J Mol Evol (1999) 49:272–281

JOURNAL OF MOLECULAR EVOLUTION

© Springer-Verlag New York Inc. 1999

The Evolution of the Calpain Family as Reflected in Paralogous Chromosome Regions

Gáspár Jékely, Peter Friedrich

Institute of Enzymology, Biological Research Center, Hungarian Academy of Sciences, H-1518 Budapest, P.O. Box 7, Hungary

Received: 17 November 1998 / Accepted: 30 April 1999

Abstract. Calpains, the Ca^{2+}-dependent intracellular proteinases, are involved in the regulation of distinct cellular pathways including signal transduction and processing, cytoskeleton dynamics, and muscle homeostasis. To investigate the evolutionary origin of diverse calpain subfamilies, a phylogenetic study was carried out. The topology of the calpain phylogenetic tree has shown that some of the gene duplications occurred before the divergence of the protostome and deuterostome lineages. Other gene doublings, leading to vertebrate-specific calpain forms, took place during early chordate evolution and coincided with genome duplications as disclosed by the localization of calpain genes to paralogous chromosome regions in the human genome. On the basis of the phylogenetic tree, the time of gene duplications, and the localization of calpain genes, we propose a model of tandem and chromosome duplications for the evolution of vertebrate-specific calpain forms. The data presented here are consistent with scenarios proposed for the evolution of other multigene families.

Key words: Calpain — Gene duplication — Tetraploidization — Chordate evolution — Paralogous chromosome regions

Introduction

Calpains, the intracellular Ca^{2+}-dependent cysteine proteinases (EC 3.4.22.17), are widely distributed among animal phyla (Sorimachi et al. 1997). As important mediators of intracellular Ca^{2+} signals, they have been postulated to play a role in many physiological processes including secretion (Li et al. 1998), development (Emori and Saigo 1994), memory (Aszódi et al. 1991), and muscle homeostasis (Huang and Forsberg 1998; Richard et al. 1995). The calpain substrates, comprising mostly transcription factors, enzymes, cytoskeletal, membrane-associated, and muscle-specific proteins, undergo regulatory proteolysis in a limited manner resulting in the altered biochemical properties of the substrate (see, e.g., Azarian et al. 1995; Di Lisa et al. 1995; Shevchenko et al. 1998; Watt and Molloy 1993; Yamazaki et al. 1997).

Several members of the calpain family have been cloned and characterized. These include ubiquitously expressed and tissue-specific isoforms, found in vertebrates, and some invertebrate homologues. The ubiquitous forms are CAPN1 and CAPN2 (μ- and *m*-calpain, respectively), which are the best-characterized mammalian isozymes, and chicken μ/*m*-calpain (Sorimachi et al. 1995b). CAPN1 and CAPN2 are heterodimeric molecules consisting of a unique large subunit and a common small subunit (CAPN4). The vertebrate tissue-specific calpains include the muscle-specific CAPN3 (also called p94, nCL-1) (Sorimachi et al. 1989), the stomach-specific CAPN8 (nCL-2) (Sorimachi et al. 1993a), the digestive tract-specific CAPN9 (nCL-4) (Lee et al. 1998), and two atypical family members, CAPN5 and CAPN6 (Dear et al. 1997). While CAPN3, (Kinbara et al. 1998b) CAPN5 and CAPN6 are monomeric, the subunit structure of CAPN8 and CAPN9 is unknown, and their association with the small subunit cannot be excluded (Lee et al. 1998; Sorimachi et al. 1993a). The

Correspondence to: G. Jékely; *e-mail:* jekely@enzim.hu

known invertebrate calpain homologues include sequences from insects (Emori and Saigo 1994; Theopold et al. 1995) and from nematodes (Andresen et al. 1991; Barnes and Hodgkin 1996; Karcz et al. 1991; Wilson et al. 1994). Apart from these, calpain-like sequences showing weak homology to canonical calpains only in the proteinase domain have been found in the mold *Aspergillus nidulans* (Denison et al. 1995), in *Saccharomyces cerevisiae,* and in *Caenorhabditis elegans* and also include the small optic lobes protein (SOL) from *Drosophila* (Delaney et al. 1991) and its human homologue, SOLH (Kamei et al. 1998).

The typical calpain species (including CAPN1, CAPN2, CAPN3, CAPN8, CAPN9, and *Drosophila* and *Schistosoma* calpains) consist of four domains (Ohno et al. 1984): domain 1 has a regulatory function, domain II is the cysteine proteinase domain, the precise function of domain III is unknown, and domain IV is responsible for Ca^{2+} binding and is similar to calmodulin. The Ca^{2+}-binding domain is thought to have been acquired via fusion with a calmodulin-like protein gene (Ohno et al. 1984) and represents a distinct module in the protein. In CAPN3 and one of the *Drosophila* calpains, CALPA, there are unique inserted regions (Sorimachi et al. 1989; Theopold et al. 1995). The small subunit (CAPN4) common to CAPN1 and CAPN2 consists of a Gly-rich N-terminal domain (domain V) and a Ca^{2+}-binding domain (domain VI) which is very similar to domain IV of the large subunit. The atypical forms (CAPN5, CAPN6, *C. elegans* tra-3) comprise domains II and III similar to other family members, a divergent domain I, and a unique domain T, which is apparently not a Ca^{2+}-binding one (Barnes and Hodgkin 1996; Dear et al. 1997).

The aim of this study was to establish the evolutionary relationships of calpain family members, to estimate the time of gene duplications, and to assess the distribution of diverse calpain isoforms in different phyla. To this end, a phylogenetic tree was constructed from the available calpain-like sequences. This methodology together with the analysis of the chromosomal localization of calpain genes shed light on the time of two episodes of major calpain gene duplications coinciding with early chordate genome duplications. These data allowed the reconstruction of the likely stages of calpain evolution, revealing a scenario of gene fusion and truncation as well as tandem versus chromosomal duplications.

