Title: Intron analyses reveal multiple calmodulin copies in Littorina

Authors: R. J. Simpson, C.S. Wilding and J. Grahame.

Present affiliations:

R. J. Simpson, School of Biology, The University of Leeds, Leeds, LS2 9JT.

C. S. Wilding, Genetics Department, Westlakes Research Institute, Westlakes Science and Technology Park, Moor Row, Cumbria, CA24 3JY.

J. Grahame, School of Biology, The University of Leeds, Leeds, LS2 9JT.

Corresponding author:

Bob Simpson

School of Biology

The University of Leeds

Leeds

LS2 9JT

Tel: +44 (0)113 343 2907

Fax: +44 (0)113 343 2835

e-mail: R.J.Simpson@leeds.ac.uk

Running head: Multiple calmodulin genes in Littorina

Abstract

Intron three and the flanking exons of the calmodulin gene have been amplified, cloned and sequenced from 18 members of the gastropod genus Littorina. From the 48 sequences, at least five different gene copies have been identified and their functionality characterized using a strategy based upon the potential protein product predicted from flanking exon data. The functionality analyses suggest that four of the genes code for functional copies of calmodulin. All five copies have been identified across a wide range of littorinid species although not ubiquitously. Using this novel approach based on intron sequences, we have identified an unprecedented number of potential calmodulin copies in Littorina, exceeding that reported for any other invertebrate. This suggests a higher number of, and more ancient, gene duplications than previously detected in a single genus.

Keywords: calmodulin, intron, Littorina, gastropod, microsatellite, minisatellite, EF-Hand

Introduction

Calmodulin is a protein ubiquitous in eukaryotes. It is likely present in all cells with a role of transducing information contained in a pulse of Ca^{2+} ions (Karabinos and Bhattacharya 2000). The calmodulin gene has a highly conserved sequence (Robson 1993). At the amino acid level, calmodulin shows 100% similarity among vertebrates (Yuasa et al. 2001) whilst vertebrate calmodulin differs from most invertebrate calmodulin by only three amino acids. Relatively high levels of homology are also found when Protista (Yazawa et al. 1981) and

fungi (Simão and Gomes 2001) are compared to mammals (e.g. rat (Nojima and Sokabe 1987)) (93% and 91% respectively).

The calmodulin protein consists of four calcium binding domains (EF-hands) produced by ancient gene duplications (Baba et al. 1984; Friedberg and Rhoads 2001), with a gene duplication of a single EF-hand being followed by fusion and then reduplication giving rise to the four domain precursor of all calmodulins (Kawasaki et al. 1998). Additional gene duplications (increasing the number of calmodulin copies) are believed to have occurred on a number of separate occasions since multiple calmodulin copies are found in a variety of taxa. In most vertebrates, calmodulin is encoded by multiple, synonymous, genes. This is an example of the 'multigene-one protein principle', prominent in vertebrates (Fischer et al. 1988; Matsuo et al. 1992). Most vertebrates, including Danio rerio (Friedberg and Rhoads 2002), tortoises (Clemmys japonica) (Shimoda et al. 2002) and mammals, have three copies of the gene encoding calmodulin (CaM I, II and III). The highest number (four) have been found in the teleost fish, medaka (Oryzias latipes) (Matsuo et al. 1992), whereas, the frog, Xenopus laevis contains only two (Chien and Dawid 1984).

In contrast, invertebrates have few, calmodulin gene copies. Only a single copy is found in Drosophila (Smith et al. 1987), the mollusc Aplysia (Swanson et al. 1990) and in Plasmodium falciparum (Robson and Jennings 1991). Two calmodulin genes have been found in the ascidian Halocynthia roretzi (Yuasa et al. 1999) and in the cephalochordate Branchiostoma (Karabinos and Bhattacharya 2000). In contrast to the 100% homology between vertebrate calmodulin copies, some variation is seen when different gene copies from a single invertebrate are compared e.g. the two copies within Branchiostoma differ by

two amino acids. Similarly the two distinct copies found in the echinoderm Arbacia punctulata (Hardy et al. 1988) also differ in amino acid sequence.

The gastropod genus Littorina contains a number of species of great ecological importance in the rocky intertidal. Reid (1996) has thoroughly reviewed the taxonomy and biogeography of this genus, recognizing 19 species. As substantial information on phylogeny of this genus is available, it represents a potentially interesting candidate for the study of calmodulin gene duplication events. We wish to ask the question - are multiple copies of calmodulin found within the genus Littorina? If so, do the numbers of gene copies vary between species? Previous studies of calmodulin have characterized gene copy number using a number of cDNA-based strategies (e.g. Hardy et al. 1988; Matsuo et al. 1992). An alternative strategy is to differentiate gene copies using intron sequences. Utilizing universal primers developed for calmodulin intron three (Côrte Real et al. 1994), we have investigated the number of calmodulin copies in this genus and the variation within these copies.

