Calmodulins and related potential calcium sensors of Arabidopsis

Elizabeth McCormack and Janet Braam

Department of Biochemistry and Cell Biology, Rice University, Houston, TX 77005–1892, USA

Author for correspondence:
Janet Braam
Tel: +1 713 348 5287
Fax: +1 713 348 5157
Email: braam@bioc.rice.edu

Received: *4 March 2003*
Accepted: *6 May 2003*

doi: 10.1046/j.1469-8137.2003.00845.x

Summary

• Calcium (Ca^{2+}) signaling is thought to orchestrate responses to cellular stimuli. The efficacy of Ca^{2+} signaling requires mediation by Ca^{2+}-binding proteins.

• The determination of the Arabidopsis genome sequence enables the identification of genes encoding potential Ca^{2+} sensors.

• Six Arabidopsis loci are defined as calmodulin (*CAM*) genes. Fifty additional genes are CAM-like (*CML*) genes, encoding proteins composed mostly of EF-hand Ca^{2+}-binding motifs, have no other identifiable functional domains, and at least 16% identical with CaM. Number and structural diversity of the EF hands are evaluated. Intron/exon boundaries, phylogenetic tree and chromosomal distribution data for the *CAMs* and *CMLs* are presented.

• Arabidopsis has 6 *CAM* genes, encoding only 3 isoforms. Maintenance of these genes suggests that they are unlikely to be fully redundant in function. Furthermore, the repeated EF hand motif is incorporated into at least 50 additional loci. The CaM relatives have altered EF hand number, organization, and predicted functional capacity. Additional structural differences and expression behaviors also indicate that the *CML* family has likely evolved distinct roles from the *CAMs*.

Key words: Arabidopsis, calmodulin, calcium, signal transduction, EF hands.

Introduction

Calcium (Ca^{2+}), at extracellular levels (~10^{-3}M), is a predicament for cells because it is toxic to their phosphate-based energy system. As a consequence, Ca^{2+} is actively pumped from the cytosol to extracellular spaces or intracellular compartments, such as the endoplasmic reticulum and vacuole, to maintain intracellular Ca^{2+} at low (~10^{-7}M) resting levels. Remarkably, evolution has turned this deadly ion into an essential signaling molecule. Cells capitalize on the steep Ca^{2+} gradient set up by intracellular removal of Ca^{2+}; gating of Ca^{2+} channels can result in rapid and dramatic (10- to 100-fold) increases in local intracellular [Ca^{2+}] ([Ca^{2+}]i). Through the linkage of extracellular events to Ca^{2+} influx, [Ca^{2+}]i changes can be used as second messengers to trigger physiological changes in response to external stimuli (Clapham, 1995).

In part because of the energy efficiency of this built-in signal source, Ca^{2+} may be one of the most widely used second messengers in eukaryotic cells. Undoubtedly, Ca^{2+} signaling has numerous, diverse and essential functions in plants as it does in animals (Berridge *et al.*, 1998). Ca^{2+} signals, in part, are generated, modified, propagated and perceived through the action of proteins that bind Ca^{2+} (Roberts & Harmon, 1992; Vogel, 1994; Ehlers & Augustine, 1999). The quintessential eukaryotic Ca^{2+}-binding protein is calmodulin (CaM). The central role of CaM in eukaryotic biology is reflected in its conservation (van Eldik & Watterson, 1998). For example, all known vertebrate CaMs are identical in amino acid sequence and share 91% amino acid identity to plant CaMs.

CaM is an unusual protein because it harbors no intrinsic activities of its own. It is a 148 amino acid protein with 4 repeating units, called EF hands; each EF hand binds a single Ca^{2+} ion (Strynadka & James, 1989). Cooperative binding sites enables CaM to act in clean on/off states (Strynadka & James, 1989; Nelson & Chazin, 1998b); a characteristic beneficial for cells and organisms that deal with environments capable of rapid changes. As a consequence of Ca^{2+} binding, CaM alters its structure (Babu *et al.*, 1985; Wriggers *et al.*, 1998). The structural changes reveal hydrophobic surfaces that serve to interact with and alter activities of target proteins

in a Ca^{2+}-dependent manner. CaM is therefore an exceptional protein, theoretically simple in form and biochemical function. And yet because of the potential to mediate Ca^{2+}-dependent regulation to multiple targets (Roberts & Harmon, 1992; Ohya & Botstein, 1994; Vogel, 1994; Snedden & Fromm, 2001), CaM harbors the ability to affect diverse cellular pathways. Furthermore, in addition to the highly conserved CaM, organisms also harbor CaM-like proteins that share the EF hand structure of CaM, but differ in ways that likely impact function, such as target specificity, subcellular localization and Ca^{2+} affinities (Roberts & Harmon, 1992; Zielinski, 1998; Braunewell & Gundelfinger, 1999; Haeseleer *et al.*, 2002; Luan *et al.*, 2002; Zielinski, 2002). These CaM-like proteins may have evolved to contribute to the diverse roles of Ca^{2+} signaling.

The full sequencing of the Arabidopsis genome reveals a striking complexity of *CAM*-like genes (The Arabidopsis Genome Initiative, 2000). Recently, 250 EF-hand encoding genes have been identified in the Arabidopsis genome and grouped into 6 classes (Day *et al.*, 2002). Here, we further characterize Group IV and V members, which we define as typical *CAMs* and C*aM-like* (*CML*) genes because they encode primarily EF hands. Analyses and comparisons of protein relatedness, Ca^{2+} binding potential, gene structures, chromosomal locations, and expression characteristics shed light on the evolutionary relationships among these *CAM* and *CML* genes and the potential functions of the encoded proteins. Genomic analyses provide the foundation for further studies aimed at defining the biochemical and physiological functions of the gene products.

Materials and Methods

Construction of alignments and trees

Sequences of CML proteins were downloaded from The Arabidopsis Information Resource (TAIR) (http://www.arabidopsis.org) and subjected to phylogenetic analysis. Alignments were constructed using the multiple sequence alignment mode of ClustalX (Thompson *et al.*, 1997). Alignments were subsequently viewed using SeqVu1.0.1 (Garvan Institute of Medical Research, Sydney, Australia) to shade protein alignments at positions with identical residues in greater than 65% of the aligned sequences. Protein trees were constructed using the neighbor-joining method (Saitou & Nei, 1987) implemented in the ClustalX program. Bootstrap analysis was performed using 200 iterations of tree building and varying the random seed generator.

Determination of amino acid percent identity among proteins

To determine the percentage identity between pairs of proteins, an alignment was performed in ClustalX (Thompson *et al.*, 1997) independently of other CaM or CML sequences. The number of identical residues throughout the alignment was summed and divided by the total number of amino acids in the shorter of the proteins being compared. This value was expressed as a percentage. This method emphasizes the total percentage identity between two proteins.

