Molecular Evolution of the Ankyrin Gene Family

Xinjiang Cai*† and Yanhong Zhang‡

*Howard Hughes Medical Institute and †Departments of Cell Biology, Biochemistry, and Neuroscience, Duke University Medical Center and ‡Department of Medicine, Duke University Medical Center

Ankyrins are membrane adaptor molecules that play important roles in coupling integral membrane proteins to the spectrin-based cytoskeleton network. Human mutations of ankyrin genes lead to severe genetic diseases such as fatal cardiac arrhythmias and hereditary spherocytosis. To elucidate the evolutionary history of ankyrins, we have identified novel ankyrin sequences in insect, fish, frog, chicken, dog, and chimpanzee genomes and explored the phylogenetic relationships of the ankyrin gene family. Our data demonstrate that duplication of ankyrin genes occurred at two different stages. The first duplication resulted from an independent evolution event specific in Arthropoda after its divergence from Chordata. Following the separation from Urochordata, expansion of ankyrins in vertebrates involved ancestral genome duplications. We did not find evidence of coordinated arrangements of gene families of ankyrin-associated membrane proteins on paralogous chromosomes. In addition, evolution of the 24 ANK-repeats strikingly correlated with the exon boundary sites of ankyrin genes, which might have occurred before its duplication in vertebrates. Such correlation is speculated to bring functional diversity and complexity. Moreover, based on the phylogenetic analysis of the ANK-repeat domain, we put forward a novel model for the putative primordial ankyrin that contains the fourth six–ANK-repeat subdomain and the spectrin-binding domain. These findings will provide guides for future studies concerning structure, function, evolutionary origins of ankyrins, and possibly other cytoskeletal proteins.

Introduction

Dynamic interactions among cytoskeletal filaments and their associated proteins regulate many fundamental processes of eukaryotic cells such as cell motility, maintenance of the structural organization and mechanical function (Pollard 2003), and generation of cell polarity (Nelson 2003). Ankyrins are a family of ubiquitously expressed membrane adaptor molecules associated with the spectrin-based cytoskeleton network and various membrane proteins (Bennett and Baines 2001). Since the initial characterization of ankyrin-R in erythrocytes (Bennett and Stenbuck 1979), studies using different model systems have revealed physiological roles of ankyrins in targeting various ion transport proteins such as the voltage-gated Na+ channel (NaV), Na+/Ca2+ exchanger (NCX), Na+/K+ ATPase (NKA), and inositol 1,4,5-triphosphate receptor (IP3R), and cell adhesion molecules such as CD44 and L1 cell adhesion molecule (L1CAM) to specialized membrane domains in striated muscle and the nervous system (Bennett and Baines 2001). Targeted disruption of ankyrin-G in the mouse cerebellum leads to the coordinate loss of NaV1.6 and L1CAM molecules at the axon initial segments (Jenkins and Bennett 2001). Mice deficient in ankyrin-B display the phenotype of a fatal cardiac arrhythmia, resulting from reduced expression level and abnormal targeting of NKA, NCX, and IP3R (Mohler et al. 2003). Ankyrins are also required for establishing lateral membrane biogenesis in polarized epithelial cells (Kizhatil and Bennett 2004). In addition to their conventional role as a scaffold protein, more recent studies have suggested that ankyrin molecules might be involved in sorting and trafficking of membrane proteins before targeting to the plasma membrane (Lopez et al. 2004; Mohler et al. 2004).

The ankyrin gene family of mammals contains three members—ANK1 (gene product ankyrin-R) (Lux, John, and Bennett 1990), ANK2 (gene product ankyrin-B) (Otto et al. 1991), and ANK3 (gene product ankyrin-G) (Kordeli, Lambert, and Bennett 1995). Canonical ankyrins comprise three highly conserved domains—an N-terminal membrane-binding domain, a central spectrin-binding domain (SBD), and a death domain. In contrast, the C-terminal domains of ankyrins are very divergent (Bennett and Baines 2001). The membrane-binding domain is composed of 24 copies of a 33-residue repeat known as the ANK-repeat. The 24 repeats are divided into four independently folded subdomains, each containing six repeats (Michaely and Bennett 1993). Atomic structures of a 12–ANK-repeat stack from human ankyrin-R (Michaely et al. 2002) have revealed that individual ANK-repeats are assembled into the characteristic “L”-shaped structure consisting of two pairs of short antiparallel α-helices followed by a long loop and an intervening β-hairpin motif that are arranged perpendicular to the helices.

SBD binds to β-spectrin with high affinity (Davis and Bennett 1984). The first 11-residue segment of the SBD of ankyrin-R seems to fold back and bind to ANK-repeat 20–24 through a peptide-in-groove interaction in an extended conformation (Michaely et al. 2002). Interaction between ankyrin-R and spectrin couples actin cytoskeleton to the plasma membrane of erythrocytes. Recently, ankyrin-B has been shown to be required for the normal localization of β2-spectrin to an intracellular compartment in neonatal cardiomyocytes (Mohler, Yoon, and Bennett 2004). In addition, studies using chimeras between ankyrin-G and ankyrin-B suggests that the SBD of ankyrin-G, but not ankyrin-B, might participate in protein pathways responsible for the lateral membrane biogenesis in human bronchial epithelial cells (Kizhatil and Bennett 2004).