Materials and methods

Sample collection and DNA extraction

Specimens of Littorina were obtained either fresh or preserved in ethanol: L. littorea (L.) were collected from Filey, North Yorkshire, L. sitkana Philippi and L. kasatka Reid, Zaslavskaya & Sergievsky from Akkeshi, Hokkaido, Japan, L. brevicula (Philippi) from Shirahama, Japan, L. mandshurica (Schrenck) from Alashimi, Hokkaido, Japan. L. subrotundata (Carpenter) from Pachena Point, Vancouver Is., B.C., Canada, L. plena Gould and L. keenae Rosewater from Baker's Beach, San Francisco, California, L. scutulata

Gould from Big Sun, California, L. aleutica Dall from Provideniya, Russia and L. natica Reid from 10 km E of Cape Kamchatskiy, Russia. Within the L. saxatilis species complex, L. saxatilis (Olivi), L. arcana Hannaford Ellis and L. neglecta Bean were collected from Old Peak, North Yorkshire, whilst L. compressa Jeffreys samples were collected from Porth Towyn, Gwynedd, Wales. Additional samples of L. saxatilis and L. compressa were used from Inishmor, Aran Is., Ireland. L. tenebrosa (Montagu) is a rare lagoonal rough periwinkle (Barnes 1993) and samples of this were collected from Golam Head, Ireland. Of the flat periwinkles L. fabalis (Turton) and L. obtusata (Linnaeus) were sampled from Old Peak. Genomic DNA (gDNA) was extracted following the method of Winnepenninckx, Backeljau, and Dewachter (1993).

PCR and sequencing of calmodulin introns

Côrte-Real et al. (1994) described the structure of the calmodulin gene in a variety of taxa, and used conserved exon sequences as primer sites for amplification of the third intron: exon-priming intron-crossing (EPIC) primers. Here the universal CAD2 and CAD3 from Côrte-Real et al. (1994) were used to amplify calmodulin intron three. 50µl PCR reactions using Reddy-Load PCR Buffer (ABgene, Epsom) with 1.5mM $MgCl_2$, 200µM of each dNTP, 25pmol of each primer, 25ng gDNA and 1U Thermoprime Taq (ABgene, Epsom) were subjected to PCR conditions of 1 × 94°C, 5min; 35 × (94°C, 1min, 52°C, 1min, 72°C, 2min); 1 × 72°C, 5min using a Perkin-Elmer 480 cycler. Amplification products were run on a 1% agarose gel for verification. Amplification products were ligated into pGEM Easy-T vector (Promega, Wisconsin) with the resultant plasmids transformed into JM109 cells (Promega, Wisconsin). Plasmids were isolated by alkaline lysis (Ausubel et al. 1992) and the insert size screened by EcoRI (New England Biolabs, Massachusetts) digestion. Plasmids containing

representative examples of each identified insert size were selected for sequencing which was performed using universal primers on an ABI377. All sequences can be found on GenBank, accession numbers AY688305 – AY688352.

Sequence analyses

Sequences were inspected, edited manually and identified by a BLAST search (Altschul et al. 1997) on the short exon regions amplified using the EPIC primers. All regions positively identified as calmodulin were aligned using ClustalX (Thompson et al. 1997) with parameters relaxed until all of the introns began and ended together. An unrooted neighbor joining phylogenetic tree based on p-distance was constructed (with exon sequence removed) in PAUP* (Swofford 1998). It was not possible to justify the use of a particular evolutionary model due to the nature of the aligned sequences therefore a tree based simply upon similarity was constructed.

Recombinant amplicons can be formed during PCR amplification, particularly during amplification of members of multi-gene families (Bradley and Hillis 1997). To address the potential problem of PCR recombinants all sequences were subjected to a Nucleic Acid Dot Matrix Plot (Bowen 1998) analysis which identifies similar sequence fragments found among different amplicons.

Functionality analyses

The putative functionality of any products was investigated using an 'EF-Hand' analysis. When the intron is excised from the total amplified sequence the remaining exon sequence

(35 nucleotide bases from exon three and 35 from exon four) provides the coding information for the production of calmodulin EF-Hand II. The binding of calcium to calmodulin is enabled by four EF-Hands, each of which have a highly conserved amino acid domain ensuring a high level of complementation (figure 1). This exon region can therefore be examined and the putative functionality of the region ascertained. Exon sequences in the EF-Hand domain were aligned against the consensus sequence and amino acids at critical sites identified to assess functionality.

Additional functional domains can be identified within introns; intron sequences may contain regulatory sequences, alternative splice sites, or play a role in maintenance of secondary structure (Prychitko and Moore 1997). Such sequences can be identified through induction of selective constraints resulting in a non-random substitution distribution. This was investigated using the methodology of Prychitko and Moore (1997) in which the number of substitutions was analyzed in windows of 10 bp. If these introns are not subject to selective constraints then the numbers of substitutions will be Poisson distributed. This analysis was conducted on all identified calmodulin copies.

Results

Multiple calmodulin introns are present in Littorina

Figure 2 shows representative amplification products from several of the species examined using the universal primers CAD2 and CAD3. Most members of the genus Littorina share a

common band at approximately 570bp, with other band sizes ranging from 390bp to approximately 900bp. Flanking exon sequences allowed identification of band sequences as calmodulin. All identified introns conformed to the GT-AG rule typical of introns (Breathnach et al. 1978; Mount 1982).

Phylogenetic analysis of all 48 calmodulin intron sequences from 18 members of the *Littorina* genus produced a tree which clustered sequences not by species but by 'gene copy' (labeled LCaM I to V) with high bootstrap values in all cases (Figure 3). A total of 13 different copies were found among the species investigated. Eight of these were found only in a single species and will not be discussed here due to the small sample size. The remaining 5 copies (LCaM I to V) were observed across a variety of species. It must be noted that the nomenclature is based upon order of discovery not upon homology found with other studies e.g. LCaM I is not identical to mammalian CaM I.