Determination of *CAM/CML* gene structure and nucleotide percent identity among *CAM* coding sequences

BAC clone ID number, genomic nucleotide sequences, identified strand of DNA used for transcription, and predicted intron/exon boundaries were determined by searching the Locus History available at the TAIR (http://www.arabidopsis.org). The obtained predicted intron/exon boundaries were used to construct scaled models of each open reading frame illustrating locations of introns and EF hand encoding nucleotides. Prior to phylogenetic analysis of *CAM* sequences, intron sequences were manually removed as indicated by the predicted intron splice sites from the GenBank database (http://www.ncbi.nlm.nih.gov) and TAIR (http://www.arabidopsis.org). *CAM* coding sequences were aligned in ClustalX (Thompson *et al.*, 1997). The number of identical nucleotides was summed, divided by the total number of nucleotides, and expressed as a percentage.

Chromosomal distribution and segmental duplications

Approximate gene locations were determined by searching for the appropriate open reading frames at TAIR (http://www.arabidopsis.org). Approximate locations for segmental duplications were estimated by scaling the map of segmental duplications (The Arabidopsis Genome Initiative, 2000) to match map of open reading frame locations. Inclusion of a *CML* within a segmental duplication was verified by comparing open reading frame names to a comprehensive list of segmental duplicated regions (The Arabidopsis Genome Initiative, 2000).

Expressed Sequence Tags

ESTs corresponding to *CAM/CML* genes were identified by performing a Locus History search at TAIR (http://www.arabidopsis.org) and by searching at The Institute for Genomic Research Arabidopsis Gene Index (http://www.tigr.org/tdb/tgi/agi/). Characteristics of *CAM/CML* expression were determined based on the types of libraries from which ESTs were derived.

Results and Discussion

Defining true Arabidopsis *CAMs*

There are six typical *CAMs* in Arabidopsis. To date, there has been confusion in the literature regarding gene names

and corresponding protein identities. For example, *CAM6* has been used to name a cDNA for which there is no corresponding genomic locus yet identified, *CAM5* has been mistakenly used as an alternative name for *CAM2* (first reported as *TCH1*), *CAM4* has been used as the name for two different genes, and CaM8–CaM14 have been used as names for proteins that do not meet strict criteria for being true CaMs (Chandra & Upadhyaya, 1993; Gawienowski *et al*., 1993; Ito *et al*., 1995; Luan *et al*., 2002). To reconcile these discrepancies in the literature, Tables 1 and 2 link the gene identification numbers to the *CAM* nomenclature. Part of the naming confusion is likely attributable to the high degree of nucleotide identity among the genes and the fact that the six distinct genomic loci encode only three distinct protein isoforms. *CAM2*, *CAM3* and *CAM5* encode identical gene products; CaM7 is different by one amino acid; CaM1 and CaM4 are identical differing from CaM7 by 4 amino acids.

Primary sequence comparisons among species (Fig. 1) lead to the prediction that the Arabidopsis CaM isoforms function as typical CaMs. The EF hands have the canonical 12-residue Ca^{2+} binding loop (Fig. 1). Ca^{2+} is bound in a pentagonal bipyramidal geometry with seven sites of coordination occurring through interactions with six amino acids, those in positions 1, 3, 5, 7, 9 and 12 (alternatively called X, Y, Z, #, -X and -Z) (Strynadka & James, 1989; Nelson & Chazin, 1998b). All of these amino acids interact with Ca^{2+} through side chain oxygens, except residue seven, which acts through its main chain oxygen. Chelation by residue nine sometimes occurs indirectly through a hydrogen-bonded water molecule. Thus, there are strong preferences for specific amino acids within the Ca^{2+}-binding loop. The X position is almost exclusively filled with aspartate (D); Y is usually aspartate (D) or asparagine (N); Z is aspartate (D), asparagine (N), or serine (S); the # position tolerates a variety of amino acids; -X also varies, but is usually aspartate (D), asparagine (N), or serine (S); -Z, which contributes two coordination sites, is nearly invariably glutamate (E). Glycine (G) at position 6 is highly conserved and is thought to provide the ability for a sharp turn within the loop. Finally, position 8 is most often isoleucine (I), which can form hydrogen bonds with the other EF loop in a pair. The cysteine (C) residue in position 7 of the first EF hand is common among plant CaMs (Zielinski, 1998), but uncommon in nonplant CaMs.

The E helix generally starts with a glutamate (E); both the E and F helices flanking the Ca^{2+}-binding loop are generally each 9 amino acids long. There is a regular distribution of hydrophobic amino acids in the E helices with a pattern of 'h**hh**h' where 'h' represents hydrophobic amino acids and '*' represents any amino acid. The pattern is similar for the F helices of hands 1 and 3, but diverges slightly in hands 2 and 4 (Fig. 1).

Based on the conservation of the consensus EF hand sequence motifs among the 3 Arabidopsis CaMs isoforms, it is predicted that these CaMs have Ca^{2+} binding behaviors that are similar to that of CaMs that have been extensively characterized (Fig. 1).

Sequence conservation of CaMs, including the three Arabidopsis CaM isoforms, is not restricted to the EF hand structures. Lysine (K) encoded at position 116, between the 3rd and 4th EF hands, is a potential site for trimethylation and is found in all but the yeast CaMs. (In mature CaMs that have the initiator methionine (M) removed, the K is at the 115 position.) Expression of a mutant CaM with an arginine (R) as amino acid 115 results in transgenic tobacco that have enhanced production of reactive oxygen species (Harding *et al*., 1997), indicating a role of this conserved amino acid in normal CaM function. In addition, the three Arabidopsis CaM isoforms are methionine (M)-rich proteins (Table 1). The average methionine content of proteins is 1.4%; whereas CaMs generally have approx. 6% methionine (Rose *et al*., 1985; Nelson & Chazin, 1998b). The unusual flexibility and polarizability of the methionine side chains are thought to contribute to CaM structure and function in two ways. When CaM binds Ca^{2+}, CaM undergoes structural alterations generating the so-called open conformation. In this form, hydrophobic regions become exposed. The properties of the methionine side chains make this configuration more energetically stable and are thought to enable the open CaM structure to adapt to both buried and solvent-exposed environments (Nelson & Chazin, 1998a). In addition, methionine residues serve to interact through strong van der Waals with numerous targets of distinct structural properties (O'Neil & DeGrado, 1990; Vogel & Zhang, 1995).

Because of this high degree of sequence similarity of the CaM1, CaM2, CaM3, CaM4, CaM5 and CaM7 proteins to known CaMs of other species, we consider these true CaMs. Further experimentation will be required to determine whether the three Arabidopsis isoforms of CaM have distinct functions or regulation.

CaM-like proteins of Arabidopsis

Using the databases (The Arabidopsis Genome Initiative, 2000), we developed a classification of genes, whose members we call CaM-likes or *CMLs*. A family of 50 *CML* genes (Tables 1 and 2 and Figs 2 and 3) encode proteins with the following characteristics: composed mostly, if not entirely of EF hands (like CaM); have no other identifiable functional domains, and share at least 16% overall amino acid identity with CaM. All but one (CML1) have at least 2 identifiable EF hand motifs.