Death domains are found in proteins such as Fas and tumor necrosis factor receptors involved in signal transduction that leads to apoptosis (Hofmann and Tsopp 1995). The role of the ankyrin death domain remains to be determined. In kidney, the death domain of ankyrin-G has been shown to interact with the death domain of Fas receptors and promote Fas-induced apoptosis in renal epithelia cells.
(Del Rio et al. 2004). The C-terminal regulatory domain is highly variable among ankyrin family members, and therefore likely accounts for distinct physiological functions of different ankyrin molecules. Indeed, replacement of the C-terminal domain of ankyrin-B with that of ankyrin-G completely abolishes the function of ankyrin-B in targeting IP_{3} and ryanodine receptors in neonatal cardiomyocytes (Mohler, Gramolini, and Bennett 2002).

Recent reports on the phylogeny of ζ-actinins of the spectrin superfamily have extended our understanding of their evolutionary origins (Dixon, Forstner, and Garcia 2003; Virel and Backman 2004). Mutations in the Caenorhabditis elegans ankyrin gene, unc-44, affect axonal guidance and result in uncoordinated movement (Otsuka et al. 1995), indicating the importance of ankyrins in animal development. However, a comprehensive analysis on the evolutionary relationship of the ankyrin family members has been hindered by the limited number of ankyrin sequences in the public database. As of December 2004, an examination of the 6,207 ANK-repeat protein sequences (IPR02110) at the InterPro protein family database (http://www.ebi.ac.uk/interpro/) revealed 31 ankyrin sequences and then pared down to only seven nonredundant ankyrin sequences.

Here, we have undertaken the first comprehensive and rigorous phylogenetic analysis of ankyrin sequences identified by extensive screenings of currently available whole-genome sequences. We have also examined the chromosomal distribution of ankyrin genes to explore the potential mechanisms leading to their expansion in vertebrates. Moreover, we have for the first time systematically examined the evolution of ankyrin repeats in terms of exon structure and repeat subdomains.

Materials and Methods

Sequence Database Searches and Multiple Sequence Alignments

BlastP (Altschul et al. 1997) searches were performed to screen the National Center for Biotechnology Information (NCBI) and UniProt/SwissProt protein databases using C. elegans ankyrin (SwissProt accession number Q17486), Homo sapiens ankyrin-B (SwissProt accession number Q01484), or their truncated sequences containing the last six ANK-repeats and the SBD. TBlastN (Altschul et al. 1997) searches for new ankyrin sequences were conducted using available genome sequences at NCBI (http://www.ncbi.nlm.nih.gov/BLAST/), the Ensembl Genome Browser (http://www.ensembl.org/Multi/blastview), and Ciona intestinalis genome v1.0 (http://genome.jgi-psf.org/ciona4/ciona4.home.html) (Dehal et al. 2002). Identified DNA segments and their neighboring regions of genomic sequences were used to search coding exons or open reading frames using the GENSCAN program (Burge and Karlin 1997) (http://genes.mit.edu/GENSCAN.html) or GENE FINDING programs at the Softberry Web site (http://www.softberry.com/berry.phtml?topic=index&group=programs&subgroup=genfind).

Protein sequence alignments were performed using ClustalX (version 1.83) (Jeanmoignon et al. 1998) and were subsequently manually edited to improve alignments in Genedoc (K. Nicholas, H. Nicholas, and Deerfield 1997) as previously described (Cai and Lytton 2004). Partial and duplicated sequences as well as alternatively spliced isoforms were excluded to confirm that only one ankyrin gene product of each type from each species was analyzed further for phylogenetic analysis.

Phylogenetic Analysis

Multiple alignment files adjusted by Genedoc were exported to files in PHYLIP format. All sequences were initially subject to the chi-square analysis for homogeneity of amino acid composition implemented in Tree-Puzzle (version 5.2) (Schmidt et al. 2002). Sequences that failed in the chi-square test or contained more than 15% gaps in the refined alignments were extruded. Protein maximum likelihood (ML) analyses were performed using PROML within the PHYLIP package (version 3.64) (Felsenstein 1996), with the Jones-Taylor-Thornton amino acid substitution matrix, a mixed four-category discrete-gamma model of among-site rate variation plus invariable sites, and global arrangements and randomized input order of sequences (10 jumbles). The η parameter and the fraction of invariable sites were estimated using Tree-Puzzle. Protein ML-distance bootstrap values by analysis of 100 resampled data sets obtained using SEQBOOT (PHYLIP package) were calculated using PUZZLEBOOT (http://www.tree-puzzle.de) with the same settings as those employed for PROML.

