J. Biochem. 123, 1180-1190 (1998)

Diversity of the Troponin C Genes during Chordate Evolution[1]

Hajime J. Yuasa,* Jos A. Cox,† and Takashi Takagi*,[2]

*Biological Institute, Graduate School of Science, Tohoku University, Sendai, Miyagi 980-8578; and †Department of Biochemistry, University of Geneva, CH-1211 Geneva 4, Switzerland

Received for publication, January 26, 1998

To elucidate the diversity of troponin C (TnC) during chordate evolution, we determined the organization of TnCs from the amphioxus, the lamprey, and the frog. Like the ascidian, the amphioxus possesses a single gene of TnC, and the fundamental gene structure is identical with the ascidian TnC. However, because alternative splicing does not occur in amphioxus, the potential for generation of TnC isoforms through this event arises only in the ascidian lineage. From the frog *Xenopus laevis*, two distinct cDNAs encoding fTnC isoforms and a single s/cTnC cDNA were determined. The duplication of the fTnC gene may be a character of only *Xenopus* or closely related species. The lamprey possesses two cDNAs each encoding fTnC and s/cTnC. The lamprey is the earliest diverged species among vertebrates, and thus it is supposed that the presence of both fTnC and s/cTnC is universal among vertebrate species, and that the gene duplication might have occurred at a vertebrate ancestor after the protochordate/vertebrate divergence. The position of the 4th intron is 3.24/0 in protochordate TnC genes, but at 3.11/2 in vertebrate fTnCs and s/cTnCs. It is suggested that the 4th intron sliding might have occurred prior to the gene duplication.

Key words: amphioxus, evolution, frog, lamprey, troponin C.

Muscle tissue is morphologically classified into two main types: striated muscle and smooth muscle. The contraction trigger of both types of muscle is identical, an increase in intracellular Ca^{2+} concentration, but the regulation systems of contraction are different. In general, striated muscle contraction is controlled by troponin complex, consisting of three protein components, troponin C, I, and T (TnC, TnI, and TnT). On the other hand, smooth muscle contraction is mainly regulated by calmodulin-dependent myosin light chain kinase, which phosphorylates the myosin regulatory light chain. These schemes appear to be universal for vertebrates muscles, as no exception has been observed.

TnC belongs to the EF-hand Ca^{2+} binding protein family and functions as the Ca^{2+} sensor of troponin complex. Two distinct isoforms of TnC, fast skeletal TnC (fTnC) and slow/cardiac TnC (s/cTnC), have been identified in mammalian and avian muscles. The former is expressed only in fast skeletal muscle, and the latter in both slow skeletal and heart muscles. These two TnC isoforms are encoded by independent genes, and the gene structures of these isoforms in human and mouse have been determined (*1-4*). In lower vertebrates, few sequences of TnC have been determined: fTnC from the frog *Rana esculenta* (*5*) and the eel *Anguilla anguilla* (*6*), and s/cTnC from the salmon *Oncorhynchus mykiss* (*7*). Because the two TnC isoforms were isolated from bony fishes, it is supposed that the presence of the fTnC and s/cTnC is general for Osteichthyes. However, the isolation of both TnC isoforms from a single species of fish has not been reported. In addition, no data is available TnC from Chondrichthyes or Agnatha.

The invertebrate chordates (also called protochordates) are composed of two subphyla, Urochordata and Cephalochordata, and are the closest species to vertebrates. The ascidian, often called the sea squirt, is a sessile tunicate belonging to Urochordata, which undergoes a radical metamorphosis during development from a tadpole-like larva to a sessile adult. The ascidian possesses three different types of muscle tissue: monocellular striated muscle of the larval tail (*8*), multinucleate smooth muscle of the adult body wall (*9, 10*), and unicellular striated muscle of the adult heart (*11*). On the other hand, the amphioxus, belonging to Cephalochordata, does not undergo metamorphosis during development. Therefore, throughout life, the amphioxus possesses only one type of muscle, monocellular striated tail muscle (*12*).

The ascidian body wall muscle, although a smooth muscle, contains troponin complex that regulates muscle contraction as in striated muscle (*13*). Recently, we have isolated the cDNAs of two TnC isoforms from the ascidian *Halocynthia roretzi* and determined their genomic structure (*14*). These two isoforms, which are the products of differential RNA processing from a single gene, do not parallel those of higher vertebrates: one is larval TnC, expressed in larval striated muscle, and the other is adult TnC, present in heart muscle and body wall smooth muscle. The intron localization of the ascidian TnC gene is identical

[1] The determined nucleotide sequences have been submitted to the DDBJ under the accession numbers D88976 (*B. lanceolatum* TnC cDNA), D88977 (*B. floridae* TnC cDNA), D88978 (*B. lanceolatum* TnC genome), AB003078 (*X. laevis* fTnCα cDNA), AB003079 (*X. laevis* fTnCβ cDNA), AB003080 (*X. laevis* s/cTnC cDNA), AB008555 (*E. japonicus* fTnC cDNA), and AB008556 (*E. japonicus* s/cTnC cDNA).

[2] To whom correspondence should be addressed. Tel: +81-22-217-6677, Fax: +81-22-263-9206, E-mail: ttakagi@mail.cc.tohoku.ac.jp

to that of the vertebrate s/cTnCs except for the 4th intron. We have also reported the primary structure of amphioxus TnC and its Ca^{2+} binding characterization (*15*). However, it is unknown whether the other TnC isoform is present in amphioxus.

In this study, we attempt to elucidate the diversity of TnC during chordate evolution by comparing the intron localization of TnC genes. First, we determine the TnC cDNAs from two species of amphioxus in different developmental stages, the adult of *Branchiostoma lanceolatum* and the larva of *B. floridae*, and the genomic structure of *B. lanceolatum* TnC. Second, we isolate the cDNAs of TnC isoforms from the lamprey *Entosphenus japonicus* and the frog *Xenopus laevis*. The positions of some introns of these TnC genes are also determined.

MATERIALS AND METHODS

Cloning of Amphioxus Troponin C cDNAs and Genomic DNA—Total RNA of adult *B. lanceolatum* was prepared by the acid guanidium thiocyanate method (*16*), and mRNA was purified with an Oligotex dT-30 Super (Japan Roche). Single-stranded cDNA was synthesized using a First-Strand cDNA Synthesis Kit (Pharmacia). The cDNA library of *B. floridae* was constructed in λZAP II (Stratagene) using mRNA prepared from 2-4-day-old larvae (*12*). The 3′-half of *B. lanceolatum* TnC cDNA was amplified by polymerase chain reaction (PCR) (*17*) using Ex *Taq* DNA polymerase (Takara). The redundant oligomer used for PCR was 5′-CARGARATGATHGARGARGTNGA-3′, where R represents A and G; H, A, C, and T; N, A, C, G, and T. This was originally designed based on the amino acid sequence QEMIEEVD (residue 51-58) of ascidian TnC (*18*). This primer was also useful for amplification of amphioxus TnC, the corresponding sequence of amphioxus being QQMIDEVD (residue 57-64) (*15*). The oligo-dT adaptor 5′-GGGATCCGAATTCT$_{17}$-3′ was used as another primer.

The 5′-upstream stretch of cDNA was determined as follows. The *Eco*RI-ended double-stranded cDNA was synthesized from mRNA using a TimeSaver cDNA Synthesis Kit (Pharmacia). The *Eco*RI Cassette (Takara) was ligated each end of cDNA. The 5′-upstream region was amplified by PCR using cassette-specific primer C1, 5′-GTACATATTGTCGTTAGAACGCG-3′, and R1 (Table I).

TABLE I. **Primers used for amplification of *B. lanceolatum* TnC genomic DNA fragments.**

Primers (positions for cDNA sequence)	Sequence
Sense primers	
F3 (−33 to −14)	TCTGGCTGTCCGTGATAAAG
F4 (35 to 54)	TCAAGGAGGAGCAGATCTCC
F2 (239 to 258)	TGGCCAGGGCCATGCAGGAC
F1 (380 to 400)	ACTTGACAGATGATGAGCTCC
F5 (within intron 5[a])	cggcgagcctgatggtaaag
Antisense primers	
R6 (within intron 2[b])	cacgtgactgtgtcagcccg
R3 (274 to 293)	GCACGTAGCTCATCGTCGGG
R2 (423 to 442)	ACCTCCCGTCCCTGTTTCCG
R7 (475 to 494)	CTACCACCGGACCTTCAGTT
R1 (939 to 962)	ACTGGAGACCAGGTTTATTAAAGG

[a]F5 corresponds to nt 6434 to 6453, within intron 5. [b]R6 corresponds to nt 2962 to 2981, within intron 2.

The *B. floridae* TnC cDNA was also amplified by PCR using a cDNA library as template. Primers used for 3′-half amplification were F4 and T7 primer, 5′-TAATACGACTCACTATAGGG-3′, and those for 5′-half amplification were R2 and T3 primer, 5′-ATTAACCCTCACTAAAGGGA-3′ (Table I).

The *B. lanceolatum* genomic DNA used for PCR amplification was prepared from a crude nuclei fraction, obtained as the precipitate of muscle protein preparation (*15*), by the conventional phenol-chloroform method. Several sets of primers were used to amplify the genomic DNA fragment by PCR. The primers were designed based on the cDNA or genomic sequence of *B. lanceolatum* TnC as listed in Table I. The strategy of amplification of the *B. lanceolatum* TnC gene is shown in Fig. 3. All the amplified products were subcloned to pCR II plasmid vector (TA-cloning kit, Invitrogen) or pUC18 for sequencing. The sequences of products were determined by the dideoxy chain termination method with Dye Primer Cycle Sequencing Kit (Applied Biosystems) using an automated DNA sequencer (Applied Biosystems 373A).