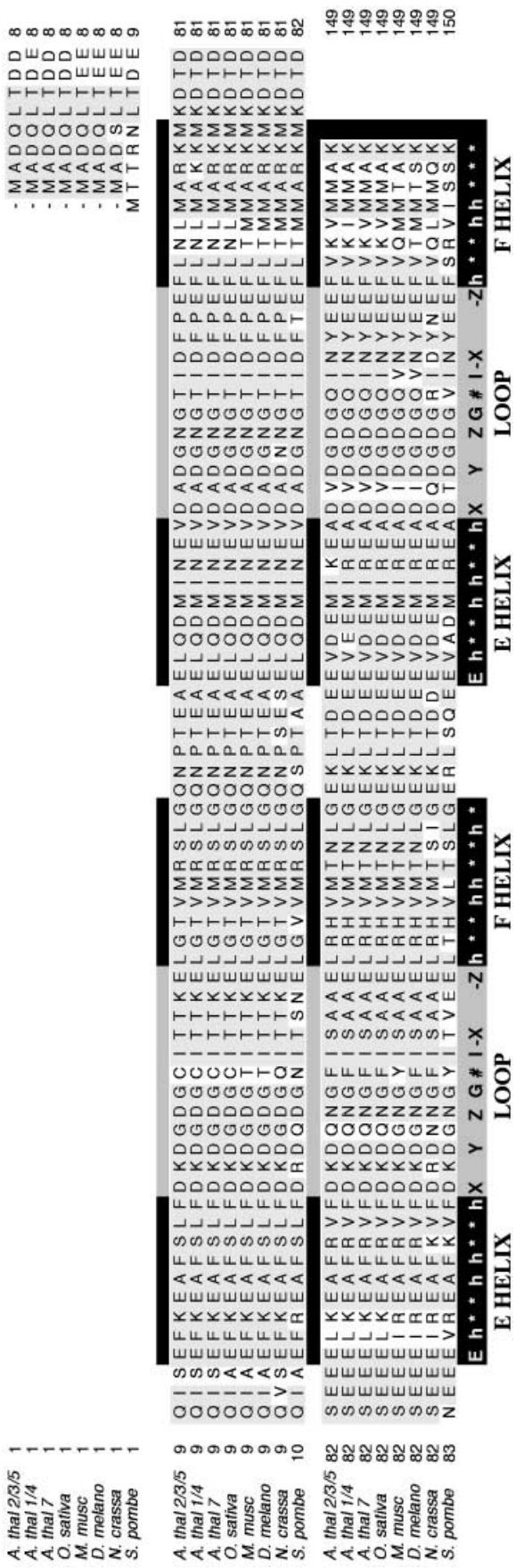
This CML class does not include a similarly large number of Arabidopsis proteins that have EF hand motifs and additional known and/or potential functional domains, such as Ca^{2+}-dependent protein kinases (CDPKs) (Roberts & Harmon, 1992), calcineurin B-like proteins (CBLs) (Luan *et al*., 2002), and *SUB1* and related *SUL1/2* genes (Guo *et al*., 2001).

Table 1 Characteristics of CaM/CML proteins

Open reading frame name	CaM/CML name	Group number defined by tree	Number of amino acids	Number of EF hands	Percentage methionine	Presence of cysteine 27	Presence of lysine 115	Potential myristoylation site	% amino acid identity to CaM2
At5g37780	CaM1	1	149	4	6.0%	+	+		96.6%
At2g41110	CaM2	1	149	4	6.0%	+	+		100.0%
At3g56800	CaM3	1	149	4	6.0%	+	+		100.0%
At1g66410	CaM4	1	149	4	6.0%	+	+		96.6%
At2g27030	CaM5	1	149	4	6.0%	+	+		100.0%
At3g43810	CaM7	1	149	4	6.0%	+	+		99.3%
At3g59450	CML1	6	148	1	2.7%				21.6%
At4g12860	CML2	6	152	4	8.6%		+		38.2%
At3g07490	CML3	6	153	4	5.9%		+		39.6%
At3g59440	CML4	6	195	4	5.6%		+		39.6%
At2g43290	CML5	6	215	4	4.7%		+		41.6%
At4g03290	CML6	6	154	4	5.8%		+		44.2%
At1g05990	CML7	6	150	4	5.3%		+	+	44.2%
At4g14640	CML8	2	151	4	4.0%	+	+		73.2%
At3g51920	CML9	2	151	4	7.9%		+		49.6%
At2g41090	CML10	2	191	4	3.7%		+		65.1%
At3g22930	CML11	2	173	4	5.2%	+	+		74.5%
At2g41100	CML12	2	324	6	4.9%				62.4%
At1g12310	CML13	3	148	3	4.0%				50.3%
At1g62820	CML14	3	148	3	4.1%		+		50.0%
At1g18530	CML15	4	157	4	4.5%				39.6%
At3g25600	CML16	4	161	4	5.0%				39.5%
At1g32250	CML17	4	166	4	1.8%				43.6%
At3g03000	CML18	4	165	4	2.4%				42.9%
At4g37010	CML19	5	167	4	6.6%				42.3%
At3g50360	CML20	5	169	4	7.1%				45.0%
At4g26470	CML21	5	248	4	3.6%			+	27.5%
At3g24110	CML22	5	229	4	3.9%				24.1%
At1g66400	CML23	8	157	4	5.7%		+		40.9%
At5g37770	CML24	8	161	4	5.0%		+	+	40.3%
At1g24620	CML25	8	186	4	4.3%			+	43.6%
At1g73630	CML26	8	163	4	4.3%				38.2%
At1g18210	CML27	8	170	4	4.1%				39.6%
At3g03430	CML28	8	83	2	4.8%				34.9%
At5g17480	CML29	8	83	2	4.8%				32.5%
At2g15680	CML30	8	187	4	4.8%				34.9%
At2g36180	CML31	8	144	4	6.3%		+		37.5%
At5g17470	CML32	8	146	4	4.8%		+		32.9%
At3g03400	CML33	8	137	3	4.4%		+		36.5%
At3g03410	CML34	8	131	4	3.1%		+		35.9%
At2g41410	CML35	9	216	4	3.2%				34.2%
At3g10190	CML36	9	209	4	2.4%				36.9%
At5g42380	CML37	6	185	3	4.3%				34.2%
At1g76650	CML38	6	177	3	5.6%				28.8%
At1g76640	CML39	6	159	4	8.2%				26.1%
At3g01830	CML40	6	146	2	3.4%				23.2%
At3g50770	CML41	6	205	4	3.4%				36.2%
At4g20780	CML42	7	191	3	2.6%				34.9%
At5g44460	CML43	7	181	3	2.2%				33.6%
At1g21550	CML44	7	155	3	2.6%				29.5%
At3g29000	CML45	7	194	2	2.6%		+		29.5%
At5g39670	CML46	7	204	2	2.9%		+		28.9%
At3g47480	CML47	7	183	2	3.8%				30.2%
At2g27480	CML48	7	186	2	2.7%				16.1%
At3g10300	CML49	7	330	2	0.9%				22.8%
At5g04170	CML50	7	315	2	1.3%				22.8%