Methods

Sequence Alignment

The protein sequences of full-length calpain large and small subunits were obtained from the GenBank (Release 105.0) and SWISS-PROT (Release 35) databases. The accession numbers are as follows: human CAPN1, X04366; chicken CAPN1, AB007775; rat CAPN1, U53858; mouse CAPN1, AF021847; chicken μ/m, X01415; chicken CAPN2, D38026; human CAPN2, M23254; mouse CAPN2, Y10139; rat CAPN2, L09120; chicken CAPN3, D38028; human CAPN3, X85030; mouse CAPN3, X92523; rat CAPN3, J05121; rat CAPN8, D14478; mouse CAPN9, U89513; human CAPN9, AF022799; rat CAPN9, RNU89514; *Drosophila melanogaster* CALPA, Z46891; *D. melanogaster* CALPB, AF062404; *Schistosoma mansoni,* M74233; *S. japonicum,* AF04407; *Caenorhabditis elegans* CPL-1, L25598; human CAPN5, Y10552; mouse CAPN5, Y10656; human CAPN6, AJ000388; mouse CAPN6, Y12582; *C. elegans* tra-3, U12921; human CAPN4, X04016; rabbit CAPN4, M13364; and bovine CAPN4, J05065. Apart from these entries, GenBank contained EST sequences representing a close homologue of the calpain small subunit. An open reading frame from the human and mouse sequences (small subunit II) could be assembled from the overlapping EST clones. The human small subunit II derived from AA234722, AA234789, AA583618, and W20468, and the mouse sequence from AA615207, AA080197, AA498874, and AA500511.

An amino acid alignment of the available sequences was created by the program ClustalW (Thompson et al. 1994) run with the default settings. The resulting alignment was edited manually. Two alignments were created from the main one: (i) one containing domains II and III of large subunit sequences and (ii) one containing the calmodulin-like domain of the small subunits and domain IV of the typical calpains. The N-terminal region of different large subunits cannot be aligned properly, therefore this regioin (whole domain I) is not included in alignment (i). Since domain T of the atypical calpains is not homologous to domain IV of the typical large subunits, this region is also excluded from alignment (i). Very divergent, calpain-like sequences (including *C. elegans* CAA84733, *D. melanogaster* SOL AAB95431, human SOLH AAC33822, *Emericella nidulans* PALB CAA91013, *C. elegans* CAB04898, *C. elegans* AAC71117, *C. elegans* CAB07236, *C. elegans* CAA96686, and *S. cerevisiae* CAA89790) could not be aligned unambiguously and were all shown to be deeply branching in the tree of conventional calpains (data not shown). These sequences were therefore not considered in this study. The alignments are available upon request from the authors.

Tree Construction

Maximum-likelihood trees were constructed with the quartet puzzling method using the Puzzle program package (Strimmer and Haeseler 1996). Maximum-likelihood distance values were calculated using the Dayhoff model of substitution and gamma-distributed rates. The gamma distribution parameter, α, was 0.90 for alignment i and 2.04 for alignment ii as estimated from the data set by Puzzle.

Chromosomal Localization

All mapping data were from the GeneCards database at http://bioinformatics.weizmann.ac.il/cards/ and the gene map of the human genome maintained by the International RH Mapping Consortium at http://www.ncbi.nlm.nih.gov/genemap/ (Deloukas et al. 1998). The localization of human CAPN8 is based on the mapping of the AA026030 and AI362355 EST clones (derived from the putative human orthologue of rat CAPN8) to the interval D1S474–D1S439 on chromosome 1.

Results

Alignment and Phylogenetic Tree of Calpains

The complex domain structure of calpains (Fig. 1) does not allow complete amino acid alignment of all family

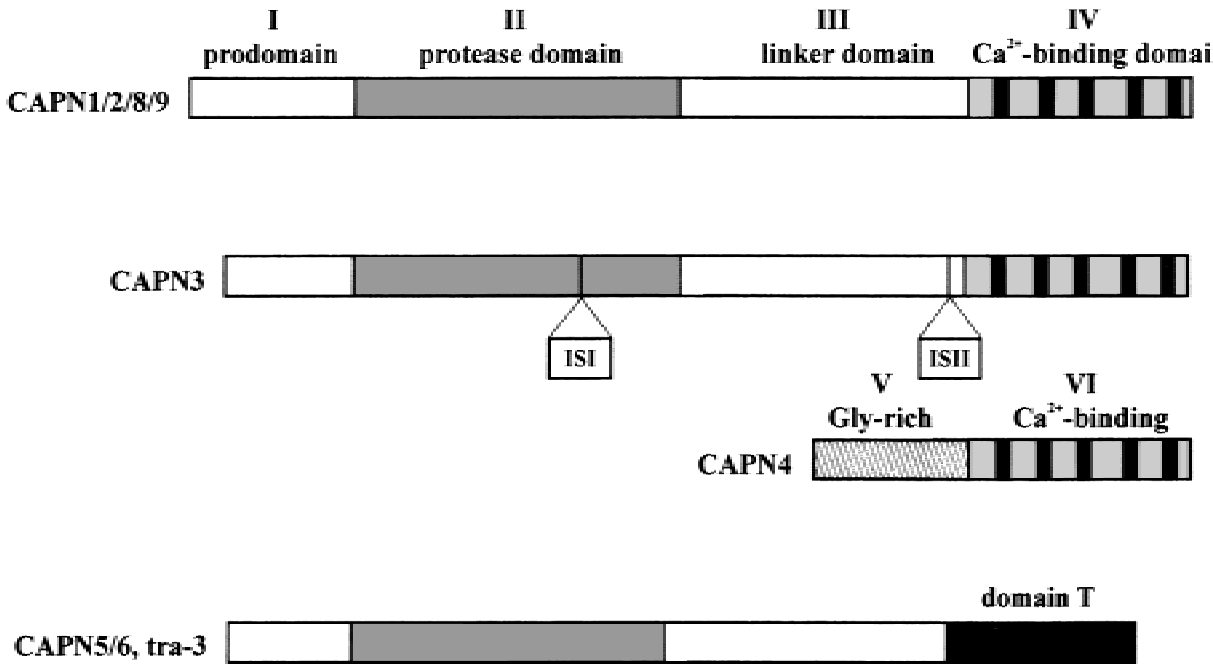


Fig. 1. Domain structure of calpain family members. The regulatory prodomain is the least conserved, while the protease and the linker domain are homologous in all large subunits. CAPN4 shows homology to large subunits only in the Ca^{2+}-binding domain. Putative calcium-binding EF-hand structures are indicated by *black boxes.* ISI and ISII are CAPN3-specific inserted regions. Domain T represents the unique, noncalmodulin-like domain of vertebrate CAPN5, CAPN6, and *C. elegans* tra-3.

members. The N-terminal regions in diverse family members are not homologous and the calmodulin-like domain is missing from CAPN5, CAPN6, *C. elegans* tra-3, and *C. elegans* CPL-1. Moreover, the small subunit shows similarity only to domain IV of typical calpains (Fig. 2). To construct phylogenetic trees from all calpain family members, the nonhomologous regions were removed from the alignment and the calmodulin-like domain was analyzed separately.