Figure 3 highlights the high levels of both inter- and intra-gene copy polymorphism whilst figure 4 indicates the high level of homology among different species for the same copy. This gene copy homology allowed alignment to be achieved using the default ClustalX (Thompson et al. 1997) parameters in almost all cases (see below for exceptions).

No evidence exists to suggest that any identified copies results from PCR recombination. The Nucleic Acid Dot Plot made every pairwise comparison among the sequences highlighting similar regions (results not shown). No similar regions were identified. As a consequence, PCR recombination is not suspected; however, we have focused our analyses and further discussion only on copies identified in multiple species since Bradley and Hillis (1997)

recommend that PCR products should not be assumed to be allelic (i.e. not in vitro recombinants) unless multiple clones or individuals are examined.

Functionality analysis

Comparisons were made across taxa to determine whether base changes found in the Littorina gene copies would cause structural alteration. Figure 5 shows an alignment of the EF-Hand II domain of calmodulin across a wide variety of taxa. The sites labeled (A – L & #) correspond to sites at which the amino acid type is critical (other sites can be any amino acid). Most of the taxa displayed show very high levels of conservation of the amino acid sequence with only a few differences in the highlighted regions. There are three taxa that show alterations within the 12 'required amino acid regions' of EF-Hand domain II (Plasmodium (Robson and Jennings 1991; Robson 1993), trypanosomes (Zimmer et al. 1988) and the sea scallop (Patinopecten) (Toda et al. 1981)), however, all are known to be functional copies.

Figure 6 shows the potential functional alterations within the different Littorina gene copies based upon the variation found within the different taxa discussed above.

All LCaM I (9 species examined) and II (15 species examined) exon sequences agree with the consensus sequence and are thus believed to be functional calmodulin copies. In the five species sequenced for LCaM III, there are four different amino acid sequences. Two amino acid sequences, belonging to L. arcana, L. littorea and L. brevicula, contain a single amino acid change compared to the consensus but it is found in the flanking region at a non-specific site and, as such, probably has no effect upon the final conformation of the protein product. The remaining two species for LCaM III (L. aleutica and L. fabalis) both contain insertions

which substantially alter the amino acid sequence in key areas of the region (L. fabalis has a STOP codon). The high level of intron homology for LCaM III at the DNA level, the high level of amino acid variation relative to other gene copies and the fact that LCaM III would have had to have lost function in two separate lineages (according to the consensus tree of Reid (1996)) strongly suggests that all of the LCaM III copies are non-functional. The LCaM IV dataset contains two different amino acid sequences (one belonging to L. arcana, and one belonging to all of the five other species studied). The substitution found in L. arcana (a threonine in place of the consensus methionine) is in a 'conserved' region of the protein and is a large protein change (hydrophobic to hydrophilic). This is the only amino acid change in any of the six species examined and could be a result of sequencing error. For this reason LCaM IV is also considered to be putatively functional, although further investigation of the indicated amino acid change is required. The two species from which sequences classed as LCaM V were obtained (L. saxatilis and L. arcana) show identical amino acid sequence which both contain three (functionally unimportant) amino acid differences to the consensus. The variation present consists entirely of substitutions and single base insertions and deletions. The distribution of these mutations occurring in each gene copy was investigated following the method of Prychitko and Moore (2000). The DNA sequence (with columns containing indels removed) was split into blocks of 10 bases and studied to determine whether the number of substitutions occurring in each block followed a Poisson distribution. For unconstrained (neutral) sequences, the number of substitutions occurring among blocks should be Poisson distributed (Prychitko and Moore 2000). The results were examined by a Chi-squared test. This test was carried out only on gene copies where there were sequences for at least five species (see Table 1).

Whilst the test shown in Table 1 is not statistically strong (as with Prychitko and Moore (2000)) the known conserved regions at the start and end of the introns (GT & AG) were not picked up by the test). It has determined that the only copy in which the distribution of mutations is not random is LCaM I. Thoughts that the microsatellite present was disrupting the results due to the different mutational mechanisms undergone by this region proved unfounded (the probability of non-random mutation distribution within LCaM I was still high after removing the microsatellite sequence from the analysis). To ensure the robustness of this result the Yates correction factor was applied to the 'LCaM I without microsatellite' result. The result was still significant ($p = 0.017$).

Sequence structure

The gross sequence structure differs among gene copies. LCaM I contains a complex microsatellite region with a variety of dinucleotide and tetranucleotide repeat motifs responsible for most of the observed length variation (ranging from 637 to 745 nucleotides). LCaM II is approximately 450 bp in length and exhibits high inter-copy variability with both substitutions and indels evident. A great deal of allelic variation is found for this intron with five alleles found within different populations of L. saxatilis (the highest level for any of the species studied). Inter-copy variability in LCaM III is also high and characterized by substitutions and small insertions and deletions. LCaM IV was aligned using the default parameters in ClustalX with alterations made by hand to conserve the minisatellite structure that is evident in this species. Three minisatellites are present, shared among the eight species examined with lengths 47, 34 and 12 bp. Also present in this gene copy is a tetranucleotide repeat (AGTC) that, with a few mismatches, contains up to 40 repeats. LCaM V was

amplified from only two closely related species (L. arcana and L. saxatilis) and unsurprisingly displays very low levels of variation (D_{xy} of 0.021).