Table 2 Characteristics of *CAM/CML* genes

Open reading frame name	*CML* name	Other name	Literature reference	Number of ESTs identified	Total number of nucleotide	Strand of DNA: Watson (W) or Crick (C)	BAC clone ID number
At5g37780		*CAM1*	41	10	1348	C	K22F20.20
At2g41110		*TCH1/CAM2*	41,42	16	813	W	T3K9.12
At3g56800		*CAM3*	40	17	939	W	T20P8.8
At1g66410		*CAM4*	20	24	1354	C	T27F4.1
At2g27030		*CAM5*	20	5	991	C	T8M16.130
At3g43810		*CAM7*	19	23	1736	C	T28A8.100
At3g59450	*CML1*			0	2083	W	F25L23.310
At4g12860	*CML2*			0	458	C	T20K18.210
At3g07490	*CML3*			0	461	W	F21O3.20
At3g59440	*CML4*			2	587	W	F25L23.300
At2g43290	*CML5*			12	647	C	T1O24.3
At4g03290	*CML6*			2	464	W	F4C21.22
At1g05990	*CML7*			1	452	W	T21E18.4
At4g14640	*CML8*	*CAM8*	17	2	2196	W	DL336OW
At3g51920	*CML9*	*CAM9*	17	6	1136	C	F4F15.30
At2g41090	*CML10*	*CABP22/CAM10*	39	10	795	W	T3K9.14
At3g22930	*CML11*	*CAM11*	16	3	1549	C	F5N5.10
At2g41100	*CML12*	*TCH3/CAM12*	42	19	1275	W	T3K9.13
At1g12310	*CML13*	*CAM13*	16	16	447	C	F5011.35
At1g62820	*CML14*	*CAM14*	16	2	1801	C	F23N19.25
At1g18530	*CML15*			0	473	W	F25I16.13
At3g25600	*CML16*			2	485	W	T5M7.5
At1g32250	*CML17*			0	500	W	F5D14.1
At3g03000	*CML18*			5	497	W	F13E7.5
At4g37010	*CML19*			7	959	W	AP22.11
At3g50360	*CML20*	*Centrin*	51	1	1081	W	F11C1.200
At4g26470	*CML21*			3	966	W	M3E9
At3g24110	*CML22*			0	984	C	MUJ8.1
At1g66400	*CML23*			6	473	C	T27F4.15
At5g37770	*CML24*	*TCH2*	42	9	485	C	K22F20.10
At1g24620	*CML25*			0	560	C	F21J9.36
At1g73630	*CML26*			2	491	W	F25P22.4
At1g18210	*CML27*			18	512	C	T10O22.19
At3g03430	*CML28*			0	252	W	T21P5.15
At5g17480	*CML29*	*APC1*	52	0	251	W	K3M16.50
At2g15680	*CML30*			1	563	W	F9O13.23
At2g36180	*CML31*			0	434	C	F9C22.11
At5g17470	*CML32*			0	440	C	K3M16.40
At3g03400	*CML33*			0	413	C	T21P5.18
At3g03410	*CML34*			0	395	C	T21P5.17
At2g41410	*CML35*	*PM129*	53	26	650	C	F13H10.4
At3g10190	*CML36*			2	629	W	F14P13.21
At5g42380	*CML37*			7	557	C	MDH9.7
At1g76650	*CML38*			4	533	C	F28O16.2
At1g76640	*CML39*			0	479	C	F28O16.1
At3g01830	*CML40*			4	440	W	F28J7.16
At3g50770	*CML41*			4	617	W	F18B3.50
At4g20780	*CML42*			1	575	C	F21C20.130
At5g44460	*CML43*			1	545	W	MFC16.12
At1g21550	*CML44*			5	467	C	F24J8.15
At3g29000	*CML45*			1	584	W	K5K13.13
At5g39670	*CML46*			10	614	W	M1J24.17
At3g47480	*CML47*			1	551	C	F1P2.30
At2g27480	*CML48*			3	933	W	F10A12.16
At3g10300	*CML49*			12	1639	W	F14P13.10
At5g04170	*CML50*			14	1842	W	F21E1.90



As summarized in Table 1, the CMLs are predicted to be relatively small proteins, ranging from 83 to 330 amino acids. Most of the CMLs (31/50) have four predicted EF hands; only one, CML12, has more than four hands (Table 1).

CML8 and CML9 have previously been called CaM8 and CaM9, respectively, in the literature (Köhler & Neuhaus, 2000; Zielinski, 2002). However, because the encoded proteins share only 73% and 49% amino acid identity with CaM, they are likely to have distinct functions from CaM. Indeed, the proteins have been shown to have different binding activities and functional complementation efficiencies of a yeast CaM null mutant (Köhler & Neuhaus, 2000; Zielinski, 2002). These results reinforce the rationale for a separate classification of CMLs from CaMs. CML8 is one of the CMLs most closely related in overall sequence to CaM (73.2% identity); yet it has been shown to be functionally distinct from CaM. Thus, the sequence variation from CaMs indicates that the CML proteins are unlikely to be true CaMs and therefore have the potential for unique, yet undiscovered, functions.

We used the neighbor-joining method analysis (Saitou & Nei, 1987) to generate a bootstrapped phylogenetic tree based on amino acid sequence similarity of the CaMs and CMLs. This analysis enables us to separate the CaM/CML family into nine groups based on apparent divergence from the typical CaMs (Fig. 2). Divergence reflects overall sequence identities to CaM2 (identical to CaM3 and CaM5) (Table 1), at least for those groups most closely aligned with CaM. For example, the CaMs that fall into group 1 share between 99.3% and 100% sequence identity to CaM2. Group 2 amino acid identities to CaM2 range from 50% to 75%, group 3 has 50% identity and group 4 ranges from 40% to 44% identity. Groups 5, 6, 7, 8 and 9 have sequence identity relatedness to CaM2 that average 35%, 35%, 28%, 37% and 36%, respectively. Groups 5, 6 and 7 have the most divergent members with percentage identities to CaM2 as low as 24%, 22%, and 16%, respectively. Although many family members show great distance from CaM based on the tree distribution, there is an overall maintenance of CaM sequence similarity that may reflect the conservation of EF hand sequences (Fig. 4).

Fig. 1 CaM isoform similarities among diverse species. The three CaM isoforms encoded by the 6 *Arabidopsis thaliana* CAM genes are aligned with CaMs predicted from other species' DNA sequences. Amino acid numberings are indicated at left and right. Note that the initiator methionine (M) is likely removed from the mature protein such that most of the mature proteins are 148 amino acids long. Sequences are positioned such that the helix-loop-helix portions of the first and third and the second and fourth EF hands, respectively, are aligned for comparison. The regions corresponding to the E helices, the Ca^{2+}-binding loops and the F helices are indicated by the black, gray, and black bars, respectively. The consensus sequences for these regions are indicated beneath the relevant sequences. 'E' stands for glutamic acid, 'h' for hydrophobic amino acid; '*' for any amino acid and 'X, Y, Z, G, #, I, -X, -Z' are defined in the text. Amino acid sequence identities are shaded.