Southern Hybridization—For Southern blot analysis, the genomic DNA was prepared from a single specimen of lyophilized *B. lanceolatum* by the conventional phenol-chloroform method. To remove polysaccharides, the genomic DNA was loaded on a DEAE-cellulose column (1×0.5 cm) which was equilibrated with TE buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA), and the column was washed with 10 volumes of TE buffer. The genomic DNA was eluted with TE buffer containing 1 M NaCl and concentrated by ethanol precipitation. The restriction enzyme-digested genomic DNA was separated on 0.7% agarose gels and transferred to nylon membranes. For the probe DNA, the 527 bp TnC cDNA was labeled with DIG-DNA Labeling Mixture (Boehringer Mannheim) by PCR using primers F3 and R7 (see Table I). Hybridization and washing were carried out according to the manufacturer's instructions

TABLE II. **Primers used for amplification of *Xenopus* and *Entosphenus* TnC cDNAs and genomic DNA fragments.**

Primers (positions for cDNA sequence)[a]	Sequence
For *Xenopus* fTnC*α*	
X f F1 (−21 to −4)	AGCTCTGTGTCCATTGCC
X f F2 (272 to 291)	CGCAGGGAAAAAGTGAAGAG
X f*α* R1 (930 to 949)	GAGTAGGCAGCTCTCTAGAG
X f R2 (77 to 94)	GAGGAAGGACCTCGCATC
X f R3 (379 to 398)	TCATCTGTGATGCTCTCCCC
For *Xenopus* fTnC*β*	
X f*β* R1 (912 to 931)	GAGAGTGATGTGCACATCAG
For *Xenopus* s/cTnC	
X s/c F1 (266 to 285)	GCAAAGGAAAATCAGAAGAA
X s/c R1 (673 to 692)	ACTGTTAACAAAGAATTTCC
X s/c R2 (373 to 392)	TCTTCTGTAATTGTCTCTCC
For *Entosphenus* fTnC	
E f F1 (−20 to −1)	ACAACCCCTTACCAAGTACC
E f F2 (278 to 297)	CGGCCGGCCAGACGGAGGAG
E f R1 (505 to 524)	TCGGTCCACGGTCCACCTGA
E f R2 (385 to 404)	AGGTCCGTGACGTTCTCGCC
For *Entosphenus* s/cTnC	
E s/c F1 (269 to 288)	GCAAGGGGAAGTCAGAAGAG
E s/c R1 (490 to 509)	GGGTATACGGAAGATTCCCA
E s/c R2 (376 to 395)	TCGTCTGTGATGTCCTCCCC

[a]The primers named F are sense (forward) primers, and those named R are antisense (reverse) primers.

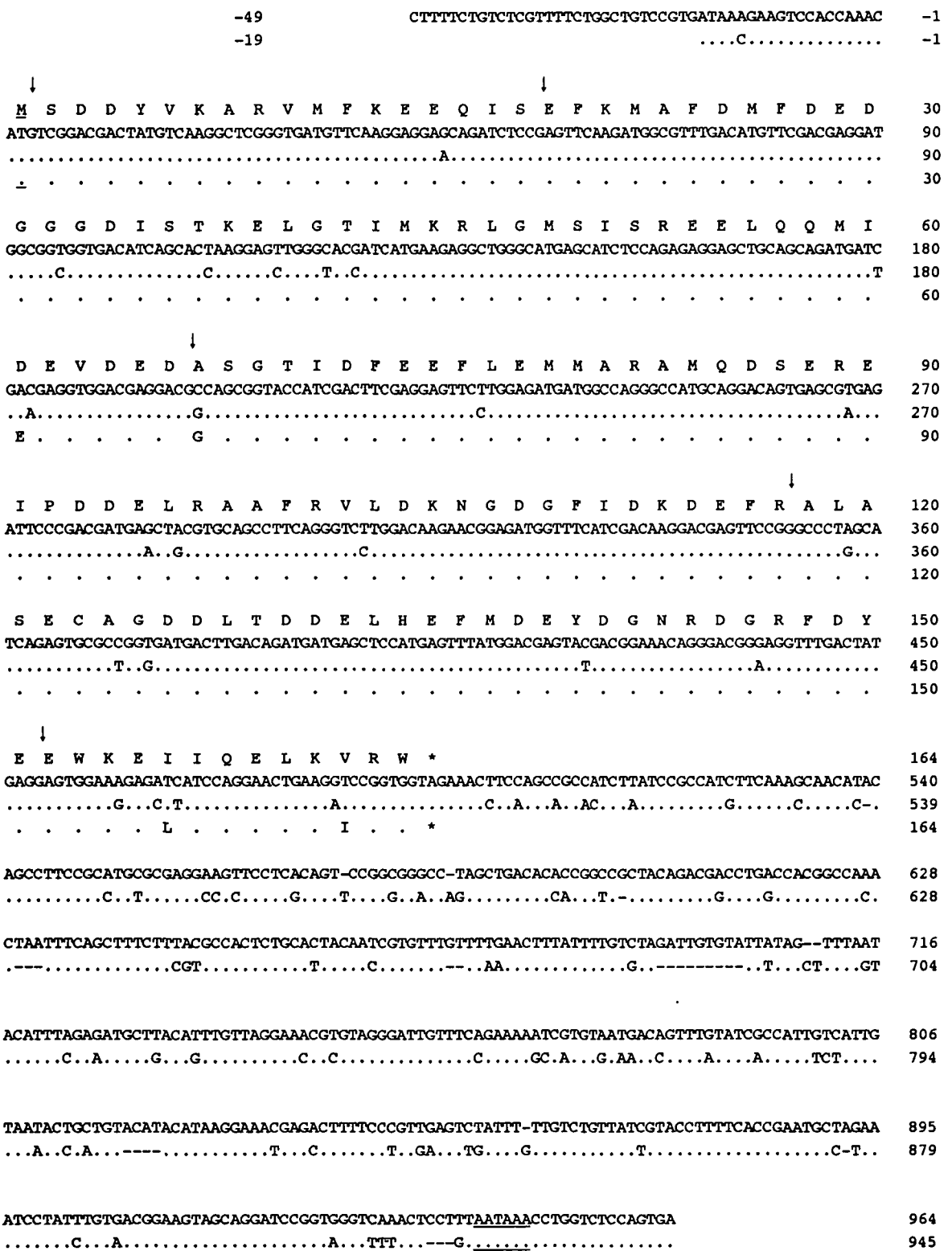

Fig. 1. **Comparison of cDNA and derived amino acid sequences of *B. lanceolatum* and *B. floridae* TnCs.** Upper, *B. lanceolatum* (adult) TnC cDNA and the deduced amino acid sequences; lower, *B. floridae* (larva) TnC cDNA and the deduced amino acid sequence. Identical nucleotides and amino acids to those in *B. lanceolatum* are indicated by dots (.). Gaps are inserted for maximal similarity and shown by bars (–). The typical polyadenylation signal (AATAAA) is underlined. The N-terminal Met (also underlined) is removed after translation. The arrows indicate the positions of introns in *B. lanceolatum* TnC gene.

(Boehringer Mannheim), and the TnC gene was detected with a DIG Luminescent Detection Kit (Boehringer Mannheim).

Isolation of Lamprey TnC and Protein Sequence Determination—The lamprey fTnC was prepared from white muscle, and the sequences of the peptides digested with lysyl endopeptidase were determined as previously described (*19*).

Cloning of Frog and Lamprey Troponin C cDNAs and Partial Genomic DNA—Poly(A)$^+$ RNA was independently purified from the white muscle and heart of the frog *X. laevis* and the lamprey *E. japonicus*, and the single-stranded cDNA was synthesized as described above. The 3′-halves of *Xenopus* and *Entosphenus* TnC isoform cDNAs were amplified by PCR using the oligo-dT adaptor and the redundant oligomer 5′-GARTTYAARGCNGCNTTYGA-3′, designed based on the consensus sequence among the vertebrates TnCs, EFKAAFE (1st-7th residues of site I, the first EF-hand site).

The 5′-upstream regions of *Xenopus* fTnC and s/cTnC cDNAs were amplified as in the case of amphioxus. The primers used were cassette-specific primer C1 and a nonredundant primer, X fα R1 (for *Xenopus* fTnCα cDNA), X fβ R1 (for *Xenopus* fTnCβ cDNA), or X s/c R1 (for *Xenopus* s/cTnC cDNA). The 5′-upstream regions of *Entosphenus* cDNAs were amplified by the 5′-RACE method (*20*). Reverse transcription was primed with E f R1 (for *Entosphenus* fTnC cDNA) or E s/c R1 (for *Entosphenus* s/cTnC cDNA) and the poly-A tail was added to the 3′-end of cDNA with terminal deoxynucleotidyl-transferase (Takara). The PCR amplifications were performed with primer sets of oligo-dT adaptor and E f R2 (for *Entosphenus* fTnC cDNA) or E s/c R2 (for *Entosphenus* s/cTnC cDNA). The primer sequences are listed in Table II.