Discussion

The universal primers based on Côrte-Real et al (1994) yielded amplicons representing at least five, and up to 13, different gene copies when applied to species of Littorina. We have concentrated our analyses on only 5 copies for which we have sequences from multiple species. Our contiguous exon sequence data indicate that four of these represent intron 3 of functional calmodulin genes and the other is found in a non-functional calmodulin-like gene. This suggests that calmodulin is multi-copy in Littorina. Whilst we recognize that our cloning strategy is not guaranteed to detect every intron copy in each species examined e.g. the number of gene copies (1-5) varied between species, we have focused our analyses only on copies identified in multiple Littorina species.

Functionality of these copies was determined through EF-Hand analysis which indicated that four of the five studied copies are functional. This technique is likely to be conservative in its determination of functionality. For example the search term ‘calmodulin’ in the Caenorhabditis elegans database (WormBase (Stein et al. 2001)) finds nine proteins, six of which are classified ‘calmodulin’ and three classed as ‘calmodulin-like’ An EF-Hand analysis found that two of these sequences code for a correct EF-Hand II domain (Accession numbers AF016429 (the only sequence with any positive expressed sequence tags (ESTs)) and Z70034). Of the remaining seven sequences, three (Accession numbers L14433, U70851 and

Z81053) contain the ‘incorrect’ amino acid at three positions whilst the remaining four sequences (accession numbers U28731, Z82282, U28740 and Z77653) contain one, one, one, and two ‘incorrect’ amino acids, respectively. Examining this short region suggests that C. elegans contains two calmodulin genes, four calmodulin-like genes and three pseudogenes (the latter classed as such due to the extremely high level of variation observed). Again EF-Hand I provided a more conservative assessment of likely functionality, the only C. elegans sequence classified as calmodulin was accession number AF016429, all the rest were either classed as calmodulin-like or pseudogenes.

The EF-Hand analysis is not the definitive answer for the number of functional copies present. The insertion of a single nucleotide upstream would alter the reading frame, thus disrupting the final product, but would go unnoticed by the EF-Hand analysis. What does hold true is that the level of gene duplication that occurred establishing this high number of copies (either functional or non-functional) is considerably higher then previously reported in invertebrates. The similar findings from the C. elegans database suggest that Littorina are not unique in this respect.

Alignments of intron sequences can potentially present problems since many of the assumptions underlying alignment programs for coding regions are not applicable. It is difficult to be certain about homology in the absence of information on function. Nevertheless, in this study introns from the same gene copies among species which diverged up to 70MYA could be aligned without difficulty. Alignment of highly diverged intron sequences has also proved possible in a similar study of β-fibrinogen intron 7 between bird orders that diverged between 55 and 90 MYA (Prychitko and Moore 2003). This suggests that, in some cases, intron alignment and subsequent analyses may not be as large an obstacle

as previously assumed, thereby making these highly variable regions a more useful tool for examining evolutionary processes than hitherto thought.

Using introns as a basis of study has elucidated levels of gene duplication that other methods of analysis would have missed e.g. cDNA analysis of this region would have found only two functional copies (LCaM I, II & IV all share identical amino acid sequence whereas LCaM V differs by one amino acid). Our observations increase the number of copies reported from almost all previously studied taxa. Most studies have determined calmodulin copy number by cDNA sequencing alone. In the few studies that have looked at introns (e.g. Ye and Berchtold 1997) they have been used purely to describe gene structure without investigating the possibility of different intron (and hence gene) sequences.

Intron analysis has highlighted higher levels of variation in invertebrates than previously suggested, indicating that the level of gene duplication within the calmodulin gene family is considerably greater, and may have occurred earlier, than previously assumed.

Acknowledgements

This research was supported by the MAST-3 program of the European Commission under contract number MAS3-CT95-0042 (AMBIOS) and a BBSRC studentship. We wish to thank Dr D. G. Reid and Dr E. G. Boulding for provision of many of the specimens. Mr G. Campbell is acknowledged for laboratory assistance.

Literature Cited

Altschul SF, Madden TL, Schaffer AA, Zhang JH, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402

Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K (1992) Current Protocols in Molecular Biology. John Wiley and Sons, New York.

Baba ML, Goodman M, Bergercohn J, Demaille JG, Matsuda G (1984) The early adaptive evolution of calmodulin. Mol. Biol. Evol. 1:442-455

Barnes RSK (1993) On the nature of the coastal lagoon winkles attributed to Littorina tenebrosa and Littorina saxatilis. Cah. Biol. Mar. 34:477-495

Beckingham K, Doyle KE, Maune JF (1987) The calmodulin gene of Drosophila melanogaster. Methods Enzymol. 139:230-247

Bowen RA (1998) Molecular Toolkit: Nucleic Acid Dot Plots. Colorado State University

Bradley RD, Hillis DM (1997) Recombinant DNA sequences generated by PCR amplification. Mol. Biol. Evol. 14:592-593

Breathnach R, Benoist C, O'Hare K, Gannon F, Chambon P (1978) Ovalbumin gene: Evidence for a leader sequence in mRNA and DNA sequences at the exon-intron boundaries. Proc. Natl. Acad. Sci. USA 75:4853-4857

Chandra A, Upadhyaya KC (1993) Structure and organization of a novel calmodulin gene of Arabidopsis thaliana. Cell. Mol. Biol. Res. 39:509-516

Chien YH, Dawid IB (1984) Isolation and characterization of calmodulin genes from Xenopus laevis. Mol. Cell. Biol. 4:507-513