Phylogenetic analyses with bootstrap values were also performed with the maximum parsimony (MP) method and the neighbor-joining (NJ) analysis as previously described (Cai and Lytton 2004), using PROTPARS, and PROTDIST and NEIGHBOR programs in the PHYLIP package (Felsenstein 1996), respectively. The Jones-Taylor-Thornton amino acid substitution matrix and a rate heterogeneity model with the gamma-distributed rates over four categories were used in the PROTDIST program. Weighted NJ analysis was performed with WEIGHBOR (Perriere and Gouy 1996), using the distance matrix calculated by PROTDIST or Tree-Puzzle. The TREEVIEW program (version 1.6.6) (Page 1996) and NJPLOT (Perriere and Gouy 1996) were used to display the phylogenetic trees.

Mapping the Genomic Locations of Members of Ankyrin and Selected Gene Families

Extensive text-based searches for the chromosomal locations of ankyrin and selected gene families in the genomes of C. elegans, Drosophila melanogaster, Gallus gallus, and H. sapiens were performed at the NCBI Map Viewer Web site (http://www.ncbi.nlm.nih.gov/mapview/). Also, each genome sequence was examined with TBlastN searches using at least one gene homolog from each gene family to reveal additional chromosomal loci of members of these gene families. The chromosomal locations were described with the genetic linkage map (centimorgans), the sequence map (kilobases), or the cytogenetic map (banding pattern).

Results and Discussion

To date, ankyrin genes have been cloned and characterized in C. elegans (Otsuka et al. 1995), D. melanogaster
Table 1
List of Ankyrin Proteins Used for Analyses