The genomic DNAs of the frog and the lamprey were prepared from their white muscles by the phenol-chloroform method. The N-terminal coding region of *Xenopus* fTnCα was amplified by PCR with primers X f F1 and X f R2. The site III regions of the frog and the lamprey TnCs genes were also amplified by PCR using primer sets of X f F2 and X f R3 (for *Xenopus* fTnCα gene), X s/c F1 and X s/c R2 (for *Xenopus* s/cTnC gene), E f F2 and E f R2 (for *Entosphenus* fTnC gene), and E s/c F1 and E s/c R2 (for *Entosphenus* s/cTnC gene). The primers sequences and positions for cDNA are listed in Table II.

RESULTS AND DISCUSSION

cDNA and Genomic Structure of TnC from Two Species of Amphioxus—The cDNA of *B. lanceolatum* TnC was amplified by PCR and the complete cDNA sequence of 1,013 nucleotides was constructed from two overlapping fragments (Fig. 1). The open reading frame is composed of 495 nucleotides and encodes a protein of 163 amino acid residues as the initial Met is removed after translation (*15*). Several amino acid differences are observed as compared with previously reported *B. lanceolatum* TnC amino acid sequence (*15*); His replaces Leu at residue 134; Asp replaces Met at residue 138; and Glu replaces Asp at residue 139. These are probably caused by peptide sequencing error, because the cDNA sequence is identical to

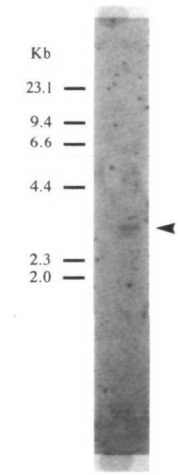

Fig. 2. **Southern blot analysis of the *B. lanceolatum* TnC gene.** Genomic DNA prepared from a single specimen of lyophilized *B. lanceolatum* was digested with *Eco*RI and hybridized to the DIG-labeled TnC cDNA. Left, size markers in kb.