Côrte Real HBSM, Dixon DR, Holland PWH (1994) Intron-targeted PCR: a new approach to survey neutral DNA polymorphism in bivalve populations. Mar. Biol. 120:407-413

Duda TF, Palumbi SR (1999) Developmental shifts and species selection in gastropods. Proc. Natl. Acad. Sci. USA 96:10272-10277

Fischer R, Koller M, Flura M, Mathews S, Strehlerpage MA, Krebs J, Penniston JT, Carafoli E, Strehler EE (1988) Multiple divergent mRNAs code for a single human calmodulin. J. Biol. Chem. 263:17055-17062

Friedberg F, Rhoads AR(2001) Evolutionary aspects of calmodulin. IUBMB Life 51:215-221

Friedberg F, Rhoads AR (2002) Multiple calmodulin genes in fish. Mol. Biol. Rep. 29:377-382

Fulton C, Lai EY, Remillard SP (1995) A flagellar calmodulin gene of Naegleria, coexpressed during differentiation with flagellar tubulin genes, shares DNA, RNA, and encoded protein sequence elements. J. Biol. Chem. 270:5839-5848

Gaunitz C, Witte H, Gauntz F (1992) Primary structure of a gene-sized DNA encoding calmodulin from the hypotrichous ciliate Stylonychia lemnae. Gene 119:191-198

Hardy DO, Bender PK, Kretsinger RH (1988) Two calmodulin genes are expressed in Arbacia punctulata, an ancient gene duplication is indicated. J. Mol. Biol. 199:223-227

Iida Y (1984) cDNA sequences and molecular evolution of calmodulin genes of chicken and eel. Bull. Chem. Soc. Japan 57:2667-2668

Jena PK, Reddy ASN, Poovaiah BW (1989) Molecular cloning and sequencing of a cDNA for plant calmodulin: signal-induced changes in the expression of calmodulin. Proc. Natl. Acad. Sci. USA 86:3644-3648

Karabinos A, Bhattacharya D (2000) Molecular evolution of calmodulin and calmodulin-like genes in the cephalochordate Branchiostoma. J. Mol. Evol. 51:141-148

Kawasaki H, Nakayama S, Kretsinger RH (1998) Classification and evolution of EF-hand proteins. Biometals 11:277-295

Lejohn HB (1989) Structure and expression of fungal calmodulin gene. J. Biol. Chem. 264:19366-19372

Ling V, Perera I, Zielinski RE(1991) Primary structures of Arabidopsis calmodulin isoforms deduced from the sequences of cDNA clones. Plant Physiol. 96:1196-1202

Liu T, Williams JG, Clarke M (1992) Inducible expression of calmodulin antisense RNA in Dictyostelium cells inhibits the completion of cytokinesis. Mol. Biol. Cell 3:1403-1413

Lukas RA, Iverson DB, Schleicher M, Watterson DM (1984) Structural characterisation of a higher plant calmodulin: Spinacia oleraces. Plant Physiol. 75:788-795

Lukas TJ, Wiggins ME, Watterson DM (1985) Amino acid sequence of a novel calmodulin from the unicellular alga Chlamydomonas. Plant Physiol. 78:477-483

Matsuo K, Sato K, Ikeshima H, Shimoda K, Takano T (1992) Four synonymous genes encode calmodulin in the teleost fish, medaka (Oryzias latipes): conservation of the multigene one-protein principle. Gene 119:279-281

Mount SM (1982) A catalogue of splice junction sequences. Nucleic Acids Res. 10:459-472

Nojima H, Sokabe H (1987) Structure of a gene for rat calmodulin. J. Mol. Biol. 193:439-445

Pieterse CM, Verbakel HM, Spaans JH, Davidse LC, Govers F (1993) Increased expression of the calmodulin gene of the late blight fungus Phytophthora infestans during pathogenesis on potato. Mol. Plant-Microbe Interact. 6:164-172

Prychitko TM, Moore WS (2000) Comparative evolution of the mitochondrial cytochrome b gene and nuclear beta-fibrinogen intron 7 in woodpeckers. Mol. Biol. Evol. 17:1101-1111

Prychitko TM, Moore WS (2003) Alignment and phylogenetic analysis of beta-fibrinogen intron 7 sequences among avian orders reveal conserved regions within the intron. Mol. Biol. Evol. 20:762-771

Reid DG (1996) Systematics and evolution of Littorina. The Ray Society, London

Robson KJH (1993) Sequence diversity in the intron of the calmodulin gene from Plasmodium falciparum. Mol. Biochem. Parasitol. 60:1-8

Robson KJH, Jennings MW (1991) The structure of the calmodulin gene of Plasmodium falciparum. Mol. Biochem. Parasitol. 46:19-34

Saporito SM, Sypherd PS (1991) The isolation and characterisation of a calmodulin-encoding gene (CMD1) from the dimorphic fungus Candida albicans. Gene 106:43-49

SenGupta B, Friedberg F, Deterawadleigh SD (1987) Molecular analysis of human and rat calmodulin complementary DNA clones - evidence for additional active genes in these species. J. Biol. Chem. 262:16663-16670

Shimoda K, Miyake T, Kimura J, Maejima K (2002) Three synonymous genes encode calmodulin in a reptile, the Japanese tortoise, Clemmys japonica. Genet. Mol. Biol. 25:43-47

Simão RDG, Gomes SL (2001) Structure, expression, and functional analysis of the gene coding for calmodulin in the chytridiomycete Blastocladiella emersonii. J. Bacteriol. 183:2280-2288