<table>
<thead>
<tr>
<th>Name</th>
<th>Organism</th>
<th>Common Name</th>
<th>Taxonomy</th>
<th>Identifier</th>
<th>Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>AgaANK-1</td>
<td>Anopheles gambiae</td>
<td>Mosquito</td>
<td>Invt.</td>
<td>55236154</td>
<td>GenBank</td>
</tr>
<tr>
<td>AgaANK-2</td>
<td>Anopheles gambiae</td>
<td>Mosquito</td>
<td>Invt.</td>
<td>Q7KD3</td>
<td>SwissProt</td>
</tr>
<tr>
<td>AmeANK</td>
<td>Apis mellifera</td>
<td>Honeybee</td>
<td>Invt.</td>
<td>48099152</td>
<td>GenBank</td>
</tr>
<tr>
<td>ChrANK</td>
<td>Caenorhabditis briggsae</td>
<td>Roundworm</td>
<td>Invt.</td>
<td>39593211</td>
<td>GenBank</td>
</tr>
<tr>
<td>CelANK</td>
<td>Caenorhabditis elegans</td>
<td>Roundworm</td>
<td>Invt.</td>
<td>Q17486</td>
<td>SwissProt</td>
</tr>
<tr>
<td>CinANK</td>
<td>Ciona intestinalis</td>
<td>Sea squirt</td>
<td>Invt.</td>
<td>26554507</td>
<td>GenBank (nt.)</td>
</tr>
<tr>
<td>DmeANK-1*</td>
<td>Drosophila melanogaster</td>
<td>Fruit fly</td>
<td>Invt.</td>
<td>24241</td>
<td>SwissProt</td>
</tr>
<tr>
<td>DmeANK-2</td>
<td>Drosophila melanogaster</td>
<td>Fruit fly</td>
<td>Invt.</td>
<td>Q7KD2</td>
<td>SwissProt</td>
</tr>
<tr>
<td>DpsANK-1</td>
<td>Drosophila pseudoobscura</td>
<td>Fruit fly</td>
<td>Invt.</td>
<td>55843078</td>
<td>GenBank (nt.)</td>
</tr>
<tr>
<td>DpsANK-2</td>
<td>Drosophila pseudoobscura</td>
<td>Fruit fly</td>
<td>Invt.</td>
<td>54641879</td>
<td>GenBank (nt.)</td>
</tr>
<tr>
<td>DyaANK-1</td>
<td>Drosophila yakuba</td>
<td>Fruit fly</td>
<td>Invt.</td>
<td>49843913</td>
<td>GenBank (nt.)</td>
</tr>
<tr>
<td>DyaANK-2</td>
<td>Drosophila yakuba</td>
<td>Fruit fly</td>
<td>Invt.</td>
<td>49841546</td>
<td>GenBank (nt.)</td>
</tr>
<tr>
<td>DreANK-B</td>
<td>Danio rerio</td>
<td>Zebratfish</td>
<td>Vt.</td>
<td>UP0000435D77</td>
<td>UniProt</td>
</tr>
<tr>
<td>DreANK-G</td>
<td>Danio rerio</td>
<td>Zebratfish</td>
<td>Vt.</td>
<td>UP0000436DF4</td>
<td>UniProt</td>
</tr>
<tr>
<td>FruANK-B</td>
<td>Fugu rubripes</td>
<td>Pufferfish</td>
<td>Vt.</td>
<td>22418278</td>
<td>GenBank (nt.)</td>
</tr>
<tr>
<td>FruANK-R*</td>
<td>Fugu rubripes</td>
<td>Pufferfish</td>
<td>Vt.</td>
<td>22423640</td>
<td>GenBank (nt.)</td>
</tr>
<tr>
<td>GgaANK-B*</td>
<td>Gallus gallus</td>
<td>Chicken</td>
<td>Vt.</td>
<td>50746759</td>
<td>GenBank</td>
</tr>
<tr>
<td>GgaANK-G</td>
<td>Gallus gallus</td>
<td>Chicken</td>
<td>Vt.</td>
<td>50749238</td>
<td>GenBank</td>
</tr>
<tr>
<td>GgaANK-R</td>
<td>Gallus gallus</td>
<td>Chicken</td>
<td>Vt.</td>
<td>58062691</td>
<td>GenBank (nt.)</td>
</tr>
<tr>
<td>TnuANK-R*</td>
<td>Tetraodon nigroviridis</td>
<td>Green spotted pufferfish</td>
<td>Vt.</td>
<td>UP0000318778</td>
<td>UniProt</td>
</tr>
<tr>
<td>XtrANK-R</td>
<td>Xenopus tropicalis</td>
<td>Clawed frog</td>
<td>Vt.</td>
<td>GENSCAN00000117559-60</td>
<td>Ensembl (Scaffold_186)</td>
</tr>
<tr>
<td>XtrANK-G</td>
<td>Xenopus tropicalis</td>
<td>Clawed frog</td>
<td>Vt.</td>
<td>GENSCAN00000077999</td>
<td>Ensembl (Scaffold_201)</td>
</tr>
<tr>
<td>MmuANK-B</td>
<td>Mus musculus</td>
<td>House mouse</td>
<td>Mam.</td>
<td>51708444</td>
<td>GenBank (nt.)</td>
</tr>
<tr>
<td>MmuANK-G</td>
<td>Mus musculus</td>
<td>House mouse</td>
<td>Mam.</td>
<td>Q61307</td>
<td>SwissProt</td>
</tr>
<tr>
<td>MmuANK-R</td>
<td>Mus musculus</td>
<td>House mouse</td>
<td>Mam.</td>
<td>Q02557</td>
<td>SwissProt</td>
</tr>
<tr>
<td>RnoANK-B</td>
<td>Rattus norvegicus</td>
<td>Norway rat</td>
<td>Mam.</td>
<td>37694350</td>
<td>GenBank (nt.)</td>
</tr>
<tr>
<td>RnoANK-R</td>
<td>Rattus norvegicus</td>
<td>Norway rat</td>
<td>Mam.</td>
<td>O70511</td>
<td>SwissProt</td>
</tr>
<tr>
<td>CfaANK-B</td>
<td>Canis familiaris</td>
<td>Dog</td>
<td>Mam.</td>
<td>57109380</td>
<td>GenBank</td>
</tr>
<tr>
<td>CfaANK-G</td>
<td>Canis familiaris</td>
<td>Dog</td>
<td>Mam.</td>
<td>57085047</td>
<td>GenBank</td>
</tr>
<tr>
<td>CfaANK-R</td>
<td>Canis familiaris</td>
<td>Dog</td>
<td>Mam.</td>
<td>57097843</td>
<td>GenBank</td>
</tr>
<tr>
<td>PrtANK-B</td>
<td>Pan troglodytes</td>
<td>Chimpanzee</td>
<td>Mam.</td>
<td>UP000036CE2C</td>
<td>UniProt</td>
</tr>
<tr>
<td>PrtANK-G</td>
<td>Pan troglodytes</td>
<td>Chimpanzee</td>
<td>Mam.</td>
<td>UP000036FA37</td>
<td>UniProt</td>
</tr>
<tr>
<td>HsaANK-B</td>
<td>Homo sapiens</td>
<td>Human</td>
<td>Mam.</td>
<td>Q01484</td>
<td>SwissProt</td>
</tr>
<tr>
<td>HsaANK-G</td>
<td>Homo sapiens</td>
<td>Human</td>
<td>Mam.</td>
<td>Q12955</td>
<td>SwissProt</td>
</tr>
<tr>
<td>HsaANK-R</td>
<td>Homo sapiens</td>
<td>Human</td>
<td>Mam.</td>
<td>P16157</td>
<td>SwissProt</td>
</tr>
</tbody>
</table>

Note—Abbreviations: Invt., invertebrate; Mam., mammal; nt., nucleotides; Vt., nonmammalian vertebrate. Naming of sequences: we describe all ankyrin members in this study using an abbreviation of genus and species name, followed by names after their phylogenetic relationship with characterized ankyrins.