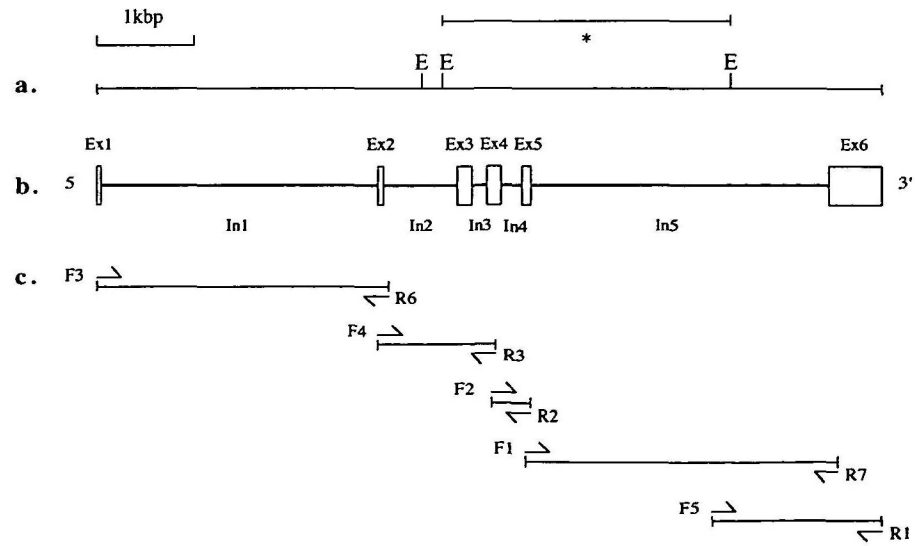

Fig. 3. **Structure of the *B. lanceolatum* TnC gene and strategies used for PCR amplification of DNA fragments.** a: *Eco*RI (E) cleavage sites. The fragment indicated by an asterisk may correspond to the band of 3.0 kbp detected by Southern blot analysis (Fig. 2). b: Exon/intron map of the *B. lanceolatum* TnC gene. Exons (Ex1-Ex6) and introns (In1-In5) are shown in boxes and thin bars, respectively. c: Strategies used for PCR amplification. The primers used for PCR are shown in Table I.

```
TCTGGCTGTCCGTGATAAAGAAGTCCACCAAACATGgtaagtcaggactcttgccatacttgaagtagtcgtctttatctgtggggacgaagttgggaaa 100
gacccggatttagcatagttaggacagcgtgaattgtggccaaggttgtatgtatgccttgtgacagaaagggaacgcgctcctgtttctggtgcttctg 200
tcagaagtcagacaatgccaatgtagacagtgtggtataaacggcgtttctaaacttataatcctaattgagccagcagggtgcgttgggtttcttgtat 300
gacttgcagtttgtagaattaatctgtaaccatagagaaacaaagactacgagtggtcattaatcctcaacatagaggttgagagtagaagcagcagatt 400
ttaccgagagataatttccagggatagtgatgtgatatcgtctcggagatgccacaaatgtcagctccgctattcacggcatggcgcatagcaggcgcca 500
tatttggcactagagcacgacccccttggtcttcggttccatccttttgaacacatgcataactcttaaacggccttaacaagcatcgggacaaaactta 600
cactgcgcggtctcgggctgtgtttcgtgatcagatgtaaaacgacattgtatttactcgggtcacattcccaacgcaatggcgtgatcgacaaatacga 700
ccgaggcaaaagtacatatttcacgacatctagtaacaagatgaggacaagtcaggtcctgtatccagaccggtccattctgattttacatttatcagta 800
tcatcagtaataaaattggagggaattttgcgtcgttgttatatgatacgtatggctgcaatcaaagccccgaacaaagatgctagcaccccatttttcc 900
ctctgctgtaagaatgcctcacatgtcagggacacttttggaatagcttgccattactgacgcaggcgcactacagtcagaccagacaaactgaaaacca 1000
gtcactgaacggatagcaatgactagtaattgcatgatcatcttccactcgatctgttgtttatttcaatatggaaatacgggcaagatgaatatacaag 1100
accgcacatgctaatttcagttcgctacccgtcaacgcactaaaaaaatcttcgtttgtttttcgaaaaaaaatgctgtttcttttaattgtagtttggt 1200
aacaaagcaccgatcatcactgagcacaggtctgattcgctggaacatcaatgatttattttgagttttggctttagaagccgttaccaccgcgaaatgt 1300
taattgggtgttggatctattcataaaggggtcgtttcaattttgtgacctgaaacgttctcaacgtagtcaacctagacagccgtcaggaatcttcacc 1400
cgcgaagaaaagccagaagtgcttcatttctatgaaaagacaaagacattccatttactaaatcgcgtcttctacacacatgaatatcttttcgcattga 1500
tatttatgcgttctgatttatacaaaatttcgcaatgctgaatatggacaaaacgaaaagatcaaggataccattgccggaaaagtcgcagtgcggctgt 1600
gtagcggagcgtaaaaatagtttccttgttgcattctacaagtaaggctgactgcgattggctcagtaaaccgcagctgtcaccgtcagtcagatgtcgg 1700
gacgaggggggggggagatgcacagattgtggtgtcaaacatatcatggataggataggatagataggccatacgggtaggtagctttggctgcatatc 1800
ggaaaggcaacatcttacaaaattatccaaggatattttcttcttcagctttccttcacattatcatcaagatatttgtatatttttgtgtaataatta 1900
ttgaatgtatcatacatcgggcttgttcttcactcagtcatcggggatgttatctgatattgatgacttttctgtaatatcagcactgttaaaatattaa 2000
acagacattcaacataacatgcataacgtccccagtcggaggggggagggggagggacagaaaactttcatacaaattaaagagtaatgtgaaaatcctca 2100
ctcagcagaaactggaccagcttcacttcattgccatgtatagcagatccttcatggcactcactggagtatgtgaaagacatttgtcagttcccctgac 2200
agtgcacttgattacacttatcataagatctagcctatataaatcttggggaactctataaccgtccatcttgggtatgtagaagtcgtttggctgacca 2300
gagtggcgcattgctgtgacagcgtttgttttatcacattttctgaatcagccttgaaggaagcggcatgtcagaggtaggccattgttgactctagagg 2400
atcccaacgtattgcttttgacaatccggtacatgacagagatatttctgctgaccttcagtgccttgggcgtaaccgatgaaaaaaccttttctctaaa 2500
acgccaaactaattatgccgcttcaacggccctgtatgcggacccgtcaggggcagtttgggtggtgatcgtttacttccttctttctggtcagtcttg 2600
tcctcacacacaatgtgtgctgacgtattttggtaaatatcagccggtactgtctcggcttggtgccagagtcatgtacaggcgggctaagtcaaccctg 2700
gaggagtgcagctcgctgtatcagctcgctcctggaggaccgaagtagtccgcatggggcgctgaagaaatctcaactccaggctaggcatgaccagcag 2800
ttggccacatccagctgtgaactttggctattgatgtgtataacaacacttcttctacattgtaaggaaactttaatgactgctgtgactttcctccctc 2900
cagTCGGACGACTATGTCAAGGCTCGGGTGATGTTCAAGGAGGAGCAGATCTCCGgtaagtcgggctgacacagtcacgtgtttcaggtgggaaaattca 3000
aaacatggatgtgcggtgtttttgtaaacaaaataacattgataggagaatattcaacttcccctccatcgtcacctcattaccgtcgtcctcttcataa 3100
tcatcattctcaaccttattataactattcacctcatcattatcctcctcctcccattaccaccaattatcatcctcagcatcattctgatcataatcaa 3200
catcatcattatcatcatcatcatcaccatcatccacatcatcatcatcatccacatcatcacagtacagagttgcactttgttcccatacaaaatgttt 3300
gccatcgaattttcgactgtccttctaccgtctttctcagaattcaaatgaccttggcttcaagtctttcttacactttgtgtcttatatataatgaag 3400
gtcctgtcatggacaagactagttttatcatctttttgaaaaggaaaagtattcaacgttattggttcagctgcggaaaggcttatttaaaatcatttaa 3500
aatgtcctcctttggtgacattcaaggatagaacatctgtgaattctaaacgtcatggcatccataatgtgaattttaatccttcaacaccctatctctg 3600
tattgactaccaggaaaaaaaaatccatttttttctatttgccctccaaccgacaatgagggaatcacatatatcattttccttgccttctaatgaaa 3700
tgtttctaaacagAGTTCAAGATGGCGTTTGACATGTTCGACGAGGATGGCGGTGGTGACATCAGCACTAAGGAGTTGGGCACGATCATGAAGAGGCTGG 3800
GCATGAGCATCTCCAGAGAGGAGCTGCAGCAGATGATCGACGAGGTGGACGAGGACGgtaagtctgtcccaactctcttcagtgggaaatgcaacaccgt 3900
taatctgtgtttggcgtcaagattactacacaggctttagatttagtcctctatagtcaaagtgcttggtatctcagacttgactcactgctgcagtttt 4000
tcgagtctatcgagtatattttctatctgttaacagCCAGCGGTACCATCGACTTCGAGGAGTTCTTGGAGATGATGGCCAGGGCCATGCAGGACAGTGA 4100
GCGTGAGATTCCCGACGATGAGCTACGTGCAGCCTTCAGGGTCTTGGACAAGAACGGAGATGGTTTCATCGACAAGGACGAGTTCCGGgttagtcatggt 4200
taactttatcttacagcatgtttaactgccatccaggccgcgcgattctgaatgtcattttccccgccagctcctaactagagattgtttacacaggcga 4300
gaaaagtgagccttcccatatatttggtgacaccaatgtgcttttggttgatttgattcctccattgatgatccttgtagtcatttgaatacctgtcctc 4400
ccccagGCCCTAGCATCAGAGTGCGCCGGTGATGACTTGACAGATGATGAGCTCCATGAGTTTATGGACGAGTACGACGGAAACAGGGACGGGAGGTTTG 4500
ACTATGAGGgtaagagccatttggcactgacgtcagctcatagcatggcatcatttccgtgtcaaaattgcatacgtaaaaacgttttggggtctatacg 4600
tgacgttttgggggtcgatcaatgcgtcacgtatgccagccctacgccacgtatggcaaaacgtacggatacgtgatgtacatatggttctgcaatttcc 4700
gtgtaaaaattacgtacataaaaaaacgttttgggggtccatacgtcacgtagtccaggctatacgtcacgtatggcatcacatatgtttgccatacgtc 4800
acgtactaaacgtgacgtacatatggtcctggcattttcgggtcgaaatatcgtactatcgaaaacgtcacgttacgtagatacaagggttgtgttaacg 4900
taatgtatgtacgcccttcaagttatcataacgcattacgttaacataacgtagtcatagttaccataacgcattacgttaacgtaacatagtcctttac 5000
gttaacgtaacgcatcacgttaacgtaacgtatgaaattatgcaaatgaggatactgacttacggtttctcgtttgtacttttatttgtgctatatgtca 5100
cataggacgttctagccaggcttcaaaaattttgtccgtactactttgtatccattggtactactttgtatccatggggtcacccattaatattgtgttc 5200
tgcggcctttaggttaataccaagccatacacaacgtaaactggtgacgtaaccagaacaccagaacacgtctattatagacgttcactctctagtgtat 5300
gtggagtttgtaacgtaacatacgtaacacaacgtaacatacgtaacattacgttaacgtggtacatttgatgctttatgtaaacgcgtaggacttgcat 5400
acgtacgtttgactcccaaaacgtcacgtattcaccaaagaacgtgacgttttttacgtatgtaaattttgacacggaaatgctgccatacacttcggat 5500
ggtttcaatgtttttgtcacatatcttggtctaaactgcggcgaagaacgacacgatatcccatgcactatgttgacaatagcgcagcactactagatag 5600
gagaactacgtatactcgatagttgaagccttgtgggtgagattgaagccatattgctgaaccatatccttgtatgattaattcaaacttttcttcaccg 5700
gctttcaccacgaatgtgacgagaatatgttcagtgatttgtaacagggctctggaaaggctttatcgactcataatctgattttaaagaaatgtttttt 5800
tctgttacctcgacataggtggttttattagatctaaataaggcaaagaaaaaataaattttcatatctgaagcgaatctaaacatattcgacagccgta 5900
caaaatgctgaaccttgcgtgactccatctatccatatttgaactgtacactgcaccgtaacgataagttttcttggactttcgttagaacaaaacatca 6000
ttctgaggtctgcccatttcgtaacctaagggccctgtcacacttgtgcgtacattcaagtgcgcatgagatccgcagtaactttggtttaagtaaatt 6100
ttgcggggaggaagttgttttctacttcgacccatcgttgtgcagtttggttatcaaaccaaagaacttacttcttcggcagcaaacttaacctaaacca 6200
aaaacatgctaatacgcaactcatgcgggcttgtgtgatcagcagactaccgtcaaaaactgtttaaactgcgttttctcaccagaccttgtatcgtgca 6300
agtactgttcaggttcattatgaattttcccaatagtcgtgatcaactacgtacaagatgttatggaaatgccggttgaatgtagtttctacacaatggt 6400
gtggttttccgctaatttttggcgacacctaagcggcgagcctgatggtaaagtagtgtgtgcagtctgcaagaattctaggaattgtacaggaatgtgt 6500
```

Fig. 4 (continued on next page)

that of the genome (see below). These amino acid changes do not effect the Ca^{2+}-binding motif.

The cDNA of *B. floridae* TnC is composed of 964 nucleotides and the open reading frame is 495 nucleotides, encoding a protein of 164 amino acid residues including the initial Met, the same number as in *B. lanceolatum* TnC as shown in Fig. 1. The nucleotide sequence of these two cDNAs is 90% identical and, within the coding regions, there are 23 nucleotide substitutions reflecting 4 amino acid differences. These substitutions seems to have no effect on Ca^{2+} binding, suggesting that both TnCs can bind three Ca^{2+} per molecule (*15*). In the case of ascidian, two isoforms of *H. roretzi* TnC, the larval and adult types, are produced by alternative splicing (*14*). However, in amphioxus TnCs, the substitutions are not restricted to a particular region, suggesting that alternative splicing has not occurred. Thus these nucleotide and amino acid substitutions are supposed to result from a species difference, and the same TnC might be expressed in larval and adult amphioxus. These results seem to reflect the fact that ascidian undergoes metamorphosis during development, but amphioxus does not.

Preparation of the genomic DNA was prepared by the phenol-chloroform method produced an unacceptably high level of contamination by polysaccharides, and the genomic DNA was not digested with restriction enzymes. The polysaccharides were removed by passing the preparation through a DEAE-cellulose column, but the amount of recovered genomic DNA was less than 1/10. This made it difficult to perform Southern analysis of *Eco*RI-digested genomic DNA, and the probe hybridized with only a single band (Fig. 2, *ca.* 3.0 kbp). This fragment seems to correspond to nucleotide positions 3541–6478 (Fig. 3a, indicated by the asterisk). According to the restriction enzyme map (Fig. 3), another fragment containing exons 1 and 2 should be detected. However, the length of exons 1 and 2 is relatively short, and the probe might not hybridize with the fragment under our conditions. Though Southern blot analysis of DNA digested with other restriction enzymes was not performed, no other band suggesting the possible existence of another TnC gene was observed. In addition, on each reaction of genomic PCR, only a single product derived from TnC gene is amplified. Thus the TnC gene seems to be present in a single copy in the genome of amphioxus.