Smith VL, Doyle KE, Maune JF, Munjaal RP, Beckingham K (1987) Structure and sequence of the Drosophila melanogaster calmodulin gene. J. Mol. Biol. 196:471-485

Stein L, Sternberg P, Durbin R, Thierry-Mieg J, Spieth J (2001) WormBase: network access to the genome and biology of Caenorhabditis elegans. Nucleic Acids Res. 29:82-86

Swanson ME, Sturner SF, Schwartz JH (1990) Structure and expression of the Aplysia californica calmodulin gene. J. Mol. Biol. 216:545-553

Swofford DL (1998) PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4.0 beta 10. Sinauer Associates, Sunderland, Massachusetts

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The CLUSTAL-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25:4876-4882

Toda H, Abe Y, Yazawa M, Yagi K (1985a) Amino acid sequence of sea cucumber calmodulin. Seikagaku 57:1037-1037

Toda H, Sakiyama F, Yokono T, Miura K, Nakamura T (1987) Amino acid sequence of Pleurotus cornucopiae calmodulin. Seikagaku 59:589-589

Toda H, Yazawa M, Kondo K, Honma T, Narita K, Yagi K (1981) Amino-acid-sequence of calmodulin from scallop (Patinopecten) adductor muscle. J. Biochem. 90:1493-1505

Toda H, Yazawa M, Sakiyama F, Yagi K (1985b) Amino acid sequence of calmodulin from wheat germ. J. Biochem. 98:833-842

Watillon B, Kettmann R, Boxus P, Burny A (1992) Cloning and characterisation of an apple (Malus domesticua (L) Borkh) calmodulin gene. Plant Sci. 82:201-212

Winnepenninckx B, Backeljau T, Dewachter R (1993) Extraction of high-molecular-weight DNA from mollusks. Trends Genet. 9:407-407

Yazawa M, Yagi K, Toda H, Kondo K, Narita K, Yamazaki R, Sobue K, Kakiuchi S, Nagao S, Nozawa Y (1981) The amino-acid-sequence of the Tetrahymena calmodulin which specifically interacts with guanylate-cyclase. Biochem. Biophys. Res. Commun. 99:1051-1057

Ye Q, Berchtold MW (1997) Structure and expression of the chicken calmodulin I gene. Gene 194:63-68

Yuasa HJ, Suzuki T, Yazawa M (2001) Structural organization of lower marine nonvertebrate calmodulin genes. Gene 279:205-212

Yuasa HJ, Yamamoto H, Takagi T (1999) The structural organization of the ascidian, Halocynthia roretzi, calmodulin genes. The vicissitude of introns during the evolution of calmodulin genes. Gene 229:163-169

Zimmer WE, Schloss JA, Silflow CD, Youngblom J, and Watterson DM (1988) Structural organization, DNA-sequence, and expression of the calmodulin gene. J. Biol. Chem. 263:19370-19383

Table:

Statistical analysis comparing mutation distribution with a random Poisson distribution for individual gene copies

Gene Copy	Chi-squared value	Probability
LCaM I	43.92	0.001
LCaM II	5.49	0.483
LCaM III	4.00	0.135
LCaM IV	0.24	0.971
LCaM I microsatellite only	7.04	0.217
LCaM I without microsatellite	67.29	0.001

Table 1.

Figure Legends:

Figure 1. A diagrammatic example of an EF-Hand domain, showing the amino acids at sites required to conserve functionality. Figure adapted with kind permission from the 'Ca-Binding Proteins Database, http://structbio.vanderbilt.edu/chazin (Melanie R. Nelson & Walter Chazin)

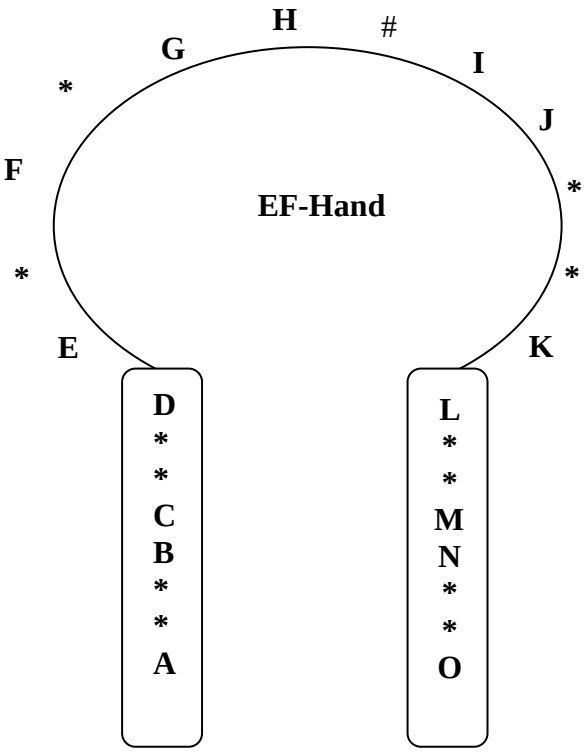
Figure 2. Typical amplification from different Littorina (and Nodilittorina (not discussed here)) species using the universal calmodulin primers of Côrte-Real et al. (1994). L = 100bp ladder (Promega, Surrey) (brightest band 500bp) NT = No Template control. Species amplicons 1-6 L. subrotundata, L. kasakta, Nodilittorina angustior, L. plena, N. natalensis, L. sitkana

Figure 3. An unrooted neighbor joining tree displaying the clustering of the different PCR products. The labels are derived from the default alignment with '*' representing allelic number, based upon order of discovery. Bootstrap percentage values are from 1000 replicates

Figure 4. Typical intron alignment among the five different copies, the sequences shown are representative of other members of the same copy.