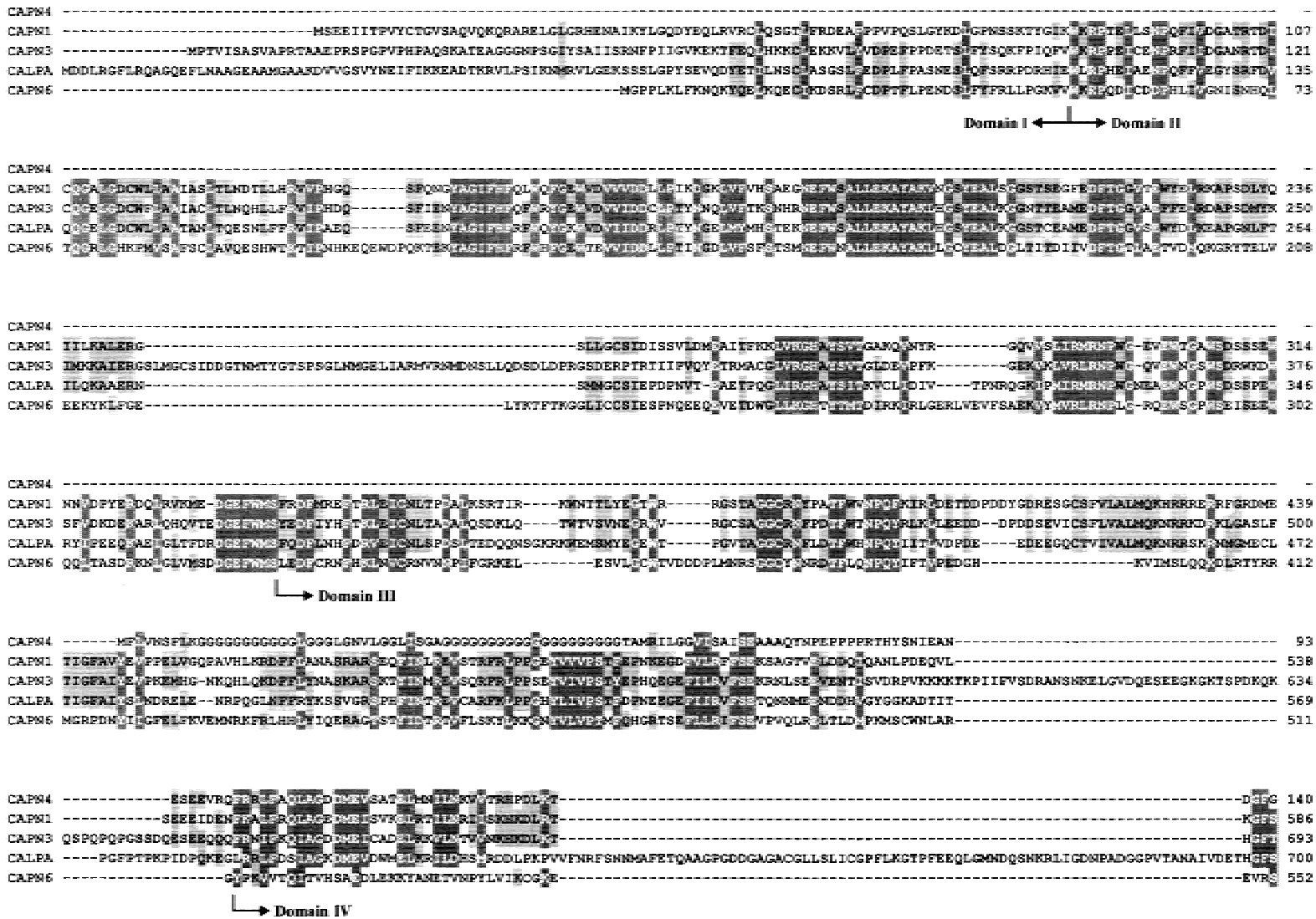
The maximum-likelihood tree (Strimmer and Haeseler 1996) of the calpain family was constructed using alignment i, which contains only domains II and III of large subunit sequences (Fig. 3A). The calpain tree has two major clades, one containing the typical calpains with a calmodulin-like domain and the other containing the atypical ones. *C. elegans* tra-3 is most likely the orthologue of vertebrate CAPN5 and CAPN6, while *Schistosoma* and *Drosophila* calpains, together with *C. elegans* CPL-1, belong to the Ca^{2+}-regulated clade, though the latter has no calmodulin-like domain (Wilson et al. 1994). The grouping of CPL-1 with Ca^{2+}-regulated calpains suggests that this sequence might have lost its calmodulin-like domain secondarily. The relationship of *Schistosoma* and *Drosophila* calpains could not be resolved in the analysis. However, the simplest assumption is that these sequences are orthologous and constitute a separate clade (Aguinaldo et al. 1997). The vertebrate canonical calpains are grouped together with 96% support, and most likely are orthologous to *Schistosoma* and *Drosophila* calpains.

To elucidate the evolutionary origin of the small subunit as well, a maximum-likelihood tree was constructed from alignment ii (Fig. 3B). In this tree the phylogeny of CAPN3, CAPN4, CAPN9, and CAPN1/2/8 could not be resolved. However, the derivation of CAPN4 (small subunit) from CAPN3 is strongly supported by the chromosomal position of the corresponding calpain genes (see below) and also by a separate clade uniting CAPN3 with CAPN4 in neighbor-joining trees with 89% bootstrap probability (not shown).

Evolution of the N-terminal Region of Calpains

The N-terminal end of calpains is very variable (Fig. 2). In spite of the overall similarity in other parts of the sequences, no acceptable alignment can be made in this region between subfamilies (except for CAPN1, CAPN2, and CAPN8). However, orthologous sequences within subfamilies show marked similarities even in this region. It seems that the N-terminal region evolved neutrally within subfamilies, while diverged more rapidly after separation of different subfamilies, i.e., after gene duplications. This suggests that this part of the protein might have had a role, together with the unique inserted se-

Fig. 2. Alignment of some representatives of the calpain family. Domain boundaries are indicated by *arrows.* CAPN4 (X04106), CAPN1 (X04366), CAPN3 (X85030), and CAPN6 (AJ000388) are human sequences, CALPA (Z46891) is a *Drosophila* calpain. Note the differences in the N-terminal regions, the nonhomologous domain T in CAPN6, and the insertion sequences in CAPN3 and CALPA.


CAPN4
CAPN1
CAPN3
CALPA
CAPN6
Domain I
Domain II
Domain III
Domain IV

quences found in certain family members, in functional diversification of calpain isozymes.