Figure 4 shows the nucleotide sequence of the *B. lanceolatum* TnC gene, which was constructed from five overlapping fragments separately amplified by PCR as shown in Fig. 3. The genomic structure shows that it is composed of 8,101 bp and divided in 6 exons by 5 introns. All introns start with gt and end with ag, and according to the nomenclature of Kretsinger and Nakayama (*21*), the intron positions are −17/0, 1.01/1, 2.13/1, 3.24/0, and 4.21/1.[3] There is no sequence discrepancy in the overlapping regions, and the nucleotide sequences of exons are exactly identical with that of cDNA. No exonic sequence is observed within intron 2 and 3, suggesting that alternative splicing does not occur.

cDNAs of TnC Isoforms from the Frog, X. laevis—From the white muscle of the frog *X. laevis*, two distinct cDNAs encoding fTnC were detected. The longer isoform, named fTnCα, is composed of 1,090 nucleotides, and the shorter isoform, named fTnCβ, is composed of 801 nucleotides (Fig. 5a). Both cDNAs encode a protein of 163 amino acid residues, and the difference in length between them lies mainly in the length of the 3′-noncoding regions. The deduced amino acid residues of fTnCα and fTnCβ are identical except that residue Arg-102 of fTnCα is changed to Cys-102 in fTnCβ. This substitution does not effect the Ca^{2+}-binding motif. These isoforms show higher homology with each other than with the fTnC of another species of frog, *R. esculenta* (*5*), suggesting that the gene duplication might have occurred only in *X. laevis* or the ancestor of closely related species to *Xenopus*. In the genus *Xenopus*,

[3] The positions of introns are indicated according to the nomenclature of Kretsinger and Nakayama (*21*). The first number indicates the number of the EF-hand site sequentially numbered from N to C. The second number (following the period) shows the number of the residue within the site, which is generally constructed from 29 residues. The last number (following the slash) is phase: 0 means the intron lies between triplet codons, 1 means between first and second nucleotides of the codon, and 2 means between second and third. For example, 4.21/1 means site IV, 21st residue and phase 1; −17/0 means phase 0, 17 residues before the beginning of site I; 3+01/1 means phase 1, 1 residue beyond site III, within the region between site III and IV.

```
ccacaggaatttcagtccacaggatgataactcaagaatgcctgaacggattgtcttcatatttttaggtgggcaggtctttgtgagacctcaaaatga 6600
ttagactttggccccccctggcaacattctatggcagtgcagcggaacttccgcttttaaaatctcgtgttctgaacgtgctatggtcatgattcttttag 6700
tttattggtagatactcttgggaaggaaaataagtcatgtaagtttttggggccccctagcggctttttggaactgtagaagctgattttgtttcaatt 6800
ttgaaaagaataactaaagaaaggttgacagatcgtcatcatttttggtttatagataacgtaagcaatggtttacataatcatacatcatttttgcaa 6900
atcagtatctcatttgcatgattgatggggaaggtttataaatctgctgcattccattataggactcaaacacgttacatatgtaattgaaaaagagatg 7000
aatatcgatagatatcaatcatgcgaacgaatacctaatctgcataatcaattccatagtggtaaatgacggggatttcattcttgcagcatttggaaga 7100
taagtaaatgtggacattattagacataaatcatgcatatgaaggcctcatttacatagtttatgaggaaatgatacaatagcttttcttgctaagacaa 7200
ggctttcgaactttgaacacgtgttatttagtagagaaggtgatcaactgatatgatttatgcaaatgagattcttatttgcatgtgtgctaagaaaatg 7300
aaaacactaaccgagaccatcgtcgccatggcaacgtccttttacgttgccaatcttgtttgggttgtttttcttgtttctaagtcatcttgtttgcctt 7400
ttatgtgcgcaaggtatcatttttgtgctcaaacaaaccagaggtctgatttggaattagaaggtgtgactgcagtcatgactgccgacatctaacttgc 7500
tgcatgtgtgttcaatacctttcttcctaatccctctacaatccactgtgatatttcagaatgattaatcaccagcacaatgctcctgttacagAGTGGAA 7600
AGAGATCATCCAGGAACTGAAGGTCCGGTGGTAGAAACTTCCAGCCGCCATCTTATCCGCCATCTTCAAAGCAACATACAGCCTTCCGCATGCGCGAGGA 7700
AGTTCCTCACAGTCCGGCGGGCCTAGCTGACACACCGGCCGCTACAGACGACCTGACCACGGCCAAACTAATTTCAGCTTTCTTTACGCCACTCTGCACT 7800
ACAATCGTGTTTGTTTTGAACTTTATTTTGTCTAGATTGTGTATTATAGTTTAATACATTTAGAGATGCTTACATTTGTTAGGAAACGTGTAGGGATTGT 7900
TTCAGAAAAATCGTGTAATGACAGTTTGTATCGCCATTGTCATTGTAATACTGCTGTACATACATAAGGAAACGAGACTTTTCCCGTTGAGTCTATTTTT 8000
GTCTGTTATCGTACCTTTTCACCGAATGCTAGAAATCCTATTTGTGACGGAAGTAGCAGGATCCGGTGGGTCAAACTCCTTTAATAAACCTGGTCTCCAG 8100
T                                                                                                       8101
```

Fig. 4. **Nucleotide sequence of the *B. lanceolatum* TnC gene.** The exons are indicated by capital letters and boxed. The sequences of introns are shown in lower-case letters.

(a)

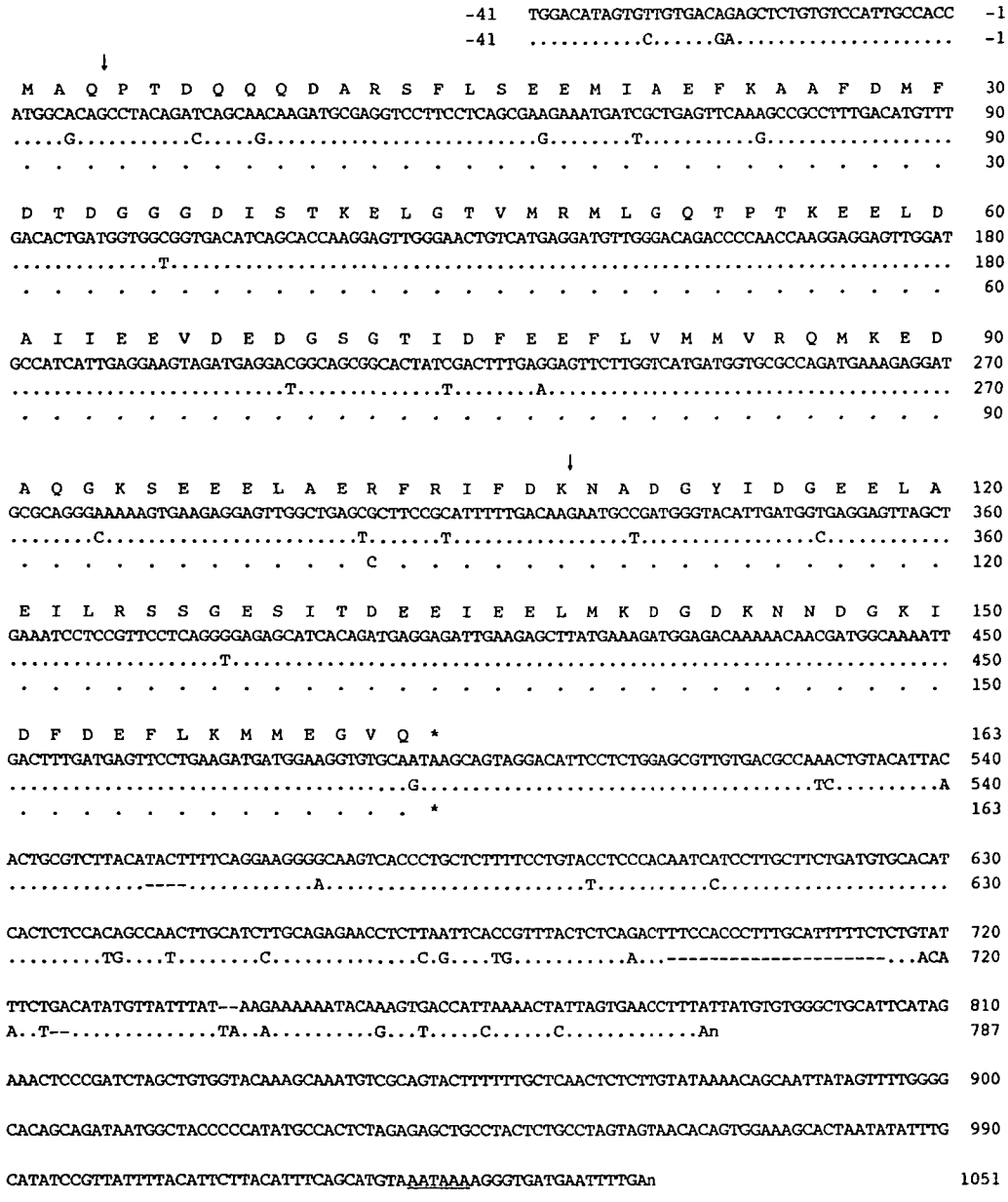

Fig. 5a

polyploid species are commonly observed (*22*), and the presence of two fTnC isoforms may arise from allelic variants of the pseudotetraploid *X. laevis*. The first intron of the fTnCα gene is composed of 439 bp, and inserted at $-18/0$ (6 bp downstream from the initiator ATG, data not shown). The intron inserted within site III coding region is composed of 138 bp, and positioned at 3.11/2 (data not shown), the same position as the mammalian and avian TnC genes.

The *Xenopus* s/cTnC cDNA was isolated from heart muscle and found to consist of 1,528 bp (Fig. 5b). The open reading frame is composed of 486 nucleotides and encodes a protein of 161 amino acid residues. This is the same length as the mammalian and avian s/cTnCs. The other cDNA encoding the s/cTnC isoform was not detected. The intron within the site III coding region is composed of 1,242 bp and inserted at 3.11/2 (data not shown).

cDNAs of TnC Isoforms from the Lamprey E. japonicus—The cDNA of the *E. japonicus* fTnC was isolated from the white muscle and found to consist of 892 nucleotides (Fig. 6a). The open reading frame is composed of 504 nucleotides and encodes a protein of 167 amino acid residues. The lamprey fTnC is four residues longer than the avian and frog fTnCs, and the longest of all known vertebrate TnCs (Fig. 7). Two of four additional amino acid residues are located at the N-terminus, and the other two residues at the C-terminal end. As in the other vertebrate TnC genes, the intron positioned at 3.11/2 (*ca.* 1.8 kbp, data not shown) is also inserted in the *Entosphenus* fTnC gene.