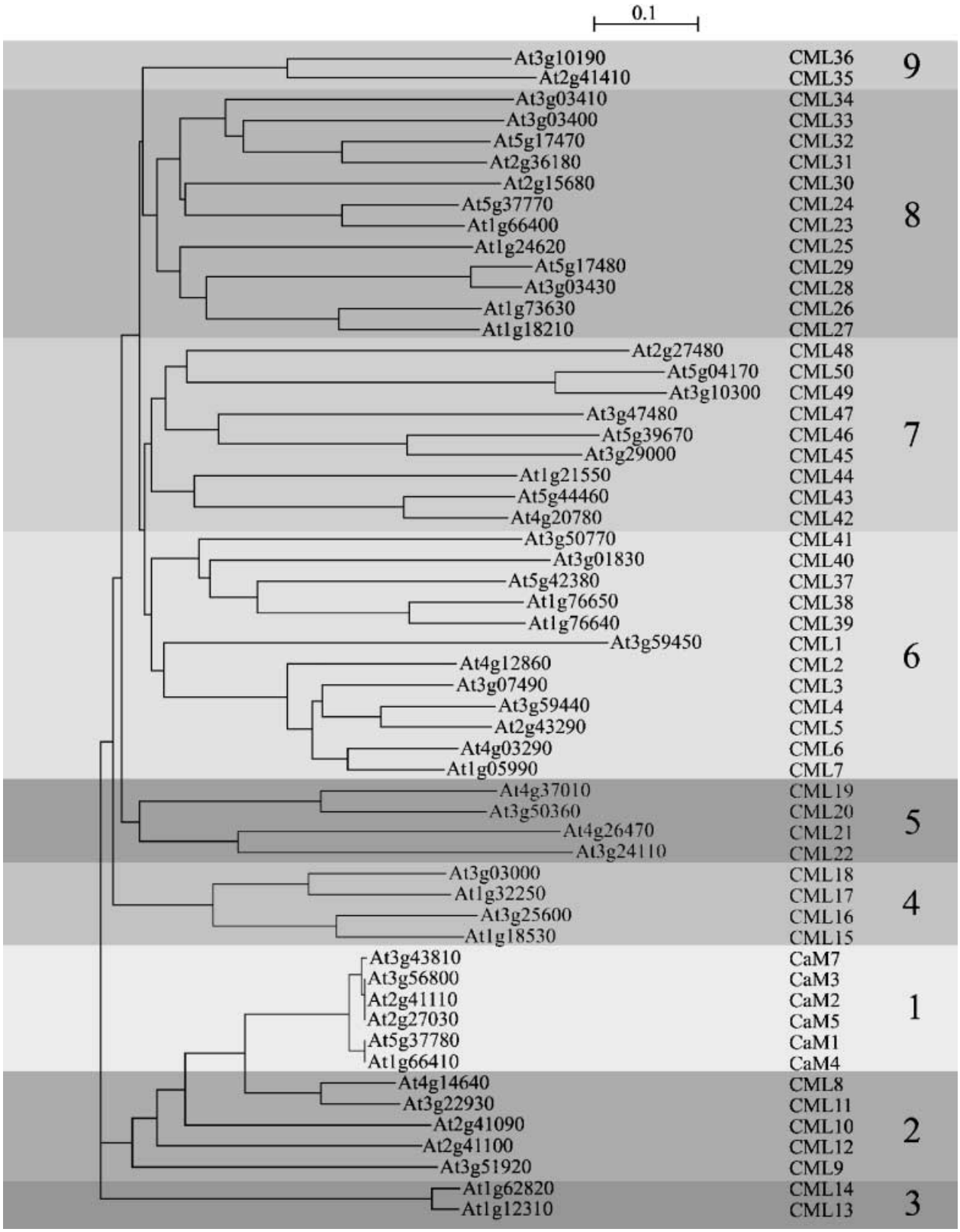


Fig. 2 Neighbor joining tree tree based on amino acid similarities segregates 9 groups of CaMs and CMLs. The amino acid sequences of the predicted CaMs and CMLs were analyzed as described in the Materials and Methods. The groupings referred to in the text and in Table 1 are indicated at the right. Both the gene identifier number and CaM/CML names are shown. The distance indicated by '0.1' refers to the percent sequence divergence as calculated by ClustalX (Thompson *et al*., 1997).

CaM2	96.6				
CaM3	96.6	100			
CaM4	100	96.6	96.6		
CaM5	96.6	100	100	96.6	
CaM7	97.3	99.3	99.3	97.3	99.3
1	CaM1	CaM2	CaM3	CaM4	CaM5

CML9	47.0			
CML10	47.7	40.4		
CML11	85.4	45.7	43.4	
CML12	52.3	49.7	54.5	48.6
2	CML8	CML9	CML10	CML11

CML16	73.9				
CML17	51.6	50.3			
CML18	51.6	51.6	71.5		
4	CML15	CML16	CML17		

CML20	67.1			
CML21	30.5	30		
CML22	24.6	26.6	36.7	
5	CML19	CML20	CML21	

CML14	94.6	
3	CML13	

CML2	23.0					
CML3	23.6	73.7				
CML4	33.1	61.8	67.3			
CML5	29.7	64.5	72.5	78.5		
CML6	25.7	61.8	66.7	64.9	68.8	
CML7	25.0	67.3	74.7	66.7	77.3	80.0
6	CML1	CML2	CML3	CML4	CML5	CML6

CML38	51.4				
CML39	52.2	76.1			
CML40	39.7	37.7	33.6		
CML41	36.2	44.1	39.6	36.3	
6	CML37	CML38	CML39	CML40	

CML43	79.6							
CML44	40.6	34.8						
CML45	27.2	29.8	30.3					
CML46	27.2	27.6	23.2	61.9				
CML47	23.0	23.8	26.5	31.7	30.1			
CML48	14.0	20.4	21.3	9.7	10.2	12.6		
CML49	17.3	18.8	18.7	21.1	16.7	18.6	30.6	
CML50	20.4	20.4	16.8	19.1	19.6	21.9	29.6	73.3
7	CML42	CML43	CML44	CML45	CML46	CML47	CML48	CML49

CML36	53.6	
9	CML35	

CML24	77.1									
CML25	43.3	46.0								
CML26	43.3	43.5	44.2							
CML27	47.7	45.3	42.4	72.4						
CML28	36.1	33.7	43.4	48.2	37.3					
CML29	32.5	35.0	37.3	45.8	37.9	82.9				
CML30	44.6	46.0	32.3	39.3	37.1	34.9	33.7			
CML31	43.1	45.8	38.2	40.3	38.2	28.9	27.7	40.3		
CML32	44.5	47.9	39.0	39.0	39.7	33.7	32.5	42.5	71.5	
CML33	43.1	40.9	36.5	34.3	35.8	31.3	30.1	40.1	48.9	51.1
CML34	40.5	42.0	32.8	33.6	33.6	33.7	33.7	35.9	45.0	47.3	43.5
8	CML23	CML24	CML25	CML26	CML27	CML28	CML29	CML30	CML31	CML32	CML33

Fig. 3 Amino acid sequence identities among proteins of distinct groups. The amino acid sequences of each member of the CaM group (group 1) and the 8 CML groups (groups 2–9) were compared. The values indicate the percent amino acid sequence identities between each given pair of proteins. The group numbers are shown in the lower left corner of each comparison table.