Calpain Genes in Paralogous Chromosome Regions

There is accumulating evidence indicating that two phases of extensive gene duplications occurred close to the radiation of vertebrates (Holland et al. 1994). The increase in gene number allowed functional diversification of duplicated genes and the origin of new morphologies and cellular behaviors characteristic of vertebrates. Paralogous chromosome regions with linkage identical to that of similar genes in the human and mouse genome indicate that the large increases in gene number were most likely due to two rounds of genome duplications (Lundin 1993, 1996). The comparison of the number of multigene family members (Holland et al. 1994; Stock et al. 1997) and the estimated total gene number in different chordate groups (Simmen et al. 1998) indicate that the first tetraploidization took place in an early chordate after branching off of the cephalochordate lineage (550–450 MYA) (Doolittle et al. 1996) and the second one after divergence of cyclostomata, but before the divergence of cartilaginous fish from the common vertebrate lineage (450–400 MYA) (Doolittle et al. 1996; Hallböök et al. 1998; Holland et al. 1994; Sharman et al. 1997). A paralogous region comprising a segment of human chromosomes 1, 11, 12, 15, and 19 also includes CAPN1, CAPN2, CAPN3, and CAPN4 (Lundin 1993) and CAPN5, CAPN8, and CAPN9 as revealed by the mapping of these genes (Deloukas et al. 1998) (Table 1). The relation of parts of human chromosomes 1 and 19 with chromosomes 6 and 9 has also been suggested (Kasahara et al. 1997). The different grouping and the imperfect alignment of these chromosome segments comprising, in some cases, parts of both chromosome arms (Table I) can be reconciled assuming chromosomal rearrangements, large regional duplications, and pericentric inversions (Lundin 1993, 1996). That the two rounds of genome duplications creating these paralogous groups occurred in early chordates is evidenced by the phylogeny of some genes in these regions and their cephalochordate or urochordate orthologues. In the trees constructed from actin (Kusakabe et al. 1997), troponin (Hastings 1997; MacLean et al. 1997), and neurotropin (Hallböök et al. 1998) genes, the prochordate sequence is a sister group of the vertebrate paralogues, indicating that tetraploidizations occurred later than the separation of the prochordate and vertebrate lineages but preceded the split of the ancestors of bony fish from the common vertebrate lineage (Hallböök et al. 1998).

The Timing of Calpain Gene Duplications

After vertebrate tra-3-like atypical family members had been discovered, it became clear that the gene duplication separating the tra-3 and the typical calpain clade preceded the divergence of the deuterostome and the protostome lineages (Dear et al. 1997). This duplication must have taken place before the divergence of these two lineages, occurring about 670 MYA (Ayala et al. 1998), but after the plant–animal and fungus–animal split, since plants and fungi lack tra-3 orthologues and Ca^{2+}-regulated calpains (Mewes et al. 1997; Wolfe et al. 1989). The plant–animal split took place about 1000 MYA (Doolittle et al. 1996; Nikoh et al. 1997), therefore the first major calpain gene duplication and later the acquisition of the calmodulin-like domain occurred somewhere between 1000 and 670 MYA.

The divergence of the ancestors of CAPN1 and CAPN2 from the ancestors of CAPN3 and CAPN4 may have coincided with the separation of the corresponding paralogous chromosome regions occurring 550–450 MYA (Doolittle et al. 1996; Lundin 1996), while the divergence of CAPN1 from CAPN2 and CAPN3 from CAPN4 most likely reflects the second tetraploidization 450–400 MYA (Doolittle et al. 1996; Lundin 1996).

The chromosomal localization of calpain family members discloses the relation of regional duplications to large-scale genome changes as well (Fig. 4). The localization of the human CAPN5 gene to the chromosome region 11q12–q14, close to CAPN1 (Matena et al. 1998), can be explained by assuming a tandem duplication before the two tetraploidizations. The first round of genome duplication could have created CAPN6, which was subsequently translocated to Xq22–25 (Matena et al. 1998). Accordingly, both CAPN5 and CAPN6 are orthologous to *C. elegans* tra-3. This duplication also created novel typical calpain genes. The further multiplication of calpain genes through another round of genome doubling was probably followed by a tandem duplication separating CAPN2 and CAPN8, which are still closely linked in the 1q41 region (within 8 cM) (Deloukas et al. 1998). The chromosomal localization of CAPN3 and CAPN4 (Table 1) suggests that the small subunit derived from CAPN3, most likely in parallel with the second round of genome duplication. Therefore the heterodimer nature of calpains is a relatively recent invention absent from invertebrates, as also evidenced by experimental data (Pintér et al. 1992; Theopold et al. 1995). Somewhat more problematic is the origin of CAPN9. Although the clade uniting it with CAPN3 is given 72% support in the calpain tree (Fig. 3A), its chromosomal location to chromosome 1, close to the D1S225 marker (Lee et al. 1998) near CAPN2 and CAPN8 (within 13 cM) (Deloukas et al. 1998), suggests that ancestral CAPN9 may have emerged as a result of a tandem duplication, probably after the first genome duplication (Fig. 4). Considering the low support of the corresponding branching pattern in the tree the linkage of CAPN9 with CAPN2 and CAPN8 can also be explained by assuming that CAPN9