(b)

```
                                                        -22  CTCCTGACCAGCATCACTTACA  -1

ATGGATGATATTTACAAAGCAGCGGTTGAACAACTTACAGAAGAGCAAAAAAAATGAGTTCAGGGCAGCCTTCGACATTTTCGTGCAAGAC   90
 M  D  D  I  Y  K  A  A  V  E  Q  L  T  E  E  Q  K  N  E  F  R  A  A  F  D  I  F  V  Q  D    30

GCTGAAGATGGCTGCATTAGCACCAAGGAACTGGGGAAGGTCATGAGAATGCTGGGGCAGAATCCCACCCCTGAGGAGTTACAGGAAATG  180
 A  E  D  G  C  I  S  T  K  E  L  G  K  V  M  R  M  L  G  Q  N  P  T  P  E  E  L  Q  E  M    60

ATAGATGAAGTGGATGAAGATGGGAGTGGCACGGTAGATTTTGATGAGTTCTTGGTTATGATGGTCCGGTGTATGAAAGACGACAGCAAA  270
 I  D  E  V  D  E  D  G  M  G  T  V  D  F  D  E  F  L  V  M  M  V  R  C  M  K  D  D  S  K    90

                                                 ↓
GGAAAATCAGAAGAAGAATTATCAGACCTTTTCCGAATGTTTGACAAAAATGCAGATGGCTACATTGACTTGGATGAGCTGAAAATGATG  360
 G  K  S  E  E  E  L  S  D  L  F  R  M  F  D  K  N  A  D  G  Y  I  D  F  D  E  L  K  M  M   120

CTTGAAGCTACTGGAGAGACAATTACAGAAGATGACATTGAGGAACTGATGCGAGATGGGGATAAAAACAATGATGGCAGGATTGATTAT  450
 L  E  A  T  G  E  T  I  T  E  D  D  I  E  E  L  M  R  D  G  D  K  N  N  D  G  R  I  D  Y   150

GATGAATTTCTGGAATTCATGAAGGGAGTTGAATAAATGTGAGATGTTAACTTCTGCCGGAGCCATGCCCCCCCCCCACCTGCATAATGAA  540
 D  E  F  L  E  F  M  K  G  V  E  *                                                        161

TTTGGGATTTTGTATTAAATATGAACATAAATGATAAATACAGAATACAACAGGGGTATTGTCCCCTACAGCTGTTATTAAACTACTACT  630
TCCAGGTTAAACATTCCTGGTGTACACATATCGACTGATATAAGCCTCACTTTCTATGGAAATTCTTTGTTAACAGTAAGGGCCAATATG  720
TCAGCCCTTATGCTTTCAAAGGGCATCCTCAGTGAGCCACCAAGGAGAAATACCTGCTCCTCCAGCAGATGATGGCTGATAGCTCACATT  810
GTCCTGTGCACAAGAGTTATACTGTGTACATAGTATTTTGTAGACTGTTTTAAAGTATTAGAGTTGAAGTGTAATGGCTGCATAATCCTG  900
CTAATAAAGCACTGTTATACTCTACTCATCCTCGGCTACCCAAACATAACATCCTTAGTGTACTGAAATTTGAGTTGAGCAACATCTGAT  990
GGGCCTCCAGTTTTATGACACATGGGAGGGACTTATCACTAAATAGGAGTGGTTTATGTGCATAACTGGGTATGGCTTGTTTAGTTCTGG 1080
GTGTGGCCTTGGGTAGTTGGGGTGTAGTCATATCTCCTTAATGTTCACATTTTACAAAGCAGACAGCCTTGGATCCAGGCACAATTTTAA 1170
TTCAGAATAAGACTAAGTATATGTAAAAAGCCTTACTCCTGACATTAACCCATTCATGGCTGAAGAGTTCAGAAGAAGTACAATACTAAA 1260
TGCAAGTGTTACAGTGGTCCGGACAAGTGCCTTCACTCCTTTGTGCTATTTTTAAACATTTTTCTAAAACCGTTTGACATTTTGTGCCCA 1350
AATAACCAGTCCATGGCATAATTCTATTTGCTTTGTGACCTTTTGCTTTTAAATGTAAATCGGTGAAAATGTGCAAAACAACATCATTGG 1440
TGGCATCAAATGCTGTCACACTGTGTCTATTAATTTGCTAATAAACTTGTTTGAAAGTTCTGTGCAn                        1506
```

Fig. 5. **cDNA and derived amino acid sequences of the *X. laevis* TnCs.** a: Upper, *X. laevis* fTnCα cDNA and the deduced amino acid sequences; lower, *X. laevis* fTnCβ cDNA and the deduced amino acid sequence. Identical nucleotides and amino acids to those in fTnCα are indicated by dots (.). Gaps are inserted for maximal similarity and shown by bars (–). The typical polyadenylation signal (AATAAA) is underlined. The arrows indicate the positions of introns. b: *X. laevis* s/cTnC cDNA and the deduced amino acid sequences. The arrow indicates the position of an intron.

The cDNA encoding the lamprey s/cTnC was isolated from the heart muscle and found to consist of 1,346 bp (Fig. 6b). The lamprey s/cTnC is composed of 162 amino acid residues, possessing an additional residue at the N-terminus as compared with the other vertebrate s/cTnCs (Fig. 7). For the lamprey s/cTnC gene, we were unable to amplify the site III coding region.

Evolution of the TnC Genes and Intron Mobility—From the above results, it is fairly certain that the protochordates possess a single TnC gene, while vertebrate species appear to possess two TnC isoforms. The constructed phylogenetic tree (Fig. 8) also suggests that the TnC gene duplication might have occurred in a vertebrate ancestor after the protochordate/vertebrate divergence. On the other hand, the duplication of the 3rd exon and expression of isoforms by alternative splicing might be a feature of the ascidian lineage.

The distributions of introns in the TnC genes of amphioxus, ascidian (*14*), and mammals (*1-4*) are listed in Table III. The positions of introns 2, 3, and 5 (placed at 1.01/1, 2.13/1, and 4.21/1) are identical in all genes. The first introns of amphioxus and mammal fTnCs are inserted at −17/0, and s/cTnCs and ascidian TnC are at −10/0. However, in the case of s/cTnCs, the first intron is located 7 residues downstream of the initiation Met, but in the ascidian TnC, as the N-terminal amino acids are deleted compared to other TnCs (Fig. 7), the position of −10.0 is just after initiation codon ATG. The insertion of the first intron just after initiation codon is a common feature of the TnC superfamily (*23*), such as calmodulins (*24-26*), myosin essential light chains (*27-31*) and Spec (*32, 33*, also see Table IV). Therefore, the position of the first intron of the ascidian TnC is assumed to be identical with the amphioxus and mammal fTnCs. The N-terminal region (before site I) does not directly participate in Ca^{2+} binding, and some insertion/deletion may be permitted provided a flame sift or stop codon does not occur. The mechanism of the sliding of first introns may be junctional sliding (*34*), the reassignment of a single upstream or downstream splice junction.

On the other hand, intron 4 is placed at 3.24/0 in two protochordate TnCs, and at 3.11/2 in vertebrate fTnCs and s/cTnCs. Thus intron 4 of TnC may have originally been placed at 3.24/0 and have slid to 3.11/2 during the evolution from protochordate to vertebrate. The gene duplication might have occurred following this slide.

The intron positions of TnC superfamily genes are listed in Table IV. As mentioned before, the first introns of TnC superfamily genes are generally inserted at just after the initiation codon. The positions of intron 2 (1.01/1) and