There are 12 examples of highly related (> 70% identity) pairs of proteins illustrated by close branches on the tree (Fig. 2) and high sequence identity (Fig. 3). The most highly related pair is composed of CML13 and CML14 that are 94.6% identical at the amino acid level (Fig. 3). These related gene pairs suggest relatively recent gene duplications.

We aligned the amino acids that compose the 172 Ca^{2+}-binding loops found in the 50 CML proteins' EF hands (additional data files) and tabulated the frequency at which amino acids are found in each site in the binding loop (Fig. 4). As is seen in true CaMs, there are strong preferences for positions 1, 3, 5, 6, 9 and 12. This indicates that the majority of EF hand motifs in the CMLs are likely to be functional, high affinity Ca^{2+}-binding sites. At the bottom of Fig. 4, a consensus sequence is shown; the positions 1, 3, 5, 6, 9, and 12 are occupied by the amino acids that most frequently fill those positions in the true CaMs. However, a subset of individual CMLs has significant sequence divergence in the Ca^{2+}-binding loops. One substitution is aspartate (D) for glutamate (E) in 12th position of the Ca^{2+}-binding loop. This substitution has been shown to increase the binding of Mg^{2+} by EF hands (Houdusse & Cohen, 1996; Cates *et al.*, 2002). At least one hand of 10 CML proteins (CML9, CML13, CML14, CML19, CML20, CML22, CML35, CML36, CML40, CML41, and CML46) has this alteration. For CML19 and CML20, this amino acid change in the second of four EF hands is the only significant alteration in the Ca^{2+}-binding loops; thus these proteins are predicted to maintain their capability of Ca^{2+}-binding but may also have increased affinity for Mg^{2+}. The coincident position of this substitution and the overall sequence similarities between CML19 and CML20 (67% amino acid identity) suggest that this E to D substitution occurred once before a duplication that led to the existence of these 2 genes (Fig. 2). The shared noncanonical sequences in CML13 and CML14 also infer a common progenitor that evolved these changes. CML13 and CML14 have additional E to D substitutions in the 12th position in the 2nd and 3rd EF hands, and the 3rd binding loop also has

	1	2	3	4	5	6	7	8	9	10	11	12	TOT
D	171	1	98	4	124	1	0	0	39	4	33	15	490
N	0	1	63	6	29	1	3	0	29	0	12	0	144
E	0	11	2	1	1	0	3	0	8	4	41	152	223
G	0	3	4	104	1	168	3	0	2	2	3	0	290
S	0	9	1	8	12	0	12	0	50	3	15	1	111
I	1	12	1	0	0	0	2	108	0	9	2	0	135
F	0	0	1	0	0	0	33	0	0	50	3	0	87
V	0	21	0	0	0	0	6	43	0	16	3	1	90
K	0	32	2	18	2	1	35	0	0	8	18	2	118
A	0	16	0	4	0	0	4	0	6	30	12	0	72
T	0	14	0	1	2	0	11	0	29	3	6	0	66
R	0	19	0	16	0	0	18	1	4	3	5	0	66
L	0	17	0	1	0	0	9	17	1	8	6	0	59
Q	0	9	0	6	0	0	6	0	1	5	5	1	33
Y	0	0	0	0	0	0	11	0	0	14	1	0	26
C	0	1	0	2	0	0	10	1	2	0	0	0	16
M	0	3	0	0	0	1	5	2	0	2	1	0	14
P	0	0	0	0	0	0	0	0	1	7	3	0	11
H	0	1	0	1	1	0	1	0	0	1	3	0	8
W	0	2	0	0	0	0	0	0	0	3	0	0	5
	D		**Dn**	**G**	**D**	**G**		**Iv**	**s**	**f**		**E**	

Fig. 4 Amino acid composition of the Ca^{2+}-binding loop in CML EF hands. The 172 predicted EF hands in the 50 CML proteins were examined. Position 1 through 12 of the Ca^{2+}-binding loop are shown at top; the single letter abbreviation for each amino acid is shown at left. The frequency at which an amino acid residue is found in each position is in the appropriate column and row. The amino acids found most frequently are shown below as a 'consensus' sequence. A capital letter is used for those amino acids that were found at that position in at least 50% of the EF hands; a lower case letter is used for those amino acids that were found at that position between 25% and 50% frequency. At right, under 'TOT' (total), are the total number of times an amino acid was found within the 172 Ca^{2+}-binding loops.

a glycine (G) in position 3 instead of an aspartate (D) or asparagine (N). These changes leave only the first Ca^{2+}-binding loop with canonical amino acids. Thus, CML13 and CML14 are predicted to be at least somewhat impaired in Ca^{2+}-binding. Similarly, the E to D substitution in the first Ca^{2+}-binding loops of CML35 and CML36 likely occurred before a gene duplication that gave rise to these two genes. CML35 also has E to D changes in the 2nd and 4th Ca^{2+}-loops and a loss of a conserved D in the third position of the second Ca^{2+}-binding loop; therefore, CML35 is predicted to have three of four hands with reduced Ca^{2+} affinities. Noncanonical amino acids in sites that are generally conserved among functional EF hands leave CML1, CML22, CML40, CML41, CML46 and CML48 with fewer than two sites likely to be fully functional in Ca^{2+} binding. The relevance of Ca^{2+} binding to the functioning of these proteins remains to be determined.

Other features of CaM are seen in subsets of the CML proteins. Nineteen of the CML genes encode the conserved lysine (K) corresponding to position 115 of CaM (Table 1), which is a potential site for trimethylation. Only the two proteins that are most closely related to typical CaM, CML8 and CML11, contain cysteine (C) residues in the first EF hand. However, nine other CMLS (CML4, CML5, CML6, CML7, CML19, CML23, CML24, CML35, CML41) have C in the 7th position of other EF hands. All but three CMLs (CML17, CML49, CML50) have greater than 2% methionine (M) (Table 1). The paucity of methionine residues in CML49 and CML50 reflects the low overall sequence relatedness of these proteins to CaM. Surprisingly, however, CML17 has over 43% sequence identity with CaM, but has only 1.8% methionine. One prediction is that these methionine-poor proteins do not act as sensor proteins and may not undergo significant conformational changes upon Ca^{2+} binding. Alternatively, they may bind only one or a few target proteins and thus may not require the flexibility of the methionine-side chains for ligand interactions.

Comparative modeling of the three dimensional structure of CML24 (also called TCH2) suggests that cysteines (C) at positions 126 and 131 may be close enough in space to have the potential to form a disulfide bond (Khan *et al.*, 1997). This potential post-translational modification suggests the possibility for regulation by re-dox state of the cell and is predicted to affect the ability of CML24 to undergo conformational changes upon loss of Ca^{2+} or target binding. Interestingly, CML23, CML25, CML26, CML27, CML33, CML35, CML36 and CML37 also have pairs of cysteines situated similarly close to each other along the primary sequence. Thus, there is the possibility that these CMLs may also have the potential to form disulfide bonds that could affect structural properties and function.

A number of EF hand containing proteins involved in synaptic activity and visual signaling regulation undergo cotranslational covalent linkage of myristate that affects subcellular localization (Ames *et al.*, 1997). CML sequences were scanned for the potential myristoylation consensus sequence G-{EDRKHPFYW}-X-X-[STAGCN]-{P} using the ProfileScan Server (http://hits.isb-sib.ch/cgi-bin/PFSCAN). Amino acids within { } are excluded from the indicated position, residues within [] are allowed at the indicated position, and X represents any amino acid. Only potential sites that begin within the first 20 amino acids were considered to be significant. Only one CML, CML21, has the strict consensus sequence for potential amino terminal myristoylation. Others, however, have potential myristoylated glycines that are near, but not directly at, the N-terminus. If these proteins (CML7,

CML24 and CML25) are proteolytically processed, it is possible that these internal glycines could be recognized for modification (http://www.cbs.dtu.dk/services/TargetP/). None of the CMLs have the carboxyl-terminal CAAX motif for prenylation. TargetP (http://www.cbs.dtu.dk/services/TargetP/; Emanuelsson *et al.*, 2000) analysis suggests that the hydrophobic-rich amino-terminal extensions of CML46 and CML47 may function as endoplasmic reticulum signal sequences. Similarly, the amino-terminal sequences of CML41 may act to direct this protein into chloroplasts. However, the great majority of the CaMs and CMLs are predicted to be cytosolic or possibly nuclear.

CAM and *CML* gene structures

The nucleotide sequence variation, ranging from 83 to 91% identity, among the *CAMs* is higher than that of the amino acid sequence variability (Table 1). The fact that multiple genes, with nucleotide variability, encode identical proteins suggests the presence of selective pressure on the strict maintenance of amino acid sequence. Arabidopsis CaMs therefore show a similar conservation of sequence that is seen in vertebrates; vertebrates also have 3 *CAM* genes, like the *CAM2*, *CAM3* and *CAM5* genes, which encode identical CaM proteins. Although this conservation of sequence could be an example of genomic redundancy, it is difficult to explain how natural selection can act to keep the protein sequences identical. If multiple *CAM* genes were truly redundant, one would expect some sequence divergence at least among the genes from the distinct vertebrate species (Toutenhoofd & Strehler, 2000). One possibility is that *CAM* genes are differentially expressed and therefore the products function with spatial or temporal specificity (Toutenhoofd & Strehler, 2000).

The six Arabidopsis *CAM* genes share the characteristic of having a single intron disrupting the coding sequence for the first EF hand at codon glycine (G) 26 (Fig. 5). Only 13 of the

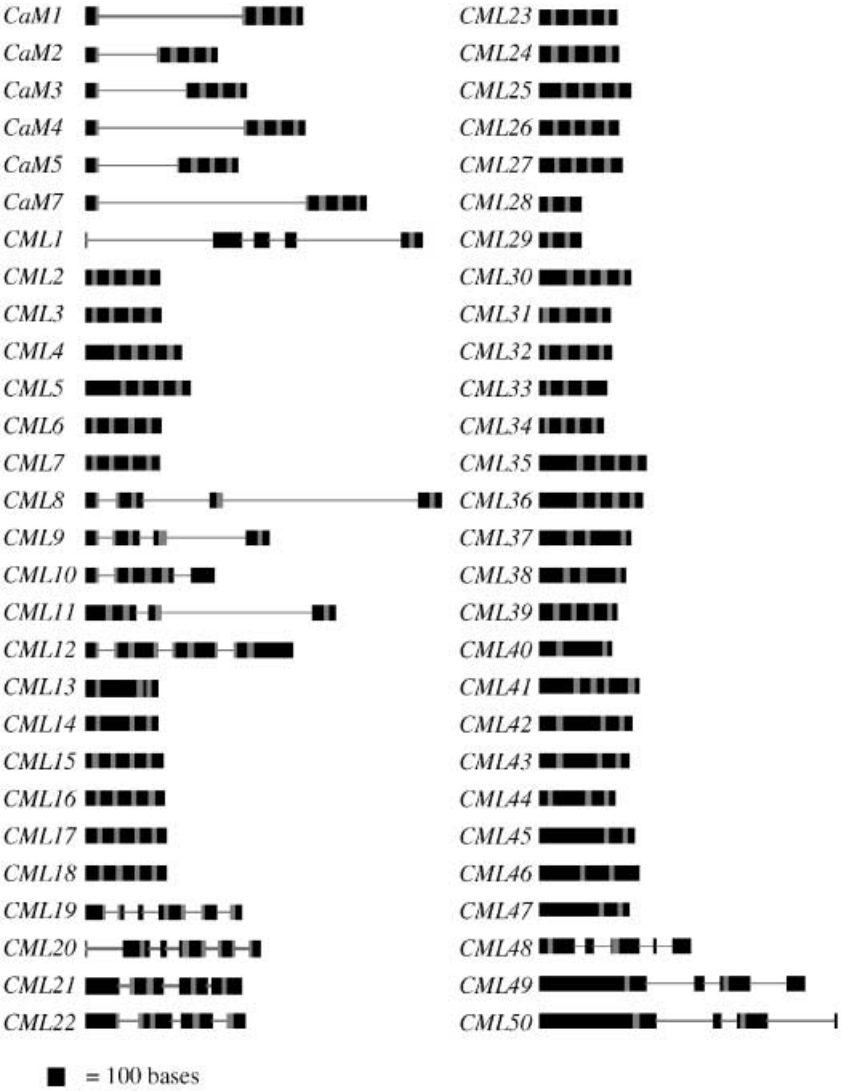


Fig. 5 Predicted presence and prediction of introns, exons and EF hand-coding sequences in the Arabidopsis CAMs and CMLs. Intron and exon boundaries were determined by comparisons of genomic DNA with cDNAs or predicted based on genomic sequences. EF hands were identified by presence of canonical sequences (see Figs 1 and 4) and alignment with related CaMs and CMLs as described in the text. Thin lines represent introns, thick bars represent exons and gray regions indicate positions of regions encoding EF hands. The size marker at bottom indicates the distance for 100 bases.