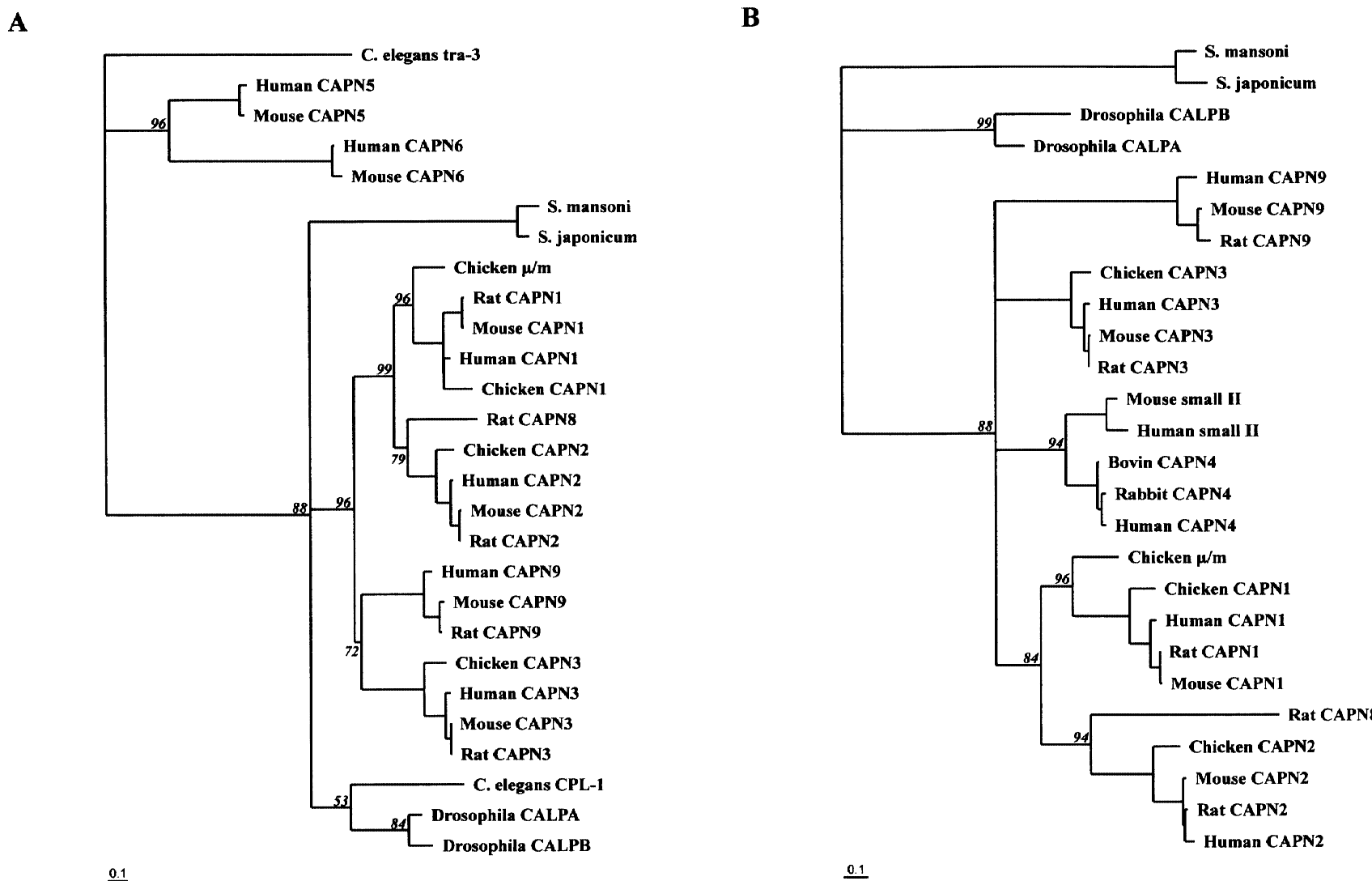


Fig. 3. Phylogenetic tree of calpain large subunits **(A)** and calmodulin-like domains **(B)** inferred by the maximum-likelihood method. Branch lengths are proportional to evolutionary distances. The scale bar indicates an evolutionary distance of 0.1 calculated using the Dayhoff model of substitution and gamma rate heterogeneity. Support for the internal branches is shown as a percentage (support for orthologous branching is not shown, for clarity).

Table 1. Localization of calpain genes to paralogous chromosome regions on human chromosomes 1, 11, 15, and 19[a]

1	11	15	19
NGFB 1p13.1	BDNF 11p13		NTF5 19q13.3
	APLP2 11q23.3		*APLP1 19q13*
APOA2 1q21–q23	APOC3 11q23		APOE 19q13.2
	APOA4 11q23		APOC4 19q13.2
	APOA1 11q23		APOC1 19q13.2
			APOC2 19q13.2
THBS3 1q21		THBS1 15q15	
TPM3 1q22–q23		*TPM1 15q22.1*	*TPM4 19q13.1*
USF1 1q22–q23			*USF2 19q13*
TNNI1 1q31.3	TNNI2 11p15.5		TNNI3 19q13.4
TNNT2 1q32	TNNT3 11p15.5		TNNT1 19q13.4
ESRRG 1q41	*ESRRA 11q12*		
TGFB2 1q41			TGFB1 19q13
CAPN2 1q41	**CAPN1 11q12–q13.1**	**CAPN3 15q15.1–q21**	**CAPN4 19q13**
CAPN8 1q41	***CAPN5 11q12–q14***		
CAPN9 1q42.1–q43			
ACTA1 1q42.1		ACTC 15q14	
RYR2 1q42.1–q43		RYR3 15q14–q15	RYR1 19q13.1
ITPKB 1q41–q43		*ITPKA 15q14–q21*	
	PLCB3 11q13	*PLCB2 15q14–q21*	
	MYO7A 11q13.5	*MYO5A 15q21*	*MYO1D 19p13.3–p13.2*
		MYO1E 15q21–q22	
		MYO1C 15q21–q22	
CHRNB2 1p21	CHRM4 11p12–p11.2	CHRNB4 15q24	
CHRM3 1q41–q44	CHRM1 11q13	CHRNA7 15q14	
		CHRNA5 15q24	
		CHRNA3 15q24	
		CHRM5 15q26	
PEPC 1q42		ANPEP 15q25–q26	PEPD 19cen–q13.11
	CCND1 11q13		*CCNE 19q13.1*
ADPRT 1q42	*ART1 11p15*		
ACTN2 1q42–q43	*ACTN3 11q13–q14*		*ACTN4 19q13*

[a] For a complete list comprising syntenic regions from the mouse genome, see Lundin (1993, 1996). The genes not included in Lundin's original list are in italics. Chromosomal localizations are from the GeneCards database and from (Deloukas et al. 1998). APO, apolipoprotein; ACTA1, actin, α1, skeletal muscle; ACTC, actin, α, cardiac muscle; ACTN, actinin; ADPRT, ADP-ribosyltransferase; ANPEP, alanyl aminopeptidase; ART1, ADP-ribosyltransferase 1; APLP, amyloid β (A4) precursor-like protein; BDNF, brain-derived neurotropic factor; CAPN, calpain; CCND1, cyclin D1; CCNE, cyclin E; CHRM, cholinergic receptor, muscarinic; CHRN, cholinergic receptor, nicotinic; ESRR, estrogen-related receptor; ITPK, inositol 1,4,5-trisphosphate 3-kinase; MYO, unconventional myosin; NGFB, nerve growth factor, β polypeptide; NTF5, neurotropin, 5; PEP, peptidase; PLCB, phospholipase C β; RYR, ryanodine receptor; TGFB, transforming growth factor β; THBS, thrombospondin; TNNI, troponin I; TNNT, troponin T; TPM, tropomyosin; USF, upstream transcription factor.

was translocated to chromosome 1 secondarily, after its separation from ancestral CAPN3.