Figure 5. The amino acid sequence across a wide range of taxa for the EF-Hand II domain of calmodulin. References a) Iida 1984; b) Matsuo et al. 1992; c) Nojima and Sokabe 1987; d) SenGupta et al. 1987; e); Chandra and Upadhyaya 1993; f) Jena et al. 1989; g) Ling et al. 1991; h) Lukas et al. 1984; i) Toda et al. 1985b; j) Watillon et al. 1992; k) Gaunitz et al. 1992; l) Lukas et al. 1985; m) Robson 1993; n) Robson and Jennings 1991; o) Zimmer et al. 1988; p) Toda et al. 1987; q) Simão and Gomes 2001; r) Lejohn 1989; s) Pieterse et al. 1993; t) Saporito and Sypherd 1991; u) Fulton et al. 1995; v) Toda et al. 1981; w) Beckingham et al. 1987; x) Smith et al. 1987; y) Toda et al. 1985a; z) Liu et al. 1992; aa) Duda and Palumbi 1999

Figure 6. The amino acid sequence across all Littorina gene copies for the EF-Hand II domain of calmodulin. LCaM I shows the correct sequence of the coding region. A dot indicates identity.



Key:

A-D, **L-O** = Hydrophobic residues
H = Glycine
* = Any residue
E-G, **J**, **K** = Calcium binding ligands
= Calcium binding ligand, provided by a backbone carbonyl
I = isoleucine or other aliphatic

Figure 1.

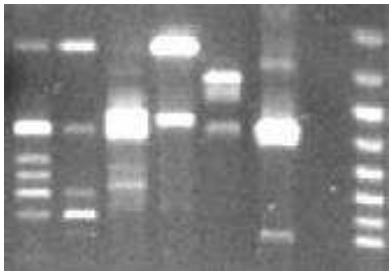


Figure 2.