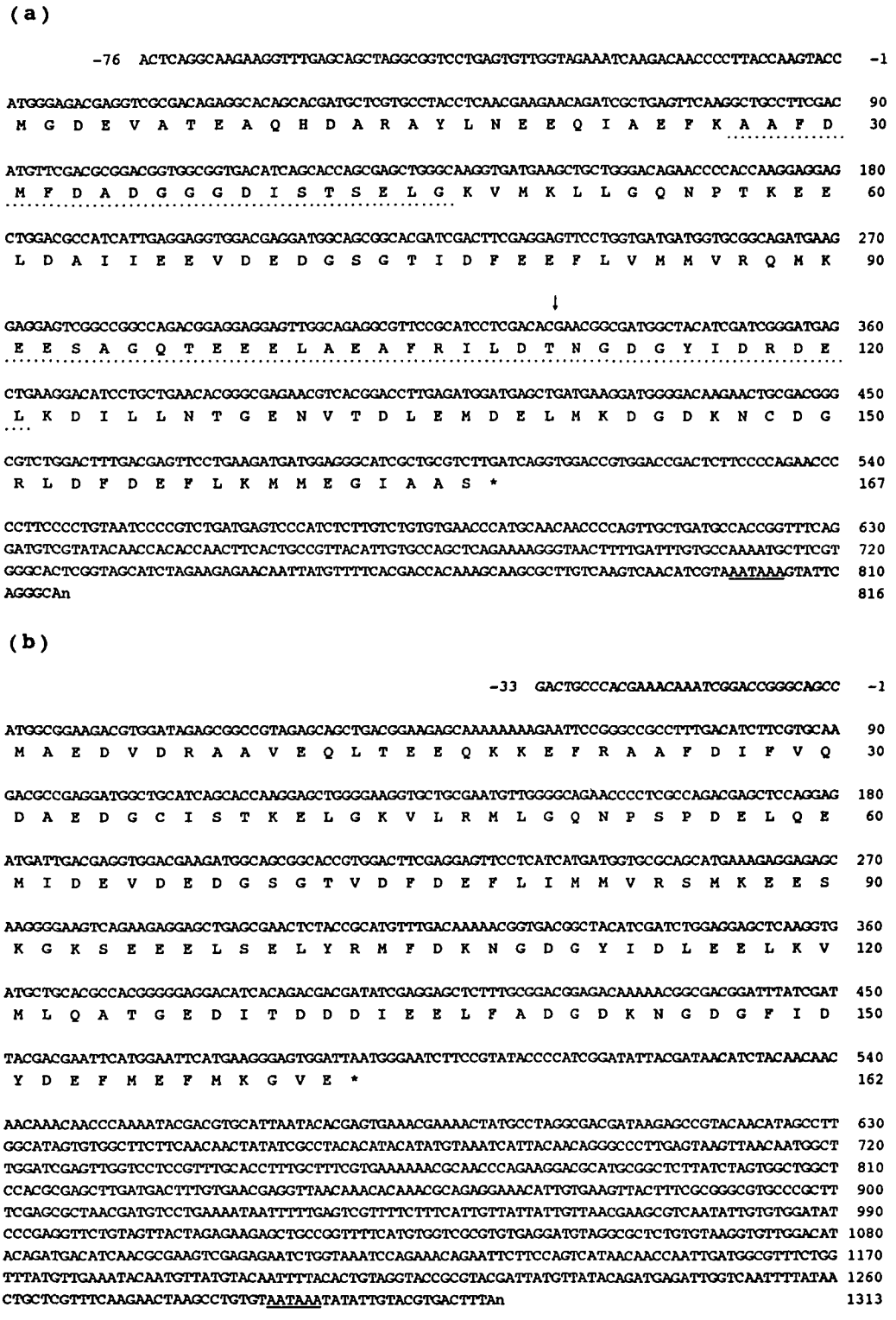

Fig. 6. **cDNA and derived amino acid sequences of the *E. japonicus* TnCs.** a: *E. japonicus* fTnC cDNA and the derived amino acid sequences. The broken-underlined peptides were determined directly by use of an automated protein sequencer. b: *E. japonicus* s/cTnC cDNA and the derived amino acid sequences. The typical polyadenylation signal (AATAAA) is underlined. The arrow indicates the position of an intron.

```
                                            site I                       site II
                                   **********_******************   *****************************
Human    f    -----MTDQQAEARSYLSEEMIAEFKAAFDMFD-ADGGGDISVKELGTVMRMLGQTPTKEELDAIIEEVDEDGSGTIDFEEFLVMMVRQ 83
Xenopus  f    --MAQPTDQQQDARSFLSEEMIAEFKAAFDMFD-TDGGGDISTKELGTVMRMLGQTPTKEELDAIIEEVDEDGSGTIDFEEFLVMMVRQ 86
Lamprey  f    MGDEVATEAQHDARAYLNEEQIAEFKAAFDMFD-ADGGGDISTSELGKVMKLLGQNPTKEELDAIIEEVDEDGSGTIDFEEFLVMMVRQ 88
Human    s/c  -----MDDIYKAAVEQLTEEQKNEFKAAFDIFVLGAEDGCISTKELGKVMRMLGQNPTPEELQEMIDEVDEDGSGTVDFDEFLVMMVRC 84
Xenopus  s/c  -----MDDIYKAAVEQLTEEQKNEFRAAFDIFVQDAEDGCISTKELGKVMRMLGQNPTPEELQEMIDEVDEDGMGTVDFDEFLVMMVRC 84
Lamprey  s/c  ----MAEDVDRAAVEQLTEEQKKEFRAAFDIFVQDAEDGCISTKELGKVLRMLGQNPSPDELQEMIDEVDEDGSGTVDFDEFLIMMVRS 85
Amphioxus     -----MSDDYVKARVMFKEEQISEFKMAFDMFD-EDGGGDISTKELGTIMKRLGMSISREELQQMIDEVDEDASGTIDFEEFLEMMARA 82
Ascidian      ------------MVEHLTEDQKSEFRTCFDIFVEDTEDGTITAKELGKLMKMLGQNPSEQELKEMVEEVDLDGSGTIDFEEFCLMMYRQ 79

                                   site III                         site IV
                           *****************************    *****************************
Human    f    MKEDAKGKSEE----ELAECFRIFDRNADGYIDPEELAEIFRAS-GEHVTDEEIESLMKDGDKNNDGRIDFDEFLKMMEGVQ--- 160
Xenopus  f    MKEDAQGKSEE----ELAERFRIFDKNADGYIDGEELAEILRSS-GESITDEEIEELMKDGDKNNDGKIDFDEFLKMMEGVQ--- 163
Lamprey  f    MKEESAGQTEE----ELAEAFRILDTNGDGYIDRDELKDILLNT-GENVTDLEMDELMKDGDKNCDGRLDFDEFLKMMEGIAAS- 167
Human    s/c  MKDDSKGKSEE----ELSDLFRMFDKNADGYIDLEELKIMLQAT-GETITEDDIEELMKDGDKNNDGRIDYDEFLEFMKGVE--- 161
Xenopus  s/c  MKDDSKGKSEE----ELSDLFRMFDKNADGYIDFDELKMMLEAT-GETITEDDIEELMRDGDKNNDGRIDYDEFLEFMKGVE--- 161
Lamprey  s/c  MKEESKGKSEE----ELSELYRMFDKNGDGYIDLEELKVMLQAT-GEDITDDDIEELFADGDKNGDGFIDYDEFMEFMKGVE--- 162
Amphioxus     MQDSEREIPDD----ELRAAFRVLDKNGDGFIDKDEFRALASECAGDDLTDDELHEFMDEYDGNRDGRFDYEEWKEIIQELKVRW 164
Ascidian      MQAQEEAKIPEREEKELSEAFRLFDLDGNGLIGWDELKAALDGT-GENVETWEVDEMMADGDKNHDSQIDYEEWVTMMKFVQ--- 156
```

Fig. 7. **Alignment of the amino acid sequences of cordate TnCs.** The alignment of amino acids mainly follows the alignment of Takagi *et al.* (*15*). The four EF-hand sites (site I-site IV) are indicated by asterisks (∗). For the ascidian, only the larval TnC was aligned.

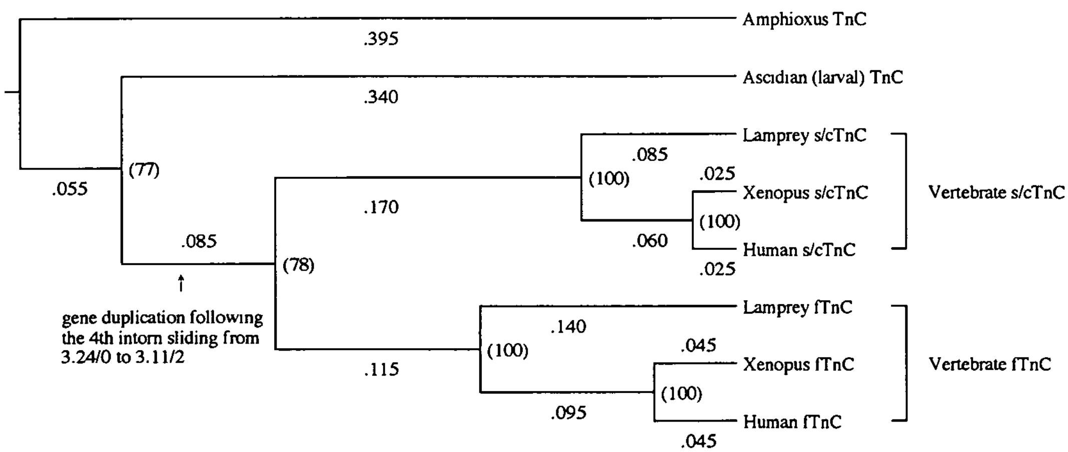

Fig. 8. **A phylogenetic tree of cordate TnCs constructed from the sequences aligned in Fig. 7.** The rooted tree was produced with the PHYLIP package (*35*) and UPGMA method was used. The numbers added to the branches show the length of each branch, and the parenthesized numbers at the forks indicate the percentage of 100 bootstrap resamplings that support these topological elements.

TABLE III. **Localization of introns of amphioxus, ascidian, and mammalian TnC genes.**

Species (type)	Introns position and phases[a]				
Amphioxus	−17/0[b]	**1.01/1**	**2.13/1**	3.24/0	**4.21/1**
Ascidian	−10/0[b]	**1.01/1**	**2.13/1**	3.24/0	**4.21/1**
Human/mouse (fTnC)	−17/0[b]	**1.01/1**	**2.13/1**	3.11/2	**4.21/1**
Human/mouse (s/cTnC)	−10/0[c]	**1.01/1**	**2.13/1**	3.11/2	**4.21/1**

[a]Introns are designated according to Kertsinger and Nakayama (*21*), related to EF-hand domains. Those with identical positions in all genes are shown in **bold**. [b]Just after initiator ATG. [c]21 bp downstream after initiator ATG.

intron 5 (4.21/1) are also highly conserved among members of the TnC superfamily. The intron 3 positions are slightly different between subfamilies, but the phases are identical. However, the positions of intron 4 are not conserved at all between subfamilies. Although the sliding mechanism is unknown, intron 4 of the TnC superfamily might have slid more easily than the other introns.