Discussion

The evolution of the calpain family is consistent with the phylogeny of several multigene families in vertebrates. For example, the analysis of the Hox genes revealed that there is a single Hox cluster in arthropods, nematodes, and amphioxus, probably two or three in lampreys and four in mammals (Holland et al. 1994). Assuming that calpain gene duplications occurred in parallel with Hox cluster duplications, similar ratios are expected for calpain genes in these phyla. Though no cephalochordate or cyclostome calpain has yet been cloned, the number of different calpain genes in arthropods, nematodes, and vertebrates is in agreement with the above assumption. Accordingly, the four vertebrate Hox clusters correspond to vertebrate-specific typical calpains, the typical calpain genes in arthropods and nematodes correspond to the single Hox cluster present in these phyla and the recently identified ParaHox cluster, resulting from an ancient cluster duplication before chordate origins (Brooke et al. 1998), is analogous to the tra-3-like calpains. However, apart from these similarities, resulting from large-scale genome evolution, the two gene families show marked differences in the pattern of tandem and chromosome duplications.

Following gene duplication the newly emerging paralogous descendants may undergo functional differentiation (Ohta 1991). This had multiple sources in the case of duplicated calpain genes. Apart from episodes of gene fusion (acquisition of the calmodulin-like domain)

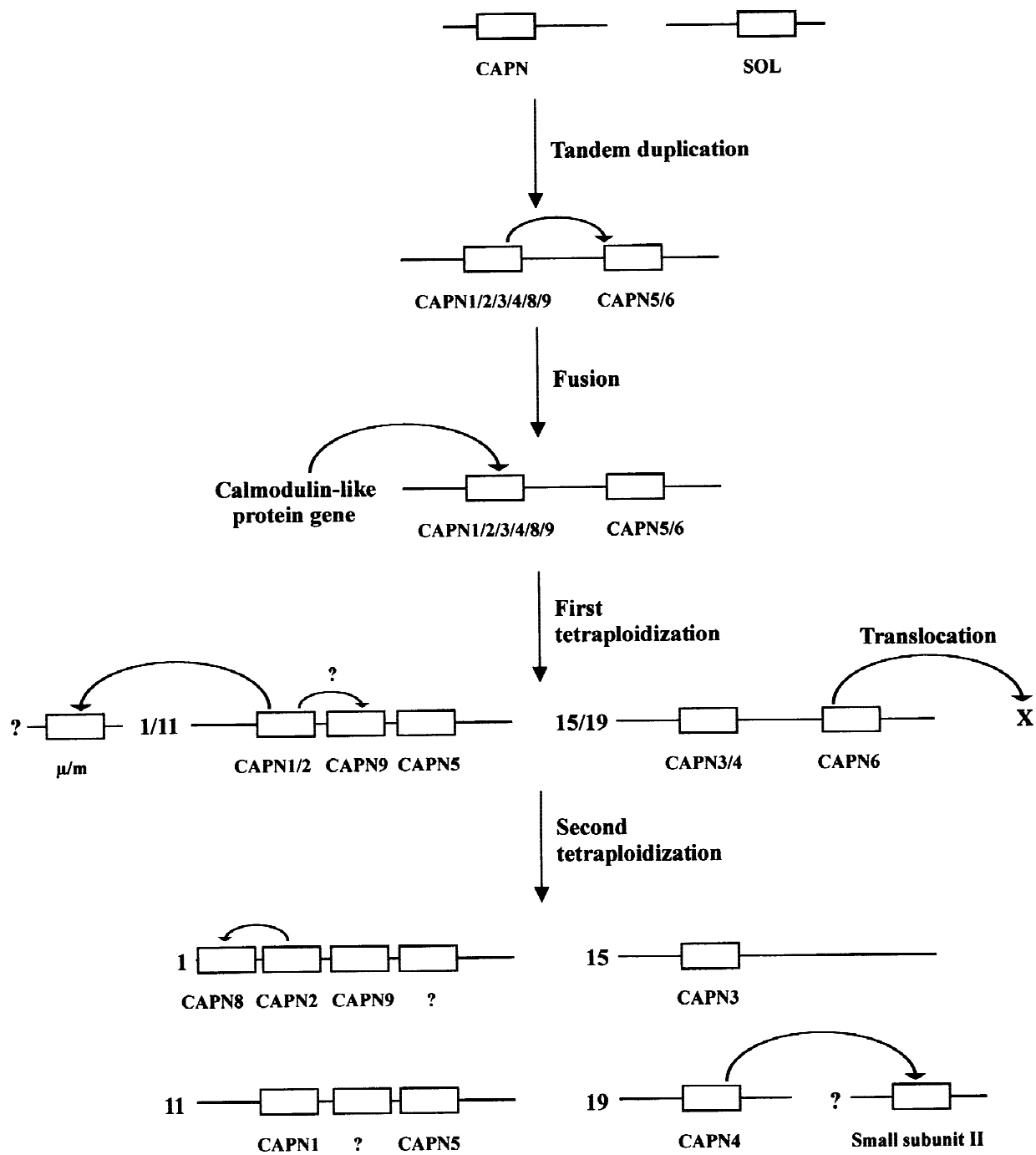


Fig. 4. Calpain gene duplications in relation to tetraploidizations during vertebrate evolution. After the separation of SOL from the ancestral calpain gene, the latter underwent a gene duplication (about 1000–670 MYA). One paralogue fused with a calmodulin-like protein before the divergence of the ancestor of protostomes and deuterostomes and founded the typical, Ca^{2+}-regulated calpain clade. The first genome duplication in an early chordate ancestor (550–450 MYA) was probably followed by the translocation of CAPN6 to the X chromosome and tandem duplications creating CAPN9 and μ/*m*-calpain. The second genome duplication (450–400 MYA) and additional tandem duplications created the now-existing vertebrate calpain isoform genes on human chromosomes 1, 11, 15, and 19. *Question marks* indicate putative calpain genes, uncertain origins, or unknown chromosomal localization.

and partial deletions (truncation of the small subunit and *C. elegans* CPL-1), the most obvious ones were the acquisition of tissue-specific expression patterns (Sorimachi et al. 1994) and the insertion of large sequences carrying new functional units (Sorimachi et al. 1995a; Sorimachi et al. 1993b; Theopold et al. 1995). The occurrence of alternative splicing in certain family members (Sorimachi et al. 1993a; Theopold et al. 1995) and the heterodimeric nature of some vertebrate calpains were also efficient means to generate distinct variants. In the latter respect the existence of a second small subunit in mammals is of particular interest. Though nothing is

known about its function, it might be associated with either CAPN8, CAPN9, or an as yet unknown calpain isoform. The large variation in the N-terminal region could also account for isozyme separation as testified by the differences in the N-terminal self-cleavage and autoregulatory mechanisms of different family members (Baki et al. 1996; Kinbara et al. 1998a; Lee et al. 1999).

When considering the conservation of large paralogous groups, the problem of what forces limit scrambling of ancient gene linkage has to be tackled. Apart from meiotic disturbances, functional constraints (e.g., similar tissue-specific regulation) might account for the selection against breaking up the paralogous regions as suggested by Lundin (Lundin 1993, 1996). It is tempting to speculate that the conserved chromosome regions containing calpain genes may also be regulated in a coordinated manner (e.g., through shared Ca^{2+}-responsive or muscle-specific regulatory elements). This regional chromosome organization might include shared enhancer elements or locus control regions (LCR) (Raguz et al. 1998; Udvardy 1999) as demonstrated for globin (Fraser et al. 1998) and Hox (Gould et al. 1997; Sharpe et al. 1998) genes. To clarify the nature of large-scale chromosome evolution and regulation and the divergent properties of calpain family members, much work is to be done, of which the aspects of calpain origins revealed here are only a minor but requisite step.

Acknowledgments. This work was supported by Grants T17633, T22069, and T29059 from the Hungarian Scientific Research Fund (OTKA). We thank Dr. Peter Tompa and Ms. Éva Bíró for their comments on the manuscript.

References

Aguinaldo AM, Turbeville JM, Linford LS, et al. (1997) Evidence for a clade of nematodes, arthropods and other moulting animals. Nature 387:489–493

Andresen K, Tom TD, Strand M (1991) Characterization of cDNA clones encoding a novel calcium-activated neutral proteinase from *Schistosoma mansoni.* J Biol Chem 266:15085–15090

Aszódi A, Müller U, Friedrich P, Spatz HC (1991) Signal convergence on protein kinase A as a molecular correlate of learning. Proc Natl Acad Sci USA 88:5832–5836

Ayala FJ, Rzhetsky A, Ayala FJ (1998) Origin of the metazoan phyla: Molecular clocks confirm paleontological estimates. Proc Natl Acad Sci USA 95:606–611

Azarian SM, King AJ, Hallett MA, Williams DS (1995) Selective proteolysis of arrestin by calpain. Molecular characteristics and its effect on rhodopsin dephosphorylation. J Biol Chem 270:24375–24384

Baki A, Tompa P, Alexa A, Molnar O, Friedrich P (1996) Autolysis parallels activation of μ-calpain. Biochem J 318:897–901

Barnes TM, Hodgkin J (1996) The *tra-3* sex determination gene of *Caenorhabditis elegans* encodes a member of the calpain regulatory protease family. EMBO J 15:4477–4484

Brooke NM, Garcia-Fernandez J, Holland PW (1998) The *ParaHox* gene cluster is an evolutionary sister of the *Hox* gene cluster. Nature 392:920–922

Dear N, Matena K, Vingron M, Boehm T (1997) A new subfamily of vertebrate calpains lacking a calmodulin-like domain: implications for calpain regulation and evolution. Genomics 45:175–184

Delaney SJ, Hayward DC, Barleben F, Fischbach KF, Miklos GL (1991) Molecular cloning and analysis of small optic lobes, a structural brain gene of *Drosophila melanogaster.* Proc Natl Acad Sci USA 88:7214–7218

Deloukas P, Schuler GD, Gyapay G, et al. (1998) A physical map of 30,000 human genes. Science 282:744–746

Denison SH, Orejas M, Arst HNJ (1995) Signaling of ambient pH in *Aspergillus* involves a cysteine protease. J Biol Chem 270:28519–28522

Di Lisa F, De Tullio R, Salamino F, et al. (1995) Specific degradation of troponin T and I by μ-calpain and its modulation by substrate phosphorylation. Biochem J 308:57–61

Doolittle RF, Feng DF, Tsang S, Cho G, Little E (1996) Determining divergence times of the major kingdoms of living organisms with a protein clock. Science 271:470–477

Emori Y, Saigo K (1994) Calpain localization changes in coordination with actin-related cytoskeletal changes during early embryonic development of *Drosophila.* J Biol Chem 269:25137–25142

Fraser P, Gribnau J, Trimborn T (1998) Mechanisms of developmental regulation in globin loci. Curr Opin Hematol 5:139–144

Gould A, Morrison A, Sproat G, White RA, Krumlauf R (1997) Positive cross-regulation and enhancer sharing: two mechanisms for specifying overlapping *Hox* expression patterns. Genes Dev 11: 900–913

Hallböök F, Lundin LG, Kullander K (1998) *Lampetra fluviatilis* neurotrophin homolog, descendant of a neurotrophin ancestor, discloses the early molecular evolution of neurotrophins in the vertebrate subphylum. J Neurosci 18:8700–8711

Hastings KE (1997) Molecular evolution of the vertebrate troponin I gene family. Cell Struct Funct 22:205–211

Holland PW, Garcia-Fernàndez J, Williams NA, Sidow A (1994) Gene duplications and the origins of vertebrate development. Development Suppl:125–133

Huang J, Forsberg NE (1998) Role of calpain in skeletal-muscle protein degradation. Proc Natl Acad Sci USA 95:12100–12105

Kamei M, Webb GC, Young IG, Campbell HD (1998) *SOLH*, a human homologue of the *Drosophila melanogaster small optic lobes* gene is a member of the calpain and zinc-finger gene families and maps to human chromosome 16p13.3 near *CATM* (cataract with microphthalmia). Genomics 51:197–206

Karcz SR, Podesta RB, Siddiqui AA, Dekaban GA, Strejan GH, Clarke MW (1991) Molecular cloning and sequence analysis of a calcium-activated neutral protease (calpain) from *Schistosoma mansoni.* Mol Biochem Parasitol 49:333–336

Kasahara M, Nakaya J, Satta Y, Takahata N (1997) Chromosomal duplication and the emergence of the adaptive immune system. Trends Genet 13:90–92

Kinbara K, Ishiura S, Tomioka S, et al. (1998a) Purification of native p94, a muscle-specific calpain, and characterization of its autolysis. Biochem J 335:589–596

Kinbara K, Sorimachi H, Ishiura S, Suzuki K (1998b) Skeletal muscle-specific calpain, p94: structure and physiological function. Biochem Pharmacol 56:415–420

Kusakabe T, Araki I, Satoh N, Jeffery WR (1997) Evolution of chordate actin genes: Evidence from genomic organization and amino acid sequences. J Mol Evol 44:289–298

Lee HJ, Sorimachi H, Jeong SY, Ishiura S, Suzuki K (1998) Molecular cloning and characterization of a novel tissue-specific calpain predominantly expressed in the digestive tract. Biol Chem 379:175–183

Lee HJ, Tomioka S, Kinbara K, et al. (1999) Characterization of a human digestive tract-specific calpain, nCL-4, expressed in the baculovirus system. Arch Biochem Biophys 362:22–31

Li HL, Feinstein SI, Liu L, Zimmerman UJ (1998) An antisense oligodeoxyribonucleotide to m-calpain mRNA inhibits secretion from alveolar epithelial type II cells. Cell Signal 10:137–142

Lundin LG (1993) Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. Genomics 16:1–19

Lundin LG (1996) Paralogous genes and metazoan macroevolution. In: New frontiers in theoretical biology. Hadronic Press, Palm Harbor, pp 3–51

MacLean DW, Meedel TH, Hastings KE (1997) Tissue-specific alternative splicing of ascidian troponin I isoforms. Redesign of a protein isoform-generating mechanism during chordate evolution. J Biol Chem 272:32115–32120

Matena K, Boehm T, Dear N (1998) Genomic organization of mouse *Capn5* and *Capn6* genes confirms that they are a distinct calpain subfamily. Genomics 48:117–120

Mewes HW, Albermann K, Bahr M, et al. (1997) Overview of the yeast genome. Nature 387:7–65

Nikoh N, Iwabe N, Kuma K, et al. (1997) An estimate of divergence time of Parazoa and Eumetazoa and that of Cephalochordata and Vertebrata by aldolase and triose phosphate isomerase clocks. J Mol Evol 45:97–106

Ohno S, Emori Y, Imajoh S, Kawasaki H, Kisaragi M, Suzuki K (1984) Evolutionary origin of a calcium-dependent protease by fusion of genes for a thiol protease and a calcium-binding protein? Nature 312:566–570

Ohta T (1991) Multigene families and the evolution of complexity. J Mol Evol 33:34–41

Pintér M, Stierandova A, Friedrich P (1992) Purification and characterization of a Ca^{2+}-activated thiol protease from *Drosophila melanogaster*. Biochemistry 31:8201–8206

Raguz S, Hobbs C, Yague E, Ioannou PA, Walsh FS, Antoniou M (1998) Muscle-specific locus control region activity associated with the human desmin gene. Dev Biol 201:26–42

Richard I, Broux O, Allamand V, et al. (1995) Mutations in the proteolytic enzyme calpain 3 cause limb-girdle muscular dystrophy type 2A. Cell 81:27–40

Sharman AC, Hay-Schmidt A, Holland PW (1997) Cloning and analysis of an HMG gene from the lamprey *Lampetra fluviatilis:* Gene duplication in vertebrate evolution. Gene 184:99–105

Sharpe J, Nonchev S, Gould A, Whiting J, Krumlauf R (1998) Selectivity, sharing and competitive interactions in the regulation of *Hoxb* genes. EMBO J 17:1788–1798

Shevchenko S, Feng W, Varsanyi M, Shoshan-Barmatz V (1998) Identification, characterization and partial purification of a thiol-protease which cleaves specifically the skeletal muscle ryanodine receptor/Ca^{2+} release channel. J Membr Biol 161:33–43

Simmen MW, Leitgeb S, Clark VH, Jones SJ, Bird A (1998) Gene number in an invertebrate chordate, *Ciona intestinalis.* Proc Natl Acad Sci USA 95:4437–4440

Sorimachi H, Imajoh-Ohmi S, Emori Y, et al. (1989) Molecular cloning of a novel mammalian calcium-dependent protease distinct from both *m*- and μ-types. Specific expression of the mRNA in skeletal muscle. J Biol Chem 264:20106–20111

Sorimachi H, Ishiura S, Suzuki K (1993a) A novel tissue-specific calpain species expressed predominantly in the stomach comprises two alternative splicing products with and without Ca^{2+}-binding domain. J Biol Chem 268:19476–19482

Sorimachi H, Toyama-Sorimachi N, Saido TC, et al. (1993b) Muscle-specific calpain, p94, is degraded by autolysis immediately after translation, resulting in disappearance from muscle. J Biol Chem 268:10593–10605

Sorimachi H, Saido TC, Suzuki K (1994) New error of calpain research. Discovery of tissue-specific calpains. FEBS Lett 343:1–5

Sorimachi H, Kinbara K, Kimura S, et al. (1995a) Muscle-specific calpain, p94, responsible for limb girdle muscular dystrophy type 2A, associates with connectin through IS2, a p94-specific sequence. J Biol Chem 270:31158–31162

Sorimachi H, Tsukahara T, Okada-Ban M, Sugita H, Ishiura S, Suzuki K (1995b) Identification of a third ubiquitous calpain species: Chicken muscle expresses four distinct calpains. Biochim Biophys Acta 1261:381–393

Sorimachi H, Ishiura S, Suzuki K (1997) Structure and physiological function of calpains. Biochem J 328:721–732

Stock DW, Quattro JM, Whitt GS, Powers DA (1997) Lactate dehydrogenase (LDH) gene duplication during chordate evolution: the cDNA sequence of the LDH of the tunicate *Styela plicata.* Mol Biol Evol 14:1273–1284

Strimmer K, Haeseler VA (1996) Quartet puzzling: A quartet maximum likelihood method for reconstructing tree topologies. Mol Biol Evol 13:964–969

Theopold U, Pintér M, Daffre S, et al. (1995) *CalpA,* a *Drosophila* calpain homolog specifically expressed in a small set of nerve, midgut, and blood cells. Mol Cell Biol 15:824–834

Thompson JD, Higgins DG, Gibson TJ (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

Udvardy A (1999) Dividing the empire: Boundary chromatin elements delimit the territory of enhancers. EMBO J 18:1–8

Watt F, Molloy PL (1993) Specific cleavage of transcription factors by the thiol protease, m-calpain. Nucleic Acids Res 21:5092–5100

Wilson R, Ainscough R, Anderson K, et al. (1994) 2.2 Mb of contiguous nucleotide sequence from chromosome III of *C. elegans.* Nature 368:32–38

Wolfe FH, Szpacenko A, McGee K, Goll DE (1989) Failure to find Ca^{2+}-dependent proteinase (calpain) activity in a plant species, *Elodea densa.* Life Sci 45:2093–2101

Yamazaki M, Ishidoh K, Suga Y, et al. (1997) Cytoplasmic processing of human profilaggrin by active μ-calpain. Biochem Biophys Res Commun 235:652–656