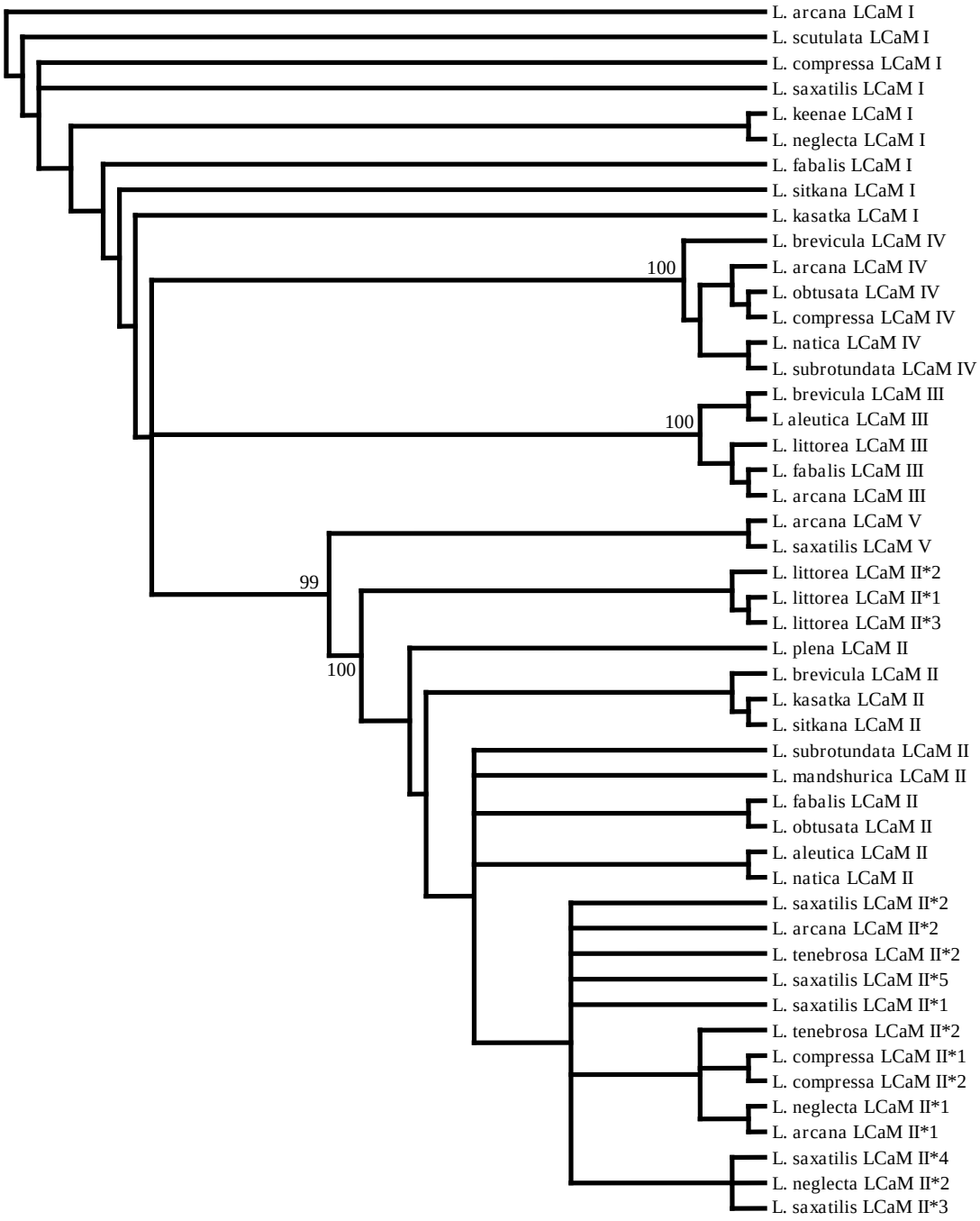

L. arcana LCaM I
L. scutulata LCaM I
L. compressa LCaM I
L. saxatilis LCaM I
L. keenae LCaM I
L. neglecta LCaM I
L. fabalis LCaM I
L. sitkana LCaM I
L. kasatka LCaM I
L. brevicula LCaM IV
100
L. arcana LCaM IV
L. obtusata LCaM IV
L. compressa LCaM IV
L. natica LCaM IV
L. subrotundata LCaM IV
L. brevicula LCaM III
100
L aleutica LCaM III
L. littorea LCaM III
L. fabalis LCaM III
L. arcana LCaM III
L. arcana LCaM V
L. saxatilis LCaM V
99
L. littorea LCaM II*2
L. littorea LCaM II*1
L. littorea LCaM II*3
L. plena LCaM II
100
L. brevicula LCaM II
L. kasatka LCaM II
L. sitkana LCaM II
L. subrotundata LCaM II
L. mandshurica LCaM II
L. fabalis LCaM II
L. obtusata LCaM II
L. aleutica LCaM II
L. natica LCaM II
L. saxatilis LCaM II*2
L. arcana LCaM II*2
L. tenebrosa LCaM II*2
L. saxatilis LCaM II*5
L. saxatilis LCaM II*1
L. tenebrosa LCaM II*2
L. compressa LCaM II*1
L. compressa LCaM II*2
L. neglecta LCaM II*1
L. arcana LCaM II*1
L. saxatilis LCaM II*4
L. neglecta LCaM II*2
L. saxatilis LCaM II*3

Figure 3.

```
L. saxatilis LCaM I      ATAAGGTATATTAAATGGCAACTAGAATGTATTCTCTTTACATTTTT
L. scutulata LCaM I      ATAAGGTATATTAAATGGCAACTAGAATGTATTCTCTTTACATTTTT
L. brevicula LCaM II     ACACTGTAATCCAAGAAGAAAATGGCACAGACCACTTTTCCTTGGGA
L. kasatka LCaM II       ACACTGTAATCAAAGAACAAATTTGCACAGACCACTTTTCCTTGGGA
L. aleutica LCaM III     A--TTTCAAGCC-AGGATTGATAAGTTC-GTATAGCCCTGGTGATGG
L. brevicula LCaM III    A--TTTCAAGCC-AGGATCGATAAGTTC-GTATATCCCTGGTGATGG
L. compressa LCaM IV     ACTTAGTAAGTTGATCAGCAAGTTAGATTGTTTGTTCATTCGTTCGT
L. obtusata LCaM IV      ACTTAGTAAGTTGATCAGCAAGTTAGATTGTTTGTTCATTCGTTCGT
L. arcana LCaM V         ATATTTGACGCCGTGCATCAGCAACAATGATTAACCCCTTGACTGG-
L. saxatilis LCaM V      ATATTTGACGCCGTGCATCAGCAACAATGATTAACCCCTTGACTGG-
```

Figure 4.

Taxa	**Reference**	**Site** (see key below)		
		A BC D	**E F GH#IJ K**	**L**
Vertebrates	a, b, c, d	LQDMINEV	DADGNGTIDFPE	FLT
Plants	e, f, g, h, i, j	....	....	..N
Ciliate	k	....	....	..S
Algae	l	.....S..	....	..M
Plasmodium	m, n	I	.T..........	...
Trypanosomes	o	....	.Q..S.......	...
Pleurotus	p	....	....	...
Blastocladiella (fungus)	q	.LV.....	....	...
Most Fungi	r, s, t	....	....	...
Naegleria	u	.H......	..........T.	...
Sea scallop	v	....	....D.......	...
Drosophila	w, x	....	....	...
Sea cucumber	y	....	....	...
Slime mould	z	....	......N.....	...
Conus	aa		....	...

Key:

A-D, **L** = Hydrophobic residues
H = Glycine
* = Any residue
E-G, **J**, **K** = Calcium binding ligands
= Calcium binding ligand, provided by a backbone carbonyl
I = isoleucine or other aliphatic

Figure 5.

Gene copy and species	**Site** (see key below)		
	A BC D	**E F GH#IJ K**	**L**
LCaM I	LQDMINEV	DADGNGTIDFPE	FLT
LCaM II	....	....	...
LCaM III L. arcana & L. littorea	.A......	....	...
L. brevicula	.S......	....	...
L. aleutica	....	.S.....SNRL.	RVPD
L. fabalis	.ARYDQRG	GR*RQRHNRL.R	VPD
LCaM IV L. saxatilis	....	....	...
L. arcana	...T....	....	...
LCaM V	....H..	..........E.	..H

Key:

A-D, **L** = Hydrophobic residues
H = Glycine
* = Any residue
E-G, **J**, **K** = Calcium binding ligands
= Calcium binding ligand, provided by a backbone carbonyl
I = isoleucine or other aliphatic

Figure 6.