* Sequences that failed in the chi-square test in Tree-Puzzle or contained more than 15% gaps in the refined alignments.

(Dubreuil and Yu 1994; Bouley et al. 2000), and H. sapiens (Lux, John, and Bennett 1990; Otto et al. 1991; Kordeli, Lambert, and Bennett 1995). Here, we were able to identify 39 nonredundant ankyrin sequences from 18 species analyzed in our study, including 15 mammalian ankyrin sequences and 24 nonmammalian ankyrin sequences. Gene names, species names, identifiers, and database origins are listed in table 1. Extensive searches of available genome sequences led to identification of novel ankyrin sequences from species in which no ankyrin sequences have been previously reported, such as marine invertebrate ascidian C. intestinalis and five nonmammalian vertebrates Danio rerio, Fugu rubripes, Tetraodon nigroviridis, Xenopus tropicalis, and G. gallus. The addition of these sequences in Urochordata and lower vertebrates is critical to explore the evolutionary history of the ankyrin family. In honeybee Apis mellifera and some fish species, additional ankyrin genes were also identified on other chromosomes or scaffolds (e.g., A. mellifera, GI 48104470). But due to incomplete genome sequencing, protein sequences encoded by these DNA fragments were too short to be included in subsequent analyses.

Duplicate Events of Ankyrin Genes Occurred at Two Different Stages

Multiple sequence alignments of 39 ankyrin sequences were manually edited, and sequences of the highly conserved regions, including primarily the membrane-binding domain and the SBD, were used for further analysis. Four sequences (see table 1) were extruded after the initial analysis with Tree-Puzzle. Repeated motifs in many proteins undergo various homogenization events (Zimmer et al. 1980), which sometimes may mask the evolutionary history normally contained in molecular sequences. Nevertheless, in this study, sequences data used in our phylogenetic analysis included both ANK-repeat domain and nonrepetitive SBD. Furthermore, results from ML, MP, and NJ approaches were compared to infer the congruent phylogeny. The unrooted phylogeny trees of the ankyrin family are shown in figure 1 and figures S1 and S2 (Supplementary Material online). Discrepancies at the terminal (mammal) branches of ankyrin-B between figure 1 and figures S1 and S2 (Supplementary Material online) were observed, implying
Bootstrap values are shown at the nodes. Trees obtained by NJ (Fig. S1) shows the evolutionary history of ankyrins from nematodes to mammals. Logenetic tree, constructed by using Tree-Puzzle and PUZZLEBOOT, inferring the reliable evolutionary analysis.

The nematode genomes (C. elegans and Caenorhabditis briggsae) contain single copies of ankyrin gene. The presence of two ankyrin members in the insect genomes suggested an initial gene duplication event occurred after the divergence into arthropoda. Clearly, the vertebrate ankyrins are grouped into three major branches, as previously defined in human (Bennett and Baines 2001). Interestingly, the two ankyrins in insect genomes do not correlate with orthologs of three vertebrate ankyrins, suggesting that the initial divergence and duplication of ankyrin might be separated from the chordate ankyrin evolution and reflect an independent arthropoda-specific late event. This hypothesis is further supported by the finding that the vertebrate urochordate C. intestinalis appears to have only a single ankyrin gene (see below).

We believe that C. intestinalis contains only one copy of ankyrin gene based on the following three reasons. Firstly, C. intestinalis appears to possess single copies of many vertebrate gene families such as Nkx2 and bHLH Hand transcription factor families (Dehal et al. 2002; Leveugle et al. 2004). It has been proposed that large-scale gene duplications occurred after the vertebrates diverged from the cephalochordates and urochordates (Holland et al. 1994). Secondly, no other homologous ankyrin sequences could be found in the current version of C. intestinalis genome sequence. Thirdly, it is unlikely that C. intestinalis might have lost two copies of ankyrin subsequently after the divergence. The C. intestinalis ankyrin (CinANK) does not branch with the three vertebrate ankyrin groups (fig. 1 and Figs. S1 and S2 [Supplementary Material online]). In contrast, if only one copy of ankyrin-B, R, or G from one of the vertebrate genomes is used in the analysis, this single copy is nevertheless grouped with one of the three vertebrate branches. The single CinANK gene subsequently evolved into three ankyrin genes, likely before the early Euteleostomi lineage, as the pufferfish T. nigroviridis contains all three vertebrate ankyrins.

Evolution of Vertebrate Ankyrins Might Involve Ancestral Genome Duplications

The phylogeny of vertebrate ankyrins was further analyzed with one outgroup chosen from invertebrates, CelANK (Figs. S3–S5, Supplementary Material online). Comparison of the overall tree topologies clearly indicates that the clades of ankyrin-G and ankyrin-B are more closely related to each other compared with the ankyrin-R clade. It seems that ankyrin-B and G were derived from an ancestral ankyrin that diverged from ankyrin-R before the duplication event.

In 1970, Susumo Ohno first proposed that vertebrate genomes evolved through rounds of large-scale duplications (Ohno 1970), possibly in the form of polyploidization. The area of ancient genome duplication in vertebrate lineage has been intensively investigated (Spring 2002; Wolfe and Li 2003). Recent phylogenetic analyses and comparative analyses of current genome duplications provide strong evidence of large segmental (en bloc) duplication of vertebrate genomes (Pebusque et al. 1998; Abi-Rached et al. 2002; McLysaght, Hokamp, and Wolfe 2002; Robinson-Rechavi, Boussau, and Laudet 2004)—but also see references suggesting coordinated arrangements of related gene families known to interact (Martin 2001; Pennisi 2001). Here, we are interested in investigating whether expansion of the vertebrate ankyrin family resulted from genome duplications. If the vertebrate ankyrin evolution was derived from genome duplications, other gene families located in the same chromosome regions should also be involved.

The single ankyrin gene in C. elegans is located on chromosome IV (cM 2.91). Members of several other gene families also placed on chromosome IV were chosen for genomic analysis, including UNC-5, choline acetyltransferase (CAT), vesicular monoamine transporter (VMAT), IP3R, and LICAM (Table S1, Supplementary Material online). UNC-5 is a transmembrane protein initially found to be involved in axon guidance in C. elegans (Leung-Hagesteijn et al. 1992; Hamelin et al. 1993). Recent studies demonstrate that members of UNC-5 family are netrin receptors mediating repulsion of axon migration (Wadsworth 2002). CAT and VMAT have been used in a previous study (Pebusque et al. 1998). IP3R and LICAM are integral membrane proteins known to interact with the ankyrin cytoskeleton network (Chen, Ong, and Bennett 2001; Mohler, Gramolini, and Bennett 2002).

We combined Map Viewer searches using keyword queries and extensive TBLastN searches of genomes from C. elegans, D. melanogaster, G. gallus, and H. sapiens.
at the NCBI Web site. The genomic locations of five gene families in four species are shown in Table S1 (Supplementary Material online). In C. elegans, each gene family contains a single copy of the gene on chromosome IV. In D. melanogaster, almost all genes remain to be a single copy except ankyrin. As discussed above, in insects, a recent and independent gene duplication event of ankyrin likely occurred. Genomic locations of these genes in C. intestinalis remain unknown because its genomic assembling has not been finished.

In vertebrates, ankyrin, VMAT, and IP₃R split into three members, and LICAM and UNC-5 evolved to have four genes (Table S1 [Supplementary Material online] and fig. 2). It is predicted that the first round of genome duplication occurred just after the divergence of Cephalochordata and Craniata ancestors. This hypothesis is supported by the finding that urochordate C. intestinalis has a nonduplicated genome (Leveugle et al. 2004). Ankyrin, UNC-5, and VMAT genes are all present on chromosome 10 and chromosome 8, along with the fibroblast growth factor receptor (FGFR) and adrenergic receptor (ADR) genes (Pebusque et al. 1998) and genes of several other families—dickkopf, EIF4E-binding protein and calcineurin catalytic unit (Table S1, Supplementary Material online). Therefore, it seems likely that at least a part of chromosomes 10 and 8 might be derived together from the first round of duplication. Duplication of VMAT-2 (10q11.2) and VMAT-3 (10q24.3-25.1) might have evolved through a centromere-proximal duplication and an intrachromosomal rearrangement mechanism on chromosome 10q11 (Crosier et al. 2002). Phylogenetic analyses have revealed that closer evolutionary relationships between HsaANK-G (chr10) and HsaANK-B (chr4) compared with HsaANK-R (chr8) (described above), UNC-5B (chr10) and UNC-5C (chr4) compared with UNC-5D (chr8) and UNC-5A (chr5) (data not shown), HsaFGFR-2 (chr10) and HsaFGFR-3 (chr4) compared with HsaFGFR-1 (chr8) and HsaFGFR-4 (chr5) (Coulier et al. 1997), and ADRA2A (chr10) and ADRA2C (chr4) compared with ADRA1A (also known as ADRA1C) (chr8) and ADRA1B (chr5) (Ruuskanen et al. 2004). Phylogeny of dickkopf, EIF4EBP, and PPP3C genes also indicates the same pattern (data not shown). These data suggest that at least a large segment of chromosomes 4 and 5 were likely duplicated from chromosomes 10 and 8 in the second round, respectively (fig. 2). It seems that a loss of related genes might have occurred after the second round of duplication, that is, no ankyrin gene could be found on chromosome 5, possibly due to functional redundancy. Analysis of yeast (Wolfe and Shields 1997) and plant genomes (Kashkush, Feldman, and Levy 2002) has suggested considerable gene loss following the duplication of the whole genome.

In contrast to the hypothesis of maintenance of functionally coupled gene families on paralogous chromosomes (Martin 2001; Pennisi 2001), IP₃R and LICAM genes, which were originally located on chromosome IV along with ankyrin, UNC-5, CAT, and VMAT in C. elegans, were separated in vertebrate genomes (Table S1 [Supplementary Material online] and fig. 2). This separation seems to have occurred before genome duplications in vertebrates. In addition, neither spectrin nor other integral membrane proteins currently known to physiologically associate with ankrys, such as NCX, ryanodine receptors, and anion exchangers, were found to coevolve through coordinated chromosome locations with ankyrin (data not shown). Interestingly, although chromosomal locations of LICAM and IP₃R genes are coordinated in the G. gallus genome, only HsaCHL1 in the LICAM gene family maintains its coordinated genomic location with HsaIP3R genes (Table S1, Supplementary Material online). HsaNrCAM and HsaNfasc genes seem to have been rearranged to chromosomes different from their IP₃R counterparts in the H. sapiens genome (also in the Mus musculus genome; data not shown). Taken together, possible loss of gene paralogs during vertebrate evolution, especially after the second round of duplications, and lineage-specific chromosomal rearrangements of gene paralogs in the recent evolution history might contribute to the complexity of the genome duplication hypothesis in vertebrates (Spring 2002; Wolfe and Li 2003).

Our observation suggests that molecular evolution of the ankyrin family likely involved ancestral segmental genome duplications in vertebrates. Due to the limited number of gene families analyzed in this study based on our primary research interests, it should be noted that the conclusion here might not necessarily imply large-scale genome duplications in vertebrates. Caenorhabditis elegans and D. melanogaster are highly derived organisms even within their own phyla, which makes it difficult to detect the
corresponding chromosome regions in invertebrates that could give rise to many other paralogous domains in vertebrates suggesting genome duplication (Spring 2002). The ongoing genome assembly project of C. intestinalis as a primitive member of Chordata may provide promising opportunities to study the genome duplication hypothesis. Large-scale genome duplications causing a severe imbalance in gene product and resulting in rapid gene losses likely contributed to the early stage of vertebrate evolution, whereas segmental (small scale) duplications occurred constantly in vertebrate evolution (X. Gu, Wang, and J. Gu 2002). Nevertheless, our observation is consistent with previous studies of segmental genome duplication of vertebrate genomes (Pebusque et al. 1998; Abi-Rached et al. 2002; McLysaght, Hokamp, and Wolfe 2002; Robinson-Rechavi, Boussau, and Laudet 2004).

Evolution of the 24 ANK-Repeats Correlated with the Exon Boundary Sites of Ankyrin Genes

A striking feature of all HsaANK genes is their unique organization of the coding exons for the 24 ANK-repeats (fig. 3). With four exceptions that one exon encodes two ANK-repeats (repeats 5–6, 10–11, 15–16, and 19–20), the exon boundaries of HsaANK within the ANK-repeat region are corresponding to exactly a single ANK-repeat for each exon (99 nt/33 amino acids each). In contrast, CelANK and DmeANK contain three large coding exons that encode more than 20 ANK-repeats. The genomic structure of CinANK shows similar features with HsaANK—18 individual ANK-repeats are encoded by single exons, while all fish DreANK genes exhibit the same exon structure as HsaANK (fig. 3). These data suggest that a common primitive ancestral ankyrin might have evolved to contain the modern exon structure after the divergence from Urochordata but before the expansion in vertebrates. Alternatively, although less likely, the modern exon structure might have been adapted from convergence evolution after the duplication in vertebrates.

The canonical ANK-repeat is composed of 33 amino acids with a characteristic L-shaped structure. Alternative splicing sites in Chordates could potentially fall within an ANK-repeat. But in order to maintain the reading frame, the activity and the rigid structure of the ANK-repeat, these sites must be consistent in the exact same location in each repeat involved, which could generate selective pressure (functional and structural constraints) on the evolution of the ankyrin exon-intron organization. The exact exon–ANK-repeat correlation might have occurred as the most economic and convenient evolution pattern under such pressure. The unique exon structures of the ANK-repeat region could then promote the diversity of ankyrin isoforms by different combinations of individual ANK-repeats as a result of alternative splicing. Such a mechanism could increase the specificity of ankyrins in mediating sorting and targeting of various membrane transporters by generating novel biological functions and/or complex cell and tissue-specific expression pattern. This idea is currently being explored in our laboratory.

Consistent with this hypothesis is the observation that ankyrin-associated integral membrane proteins such as NCX also undergo considerable tissue-specific alternative splicing in the large cytoplasmic loops speculated to bind ankyrin (Li et al. 1993; Quednau, Nicoll, and Philipson 1997). Interestingly, these alternative splice sites of NCX have been also shown to be more evident in vertebrates than in invertebrates (Marshall et al. 2005). The evolution of complex physiology of vertebrates might require more precise regulation of targeting of diverse membrane proteins including different splice isoforms of the same gene to various specialized membrane domains of the cell. Characterization of the putative ankyrin isoforms with different combination of ankyrin repeats would provide novel insights into novel cellular pathways essential for vertebrate physiology.

Attempt to Identify the Evolutionary Origin of Ankyrins

Previous studies of the spectrin phylogeny suggested that once the number of spectrin repeats became constant, repeat sequences seemed to have evolved independently (Thomas et al. 1997), and therefore, sequences from these fixed repeats might be reliable in building accurate phylogeny trees. The full-length ankyrins all contain exactly the same copies (24) of ANK-repeats in each molecule. The 24 ANK-repeats of ankyrins form four functionally important six repeat subdomains as basic protein interaction sites (Michaely and Bennett 1993). Phylogenetic analysis of these six–ANK-repeat sequences from four representative species was conducted with a protokaryotic six–ANK-repeat
sequence as an outgroup (SwissProt Q8QOU0) (fig. 4 and Figs. S6 and S7 [Supplementary Material online]). Although the position of its clade in the trees was consistent in different phylogenetic analyses, the first six-repeat sequences at the N-terminus of the ANK-repeat domain did not generate reliable grouping pattern within its own clade. This is possibly due to lower conservation of the external protein repeats (Bork 1993), which might have masked the evolutionary history between them. In all these trees, however, subdomain 4 containing last six repeats (repeats 19–24) shows consistent distant evolutionary relationship with other three subdomains. Recent crystal structure of repeats 13–24 of *H. sapiens* ankyrin-R suggests that the last six-repeat subdomain might play a role in fixing the orientation of the SBD (Michaely et al. 2002). Because the spectrin-binding function is an essential component of ankrysins, we propose a novel model for the primordial ankyrin molecule predating the nematode CelANK—the putative primordial ankyrin contains primarily the last six–ANK-repeat subdomain and the SBD.

Subsequently, TBlastN searches were performed to screen the NCBI genome databases of bacteria, archaea, and unicellular eukaryotes using truncated sequences of CelANK and HsaANK-B containing the last six ANK-repeats and the SBD, as well as their full-length sequences. Although our searches confirmed the abundance of ANK-repeat containing sequences in the genome databases (Bork 1993; Mosavi et al. 2004), no significant sequence similarities with the SBD could be identified in these sequences. The identification of bacterial homologs of eukaryotic cytoskeletal proteins has advanced our understanding of bacterial cell biology (Gitaï 2005). To date, however, there is no evidence from genomic database searches to support the presence of ankrysins outside the metazoan phyla. Possibly, ankyrin and spectrin proteins evolved during the transition to the metazoans to form spatially organized cellular domains and to facilitate cell-cell and intracellular communications (Bennett and Baines 2001). But it should be pointed out that, as demonstrated in structural and functional comparison between prokaryotic and eukaryotic cytoskeletal proteins (Amos, van den Ent, and Lowe 2004), the SBD sequence of the primordial ankyrin molecule might be too divergent to allow identification by the current database search model based on pairwise sequence comparisons. This problem could potentially be resolved by combining sequence alignments with comparisons of supposed three-dimensional structure of functional domains (Stevens 2005), which requires determination of the structure of the SBD in the future.

In summary, this work has presented a comprehensive bioinformatics study of the ankyrin gene family, integrating genomic and phylogenetic analyses. Ankyrin genes were duplicated separately in arthropod and vertebrate lineages. Expansion of the vertebrate ankrysins appears to have resulted from segmental genome duplications. A novel model for the putative ankyrin ancestor has been proposed. These findings build a foundation for future studies concerning structural, functional, and evolutionary analyses of ankysins and possibly other cytoskeletal proteins.

**Note Added in Proof**

While this manuscript was under review, Hopitzan, Baines, and Kordeli (2005) reported the gain-of-function evolution of an obscurin/titin-binding–related domain in vertebrate ankrysins that are known to interact with two myofibrillar proteins, obscurin, and titin (Hopitzan et al. 2005). They described a very similar evolution pattern of ankrysins (see their fig. 4) based on the phylogenetic analysis of nucleotide sequences of the ZU-5 domain, which was consistent with our evolutionary analyses of ankysins presented in this manuscript.

**Supplementary Material**

Figures S1–S7 and Table S1 are at available at Molecular Biology and Evolution online (http://www.mbe.oxfordjournals.org/).

**Acknowledgments**

X.C. thanks Vann Bennett (in whose laboratory this work was done) for his support and helpful discussions and Andrew Ravanelli for critical reading of this manuscript. This work was supported by grants from the Howard Hughes Medical Institute and Johnson and Johnson to Vann Bennett.

**Literature Cited**


chimeras in rescue of abnormal inositol 1,4,5-trisphosphate and ryanodine receptor distribution in ankyrin-B (−/−) neonatal cardiomyocytes. J. Biol. Chem. 277:10599–10607.


Claudia Schmidt-Dannert, Associate Editor

Accepted November 2, 2005