We wish to thank Dr. L.Z. Holland of Scripps Institution of Oceanography, University of California San Diego, for providing us with the *B. floridae* larval cDNA library.

REFERENCES

1. Parmacek, M.S. and Leiden, J.M. (1989) Structure and expression of the murine slow/cardiac troponin C gene. *J. Biol. Chem.* **264**, 13217-13225
2. Gahlmann, R. and Kedes, L. (1990) Cloning, structural analysis, and expression of the human fast twitch skeletal muscle troponin C gene. *J. Biol. Chem.* **265**, 12520-12528
3. Schreier, T., Kedes, L., and Gahlmann, R. (1990) Cloning,

TABLE IV. **The intron positions of TnC superfamily genes (based largely on Ref. *21*).**

Subfamily species (type)	Positions of introns[a]				
TnC					
Human(f)	−17/0(M)	1.01/1	2.13/1	3.11/2	4.21/1
Human(s/c)	−10/0	1.01/1	2.13/1	3.11/2	4.21/1
Amphioxus	−17/0(M)	1.01/1	2.13/1	3.24/0	4.21/1
Ascidian	−10/0(M)	1.01/1	2.13/1	3.24/0	4.21/1
Calmodulin					
Human(III)	−10/0(M)	1.01/1	2.13/1	3.12/0	4.21/1
Rat(I,III)	−10/0(M)	1.01/1	2.13/1	3.12/0	4.21/1
Rat(II)	−10/0(M)	1.01/1	−	3.12/0	4.21/1
Chicken	−10/0(M)	1.01/1	2.13/1	3.12/0	4.21/1
Myosin essential light chain					
Human	−09/0(M)	1.01/1	2.12/1	3+01/0	4.21/1
Rat(L1)	−09/0(M)	1.01/1	2.12/1	3+01/0	4.21/1
Rat(L4)	−08/0(M)	1.01/1	2.12/1	3+01/0	4.21/1
Chicken(L1)	−09/0(M)	1.01/1	2.12/1	3+01/0	4.21/1
Chicken(L3)	−08/0(M)	1.01/1	2.12/1	3+01/0	4.21/1
Spec					
Sea urchin	−12/0(M)	1.01/1	2.13/1	3.18/2	4.21/1
Parvalbumin					
Human	−	−	2.11/1	3.23/2	4.21/1
Rat	−	−	2.11/1	3.23/2	4.21/1

[a](M) shows the intron is inserted just after the initiation codon, ATG. (−) shows the absence of intron.

structural analysis, and expression of the human slow twitch skeletal muscle/cardiac troponin C gene. *J. Biol. Chem.* **265**, 21247-21253
4. Parmacek, M.S., Bengur, A.R., Vora, A.J., and Leiden, J.M. (1990) The structure and regulation of expression of the murine fast skeletal troponin C gene. *J. Biol. Chem.* **265**, 15970-15976
5. van Eerd, J.-P., Capony, J.-P., Ferraz, C., and Pechere, J.-F. (1978) The amino-acid sequence of troponin C from frog skeletal muscle. *Eur. J. Biochem.* **91**, 231-242
6. Francois, J.-M., Gerday, C., Prendergast, F.G., and Potter, J.D. (1993) Determination of the Ca^{2+} and Mg^{2+} affinity constants of troponin C from eel skeletal muscle and positioning of the single tryptophan in the primary structure. *J. Muscle Res. Cell Motil.* **14**, 585-593
7. Moyes, C.D., Borgford, T., LeBlanc, L., and Tibbits, G.F. (1996) Cloning and expression of salmon cardiac troponin C: titration of the low-affinity Ca^{2+}-binding site using a tryptophan mutant. *Biochemistry* **35**, 11756-11762
8. Satoh, N. (1994) *Developmental Biology of Ascidian*, Cambridge University Press, New York
9. Shinohara, Y. and Konishi, K. (1982) Ultrastructure of the body-wall muscle of the ascidian *Halocynthia roretzi*: Smooth muscle cell with multiple nuclei. *J. Exp. Zool.* **221**, 137-142
10. Terakado, K. and Obinata, T. (1987) Structure of multinucleated smooth muscle cells of the ascidian *Halocynthia roretzi*. *Cell Tissue Res.* **247**, 85-94
11. Kalk, M. (1970) The organization of a tunicate heart. *Tissue Cell* **2**, 99-118
12. Holland, L.Z., Pace, D.A., Blink, M.L., Kene, M., and Holland, N.D. (1995) Sequence and expression of amphioxus alkali myosin light chain (AmphiMLC-alk) throughout development: implications for vertebrate myogenesis. *Dev. Biol.* **171**, 665-676
13. Endo, T. and Obinata, T. (1981) Troponin and its components from ascidian smooth muscle. *J. Biochem.* **89**, 1599-1608
14. Yuasa, H.J., Sato, S., Yamamoto, H., and Takagi, T. (1997) The structure of ascidian, *Halocynthia roretzi* troponin C gene. *J. Biochem.* **121**, 671-676
15. Takagi, T., Petrova, T., Comte, M., Kuster, T., Heizmann, C.W., and Cox, J.A. (1994) Characterization and primary structure of amphioxus troponin C. *Eur. J. Biochem.* **221**, 537-546
16. Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**, 156-159
17. Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-491
18. Takagi, T. and Konishi, K. (1983) Amino acid sequence of troponin C obtained from ascidian (*Halocynthia roretzi*) body wall muscle. *J. Biochem.* **94**, 1753-1760
19. Yuasa, H.J., Sato, S., Yamamoto, H., and Takagi, T. (1997) Primary structure of Troponin I isoforms from the ascidian *Halocynthia roretzi*. *J. Biochem.* **122**, 374-380
20. Frohman, M.A., Dush, M.K., and Martin, G.R. (1988) Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85**, 8998-9002
21. Kretsinger, R.H. and Nakayama, S. (1993) Evolution of EF-hand calcium-modulated proteins. IV. Exon shuffling did not determine the domain compositions of EF-hand proteins. *J. Mol. Evol.* **36**, 477-488
22. Kobel, H.R. and Du Pasquier, L. (1986) Genetics of polyploid *Xenopus*. *Trends Genet.* **2**, 310-315
23. Carpenter, C.D., Bruskin, A.M., Hardin, P.E., Keast, M.J., Anstrom, J., Tyner, A.L., Brandhorst, B.P., and Klein, W.H. (1984) Novel proteins belonging to the troponin C superfamily are encoded by a set of mRNA in sea urchin embryos. *Cell* **36**, 663-671
24. Nojima, H. and Sokabe, H. (1987) Structure of a gene for rat calmodulin. *J. Mol. Biol.* **193**, 439-445
25. Nojima, H. (1989) Structural organization of multiple rat calmodulin genes. *J. Mol. Biol.* **208**, 269-282
26. Koller, M., Schnyder, B., and Strehler, E.E. (1990) Structural organization of the human *CaMIII* calmodulin gene. *Biochim. Biophys. Acta* **1087**, 180-189
27. Nabeshima, Y., Fujii-Kuriyama, Y., Muramatsu, M., and Ogata, K. (1984) Alternative transcription and two modes of splicing result in two myosin light chains from one gene. *Nature* **308**, 333-338
28. Periasamy, M., Strehler, E.E., Garfinkel, L.I., Gubits, R.M., Ruiz-Opazo, N., and Nadal-inard, B. (1984) Fast skeletal muscle myosin light chains 1 and 3 are produced from a single gene by a combined process of differential RNA transcription and splicing. *J. Biol. Chem.* **259**, 13595-13604
29. Robert, B., Daubas, P., Akimenko, M.-A., Cohen, A., Garner, I., Guenet, J.-L., and Buckingham, M. (1984) A single locus in the mouse encodes both myosin light chains 1 and 3, a second locus corresponds to a related pseudogene. *Cell* **39**, 129-140
30. Lenz, S., Lohse, P., Seidel, U., and Arnold, H.-H. (1989) The alkali light chains of human smooth and nonmuscle myosins are encoded by a single gene. *J. Biol. Chem.* **264**, 9009-9015
31. Henderson, S.A., Spencer, M., Sen, A., Kumar, C., Siddiqui, M. A.Q., and Chien, K.R. (1989) Structure, organization, and expression of the rat cardiac myosin light chain-2 gene. *J. Biol. Chem.* **264**, 18142-18148
32. Hardin, S.H., Carpenter, C.D., Hardin, P.E., Bruskin, A.M., and Klein, W.H. (1985) Structure of the Spec1 gene encoding a major calcium-binding protein in the embryonic ectoderm of the sea urchin, *Strongylocentrotus purpuratus*. *J. Mol. Biol.* **186**, 243-255
33. Hardin, P.E., Angerer, L.M., Hardin, S.H., Angerer, R.C., and Klein, W.H. (1988) Spec2 genes of *Strongylocentrotus purpuratus*. *J. Mol. Biol.* **202**, 417-431
34. Craik, C.S., Rutter, W.J., and Fletterick, R. (1983) Splice junctions: association with variation in protein structure. *Science* **220**, 1125-1129
35. Felsenstein, J. (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle