage, creating the  $\alpha$ 1A and  $\alpha$ 1B subfamilies. Further duplications within each of the branches generated the additional  $\alpha$ 1 genes found in the zebrafish awaits further genomic analysis of other fish.

Combining data from the phylogenetic analysis and mapping, a clear picture of the gene duplication events in teleosts emerges (Fig. 4). The first duplication, which created the  $\alpha$ 1A and  $\alpha$ 1B subfamilies, was likely a tandem duplication because (1) it preceded the known large-scale duplication, which occurred after divergence of the lineage leading to eels, and (2)  $\alpha$ 1A and  $\alpha$ 1B both map to the same small interval on LG 1 (Fig. 2).

The duplication giving rise to the two  $\alpha$ 3 paralogs and the duplication in the  $\alpha 1B$  subfamily are likely due to the teleost large-scale duplication. Both the linkage data and the phylogenetic analysis favor this interpretation. The two  $\alpha 3$ paralogs map to LG 16 and LG 19, linked to the duplicated Hoxa clusters (Amores et al. 1998; Gates et al. 1999; Woods et al. 2000), and the duplication occurred prior to divergence of the lineage leading to *Tilapia*. Similarly, the  $\alpha 1B$  paralogs map to LG 1 and LG 9, which have been suggested to contain paralogous regions, as teleost-specific members of other gene families, such as distalless and engrailed, map to these chromosomes (Gates et al. 1999; Woods et al. 2000). Here too, the phylogenetic position of the duplications is consistent with the large-scale duplication. Duplicates of the  $atp\alpha 1A$  lineage are unlikely to be due to the large scale gene duplication, as both duplications follow the species divergence node with Tilapia. Radiation hybrid mapping of all three genes indicates that they lie in the same area of LG 1, suggesting that they arose by tandem duplication.

#### Divergence of Expression and the DDC Model

Gene duplications have been proposed to be an important mechanism driving diversification during evolution. According to classical models, duplicated genes can accumulate mutations and acquire novel functions. Alternatively, they may be lost, such as one of the *Hoxd* clusters in zebrafish, or may degenerate to a pseudogene, as *Hoxaa-10* has in zebrafish (Stellwag 1999). According to the DDC model (Force et al. 1999), some degenerative mutations may actually favor the preservation of both paralogs through complementation of their subfunctions.

Thus far, evidence for the DDC model has come from genes encoding developmental regulators. The zebrafish genome contains two *engrailed1* genes expressed in two distinct locales of a few cells each, in the fin buds and spinal cord (Force et al., 1999). The expression pattern of the duplicated *sox11* genes in zebrafish is also partitioned. Whereas the single *sox11* gene is expressed throughout the somites in the mouse, two zebrafish paralogs share this expression domain; *sox11a* is expressed anteriorly and *sox11b* posteriorly (de Martino et al., 2000). Here we present data that extend this model to genes essential for survival of every cell in the body. Interestingly, data from studies of the tissue-specific expression of



Evolution of Na+/K+ ATPase  $\alpha$  Subunit Genes

**Figure 3** Whole-mount expression of  $atp\alpha 1A1$ ,  $atp\alpha 1B1$  and  $atp\alpha 2$  during zebrafish development. Parallel expression analysis of zebrafish sodium pump catalytic subunits at the 17 somite, 24 hpf, 48 hpf, and 96 hpf stages. At the 17 somite stage,  $atp\alpha 1A1$  is expressed in the intermediate mesoderm and at lower levels in the otic placode (*A*,*D*);  $atp\alpha 1B1$  is expressed in the otic placode, optic cup and throughout the brain (*B*,*E*);  $atp\alpha 2$  is expressed in the somites (*C*,*F*). At 24 hpf,  $atp\alpha 1A1$  expression is seen in the nephric duct (*G*) and lower levels in the otic vesicle;  $atp\alpha 1B1$  is expressed in the brain and nephric duct (*H*), whereas  $atp\alpha 2$  expression remains exclusive to the muscle lineage (*I*). At 48 hpf,  $atp\alpha 1A1$  expression is detected in the otic vesicle and nephric duct (*J*);  $atp\alpha 1B1$  expression remains in the brain, but is now also seen in the fin buds (*K*) and the heart (asterisk in *L*);  $atp\alpha 2$  is also detected in the fin buds (*M*). (*N*, *O*) Expression of  $atp\alpha 1B1$  at 96 hpf, note the absence of  $atp\alpha 1A1$  expression in the liver and the strong expression of  $atp\alpha 1B1$  in the brain. (im) Intermediate mesoderm; (op) otic placode; (oc) optic cup; (nd) nephric duct; (ov) otic vesicle; (fb) fin bud; (pt) pronephric tubule; (g) gut; (lv) liver.

isozymes of lactate dehydrogenase and alcohol dehydrogenase (Li et al. 1983; Edenberg 2000) are also consistent with a DDC mechanism; future genetic mapping and evolutionary anlyses of these genes may provide further insights into the molecular mechanisms of DDC.

The DDC model predicts that duplicate genes partition functions normally performed by the original gene. Our expression results for  $atp\alpha 1A1$  and  $atp\alpha 1B1$  are consistent with this model. The  $\alpha 1$  isoform in mice is expressed in the kidney, heart, brain, and gut. In zebrafish, the  $atp\alpha 1A1$  gene is the dominant kidney isoform expressed very early in nephrogenesis and at very high levels. The  $atp\alpha 1B1$  gene is highly expressed in the brain. The other  $\alpha 1$  paralog ( $atp\alpha 1A1$ ) is also expressed in the brain, but much later in development and at lower levels. Only *atp* $\alpha$ *1B1* expression is detected in the heart and liver. The model further predicts that the broader the expression domains associated with a gene, the more likely duplicated paralogs will be maintained. The  $\alpha$ *1* gene in mammals is ubiquitously expressed, performing essential functions in all cell types of the organism. In other words, it contains the greatest potential number of subfunctions. Consistent with this prediction, the  $\alpha$ *1* family in zebrafish contains the greatest number of paralogs with five.

Prior data for the *engrailed1* genes indicate that partitioning of expression domains can follow a chromosomal duplication event. Our mapping, expression, and evolutionary data for the  $\alpha$  family members are consistent with these prior observations and reveal, as well, examples of subfunctional-



**Figure 4** Diagram illustrating the gene duplication events for the  $\alpha 1$  gene in zebrafish. (A) A tandem duplication generated two  $\alpha 1$  paralogs as follows:  $\alpha 1A$ , shown in gray and  $\alpha 1B$ , hashed. (B) The tandem is duplicated by the large scale duplication in the bony fish lineage  $\alpha 1A$  and  $\alpha 1B$ . (C) Present-day zebrafish. One duplicate pair is found on LG 1, where the  $\alpha 1A$  paralog has undergone two further gene duplications to generate three genes ( $\alpha 1A1$ ,  $\alpha 1A2a$ , and  $\alpha 1A2b$ ). We do not have evidence for the existence of the other  $\alpha 1A$  paralog. It may have been lost or remained as a pseudogene and is indicated by a question mark. The  $\alpha 1B$  paralogs have give rise to the  $\alpha 1B1$  gene on LG 1 and  $\alpha 1B2$  on LG 9. The arrangement of the genes on the chromosome is arbitrary.

ization following a tandem duplication event. These data lend further support to the DDC model, extending it to ubiquitously expressed genes and to tandem duplications.

# **METHODS**

#### Zebrafish Lines

Fish were raised and maintained in the Cardiovascular Research Center fish facility at the Massachusetts General Hospital. Embryos were kept in E3 medium at 28.5°C and staged according to somite number or hours postfertilization (hpf) (Kimmel et al. 1995). Wild-type fish were of the Tübingen background (obtained from Dr. Christianne Nüsslein-Volhard, Tübingen).

#### cDNA Cloning and Sequence Analysis

We identified an expressed sequence tag (accession no. AA494679) encoding a zebrafish  $\alpha$  subunit in the public databases and PCR primers were designed to amplify a portion

of this coding sequence. We used the PCR product to screen  $5 \times 10^6$  plaques at low stringency from a 24-h zebrafish cDNA XZAP Express library (Stratagene). We selected 20 plaques for further study of ~300 positives obtained. Ten strong positives and ten weaker positives were selected and rescued. Full-length clones were completely sequenced on both strands and characterized. We retrieved homologs from GenBank by BLAST search, and aligned the amino acid sequences with CLUSTALW (Thompson et al. 1994). Regions of uncertain homology around gaps were excluded from subsequent phylogenetic analysis by use of PROTML (Adachi and Hasegawa 1996). The tree was rooted with invertebrate sequences that were later excluded from the analyses of the vertebrate sequences to maximize the number of unambiguously homologous sequence positions. The topologies of  $\alpha 1$ ,  $\alpha 2$ , and  $\alpha 3$  subtrees were determined in independent analyses, and a global analysis allowed determination of the relative branching order of the three paralogs. Branching order in those parts of the tree that are important for our interpretation of the nature of the nodes was tested explicitly by use of the user tree option in PROTML.

#### In Situ Hybridizations

Embryos of the appropriate stage were fixed in 4% paraformaldehyde, stored in methanol and processed as described previously (Jowett 1999). The cDNA clones (in pBK-CMV) were linearized with *Bam*HI and T7 RNA polymerase used to transcribe antisense DIG-labeled riboprobes.

### Radiation Hybrid Mapping and Genotyping

We designed primers to amplify a portion of the 3' untranslated region of each  $\alpha$  subunit. We also mapped GenBank entries representing other  $\alpha$ -like genes. In Table 1, we show our proposed nomenclature consistent with the phylogenetic data presented in this study.

The cycling profile was as follows: 90 sec of denaturation at 94°C, 30 cycles with 30 sec at 94°C, 30 sec at 60°C, and 1 min at 72°C. Somatic cell radiation hybrid mapping was carried out by use of the Goodfellow T51 panel (Research Genetics) and map positions calculated by use of the RH mapping service at Boston's Children's Hospital (http://genetics. med.harvard.edu/~zonlab/).

# ACKNOWLEDGMENTS

We thank Alex Therien for critical comments on earlier versions of this manuscript and for helpful discussions and Sarah Childs for helpful criticism of the manuscript. This work was supported in part by National Institutes of Health grants RO1DK55383 and RO1HL63206 to M.C.F.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be

Gene	Forward primer	Reverse primer
atpα1A1 (AY028629)	TCAAACCGTGATGGTGGTTTTGCG	AGTAGTATGTTTCTTGTTCCACCC
atpα1A2a (11096275)	CATGGAAGTTGCTGTCAGAATGT	AAGAGAGCGTCTTCTGGCTTTAG
atpα1A2b (9789575)	ACTGCCTTACCATACGCTCTTATC	ACACCAGATTCTTTATCGCTTTTC
atpa1B1 (AY028630)	TGCATGGTTGTCTCGCTGCTCGG	AAGGCACGCAATGTCTTGGTTCACG
atpα1B2 (11067032)	TACTAGTCAAGATGTCCGCAGTTC	CGCTTTTCAAGTCTATCTGTCTCA
atpα2 (AY028628)	GCTGGGTGGAAATAGAGACCT	GACACGATTGTCTGATCCTTCA
$atp\alpha 3A$ (11067030)	TTGTTTCCATCTCTTCTCTCTCC	CGTGTGAAAATACATGCAAAGAGT
atpα3B (11096277)	AGAAACAACCATGTTACAAATCCA	GGTAAACCTTACCGATGTGTTTCT

The GenPept number for a subunit is shown in parentheses following the name. Accession numbers are given for the genes cloned for this study.

#### Evolution of Na+/K+ ATPase $\alpha$ Subunit Genes

hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

# NOTE ADDED IN PROOF

An analysis of zebrafish  $\alpha$  subunit genes has also been performed by Rajarao et al. (2001).

# REFERENCES

- Adachi, J. and Hasegawa, M. 1996. Programs for molecular phylogenies based on maximum likelihood. Institute for Statistical Mathematics, Tokyo, Japan.
- Amores, A., Force, A., Yan, Y.L., Joly, L., Amemiya, C., Fritz, A., Ho, R.K., Langeland, J., Prince, V., Wang, Y., et al. 1998. Zebrafish hox clusters and vertebrate genome evolution. *Science* 282: 1711–1714.
- Blanco, G. and Mercer, R.W. 1998. Isozymes of the Na-K-ATPase: Heterogeneity in structure, diversity in function. *Am. J. Physiol.* 275: F633–F650.
- de Martino, S., Yan, Y.L., Jowett, T., Postlethwait, J.H., Varga, Z.M., Ashworth, A., and Austin, C.A. 2000. Expression of sox11 gene duplicates in zebrafish suggests the reciprocal loss of ancestral gene expression patterns in development. *Dev. Dyn.* 217: 279–292.
- Edenberg, H.J. 2000. Regulation of the mammalian alcohol dehydrogenase genes. *Prog. Nucleic Acid Res. Mol. Biol.* 64: 295–341.
- Emery, A.M., Ready, P.D., Billingsley, P.F., and Djamgoz, M.B. 1995. A single isoform of the Na+/K(+)-ATPase alpha-subunit in Diptera: Evidence from characterization of the first extracellular domain. *Insect Mol. Biol.* **4**: 179–192.
- Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., and Postlethwait, J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151: 1531–1545.
- Gates, M.A., Kim, L., Egan, E.S., Cardozo, T., Sirotkin, H.I., Dougan, S.T., Lashkari, D., Abagyan, R., Schier, A.F., and Talbot, W.S. 1999. A genetic linkage map for zebrafish: Comparative analysis and localization of genes and expressed sequences. *Genome Res.* 9: 334–347.
- Geisler, R., Rauch, G.J., Baier, H., van Bebber, F., Brobeta, L., Dekens, M.P., Finger, K., Fricke, C., Gates, M.A., Geiger, H., et al. 1999. A radiation hybrid map of the zebrafish genome. *Nat. Genet* 23: 86–89.
- Glynn, I.M. 1993. Annual review prize lecture. 'All hands to the sodium pump'. *J. Physiol.* **462:** 1–30. James, P.F., Grupp, I.L., Grupp, G., Woo, A.L., Askew, G.R., Croyle, J. 1997. The solid set of the subscript of the subsc
- James, P.F., Grupp, I.L., Grupp, G., Woo, A.L., Askew, G.R., Croyle, M.L., Walsh, R.A., and Lingrel, J.B. 1999. Identification of a specific role for the Na,K-ATPase alpha 2 isoform as a regulator of calcium in the heart. *Mol. Cell* **3**: 555–563. Jewell, E.A., and Lingrel, J.B. 1992. Chimeric rat Na,K-ATPase alpha
- Jewell, E.A., and Lingrel, J.B. 1992. Chimeric rat Na,K-ATPase alpha 1/alpha 3\* isoforms. Analysis of the structural basis for differences in Na+ requirements in the alpha 1 and alpha 3\* isoforms. Ann. NY Acad. Sci. 671: 120–132.
- Jowett, T. 1999. Analysis of protein and gene expression. *Meth. Cell. Biol.* **59:** 663–685.
- Kawakami, K., Nojima, H., Ohta, T., and Nagano, K. 1986. Molecular cloning and sequence analysis of human Na,K-ATPase beta-subunit. *Nucleic Acids Res.* 14: 2833–2844.
- Kimmel, C., Ballard, W., Kimmel, S., Ullmann, B., and Schilling, T. 1995. Stages of embryonic development of the zebrafish. *Dev. Dyn.* 203: 253–310.
- Li, S.S., Fitch, W.M., Pan, Y.C., and Sharief, F.S. 1983. Evolutionary relationships of vertebrate lactate dehydrogenase isozymes A4 (muscle), B4 (heart), and C4 (testis). *J. Biol. Chem.* 258: 7029–7032.
- Lingrel, J.B. and Kuntzweiler, T. 1994. Na+,K(+)-ATPase. J. Biol. Chem. 269: 19659–19662.
- Malik, N., Canfield, V.A., Beckers, M.C., Gros, P., and Levenson, R.

1996. Identification of the mammalian Na,K-ATPase 3 subunit. *J. Biol. Chem.* **271:** 22754–22758.

- Martin-Vasallo, P., Dackowski, W., Emanuel, J.R., and Levenson, R. 1989. Identification of a putative isoform of the Na,K-ATPase beta subunit. Primary structure and tissue-specific expression. *J. Biol. Chem.* **264**: 4613–4618.
- Mercer, R.W., Biemesderfer, D., Bliss, D.P., Jr., Collins, J.H., and Forbush, B.D. 1993. Molecular cloning and immunological characterization of the gamma polypeptide, a small protein associated with the Na,K-ATPase. J. Cell. Biol. 121: 579–586.
- Ohno, S. 1970. *Evolution by Gene Duplication*. Springer-Verlag, Heidelberg, Germany
- Postlethwait, J.H., Woods, I.G., Ngo-Hazelett, P., Yan, Y.L., Kelly, P.D., Chu, F., Huang, H., Hill-Force, A., and Talbot, W.S. 2000. Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res.* **10**: 1890–1902.
- Postlethwait, J.H., Yan, Y.L., Gates, M.A., Horne, S., Amores, A., Brownlie, A., Donovan, A., Egan, E.S., Force, A., Gong, Z., et al. 1998. Vertebrate genome evolution and the zebrafish gene map. *Nat. Genet.* 18: 345–349.
- Price, E.M., Rice, D.A., and Lingrel, J.B. 1990. Structure-function studies of Na,K-ATPase. Site-directed mutagenesis of the border residues from the H1-H2 extracellular domain of the alpha subunit. J. Biol. Chem. 265: 6638–6641.
- Prince, V.E., Joly, L., Ekker, M., and Ho, R.K. 1998. Zebrafish hox genes: Genomic organization and modified colinear expression. *Development* 125: 407–420.
- Rajarao, S.J.R., Canfield, V.A., Mohideen, M.-A.P.K., Yan, Y.-L., Postlethwait, J.H., Cheng, K.C., and Levenson, R. 2001. The repertoire of Na, K-Atpase  $\alpha$  and  $\beta$  subunit genes expressed in the zebrafish, Danio rerio. *Genome Res.* **11**: 1211–1220.
- Shamraj, O.I. and Lingrel, J.B. 1994. A putative fourth Na+,K(+)-ATPase alpha-subunit gene is expressed in testis. Proc. Natl. Acad. Sci. 91: 12952–12956.
- Shimoda, N., Knapik, E.W., Ziniti, J., Sim, C., Yamada, E., Kaplan, S., Jackson, D., de Sauvage, F., Jacob, H., and Fishman, M.C. 1999. Zebrafish genetic map with 2000 microsatellite markers. *Genomics* 58: 219–232.
- Shull, G.E., Greeb, J., and Lingrel, J.B. 1986. Molecular cloning of three distinct forms of the Na+,K+-ATPase alpha-subunit from rat brain. *Biochemistry* 25: 8125–8132.
- Sidow, A. 1992. Diversification of the Wnt gene family on the ancestral lineage of vertebrates. *Proc. Natl. Acad. Sci.* 89: 5098–5102.
- ——.1996. Gen(om)e duplications in the evolution of early vertebrates. *Curr. Opin. Genet. Dev.* **6**: 715–722.
- Stellwag, E.J. 1999. Hox gene duplication in fish. Semin. Cell. Dev. Biol. 10: 531–540.
- Suga, H., Koyanagi, M., Hoshiyama, D., Ono, K., Iwabe, N., Kuma, K., and Miyata, T. 1999. Extensive gene duplication in the early evolution of animals before the parazoan-eumetazoan split demonstrated by G proteins and protein tyrosine kinases from sponge and hydra. J. Mol. Evol. **48**: 646–653.
- Sweadner, K.J. 1989. Isozymes of the Na+/K+-ATPase. *Biochim. Biophys. Acta* **988**: 185–220.
- Therien, A.G. and Blostein, R. 2000. Mechanisms of sodium pump regulation. *Am. J. Physiol. Cell. Physiol.* **279**: C541–C566.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22: 4673–4680.
- Woods, I.G., Kelly, P.D., Chu, F., Ngo-Hazelett, P., Yan, Y.L., Huang, H., Postlethwait, J.H., and Talbot, W.S. 2000. A comparative map of the zebrafish genome. *Genome Res.* **10**: 1903–1914.
- Zhang, Z., Devarajan, P., Dorfman, A.L., and Morrow, J.S. 1998. Structure of the ankyrin-binding domain of alpha-Na,K-ATPase. *J. Biol. Chem.* **273:** 18681–18684.

Received April 12, 2001; accepted in revised form June 4, 2001.



# Partitioning of Tissue Expression Accompanies Multiple Duplications of the Na+/K+ ATPase $\,\alpha$ Subunit Gene

Fabrizio C. Serluca, Arend Sidow, John D. Mably, et al.

Genome Res. 2001 11: 1625-1631 Access the most recent version at doi:10.1101/gr.192001

# License

**Email Alerting** Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here**.



To subscribe to Genome Research go to: https://genome.cshlp.org/subscriptions