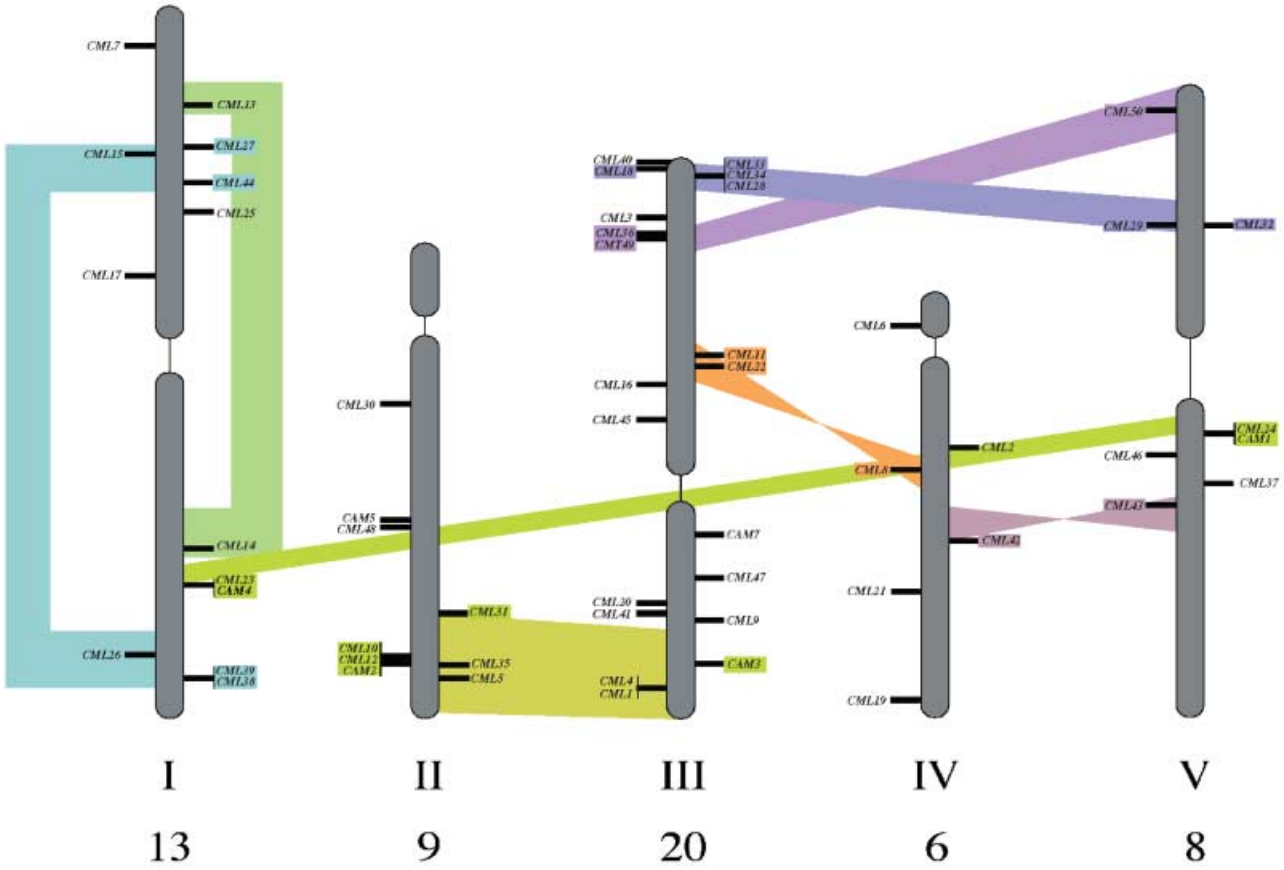


Fig. 6 The CAMs and CMLs are distributed among the 5 Arabidopsis chromosomes. The arms of the 5 Arabidopsis chromosomes are indicated as rounded gray bars; centromeric sites dividing arms are represented by thin connecting lines. The positions of the *CAM* and *CML* genes are indicated. Regions of predicted segmental duplication (http://www.arabidopsis.org) are indicated by color-specific shading. The *CAM* and *CML* names and numbers are shaded for those genes found within predicted duplication regions. Genes with names at left of chromosomes are on the Watson strand; genes with names at right of chromosomes are on the Crick strand (see also Table 2).

50 *CMLs* are interrupted by introns, the location for five of these (*CML8*, *CML9*, *CML10*, *CML12* and *CML22*) is comparable to *CAM* G26 position (Fig. 4). Interestingly, the 2nd and 3rd introns of *CML12* are also in comparable positions, suggesting that the gene region encoding the second and third pairs of hands was derived from the 5′ end of the gene encoding the first pair of hands. Indeed sequence similarity comparisons of EF domains are consistent with this idea (additional data files).

Chromosomal distribution

The family members are distributed on the five chromosomes (Fig. 6). Thirty-five of the 56 *CAMs* and *CMLs* are in regions of the genome that are thought to be derived from segmental duplications (The Arabidopsis Genome Initiative, 2000). These regions include pairs of genes that encode identical CaMs (*CAM1* and *CAM4*; *CAM2* and *CAM3*) and highly related *CMLs* that pair in the phylogenetic tree (Fig. 3), including *CML4/CML5*, *CML8/CML11*, *CML13/CML14*, *CML23/CML24*, *CML26/CML27*, *CML28/CML29*, *CML42/CML43* and *CML49/CML50*. These *CML* pairs encode proteins that share between 72% and 95% amino acid identity. There are also six chromosomal sites of tandemly arranged *CAMs* and/or *CMLs* (Fig. 6). Four of the pairs cluster close together in the phylogenetic tree (Fig. 3) (*CML1/CML4*, *CML10/CML12*, *CML33/CML34*, and *CML38/CML39*). *CML24* and *CAM1* are adjacent on chromosome 5, and the gene most related to *CML24*, *CML23*, lies in tandem with *CAM4*, the gene most related to *CaM1* (CML23 and CML24 are 77% identical; CaM1 and CaM4 are 100% identical, Fig. 3). Thus, most likely, there was a local duplication and divergence followed by a segmental duplication. Another *CAM* gene, *CAM2* is found in a tandem grouping with *CML12* and *CML10* on chromosome 2. *CML10* is the closest relative to *CML12*.

CAM and *CML* expression

Identification of expressed sequence tags (ESTs) corresponding to the *CAM/CML* genes provides evidence for *CAM/CML* expression. In addition, because many of the EST libraries

Table 3 Evidence for CAM/CML expression

ORF NAME	*CAM/CML*	Roots	Flowers and /or Siliques	Seed	Rosette	Liquid Cultured Seedlings	Stress and/or Hormones
at5g37780	*CAM1*	1		1	1		4
at2g41110	*CAM2*	5	2				5
at3g56800	*CAM3*		1				
at1g66410	*CAM4*		6				5
at2g27030	*CAM5*	2	1	1			4
at3g43810	*CAM7*	2	3				4
at3g59450	*CML1*						
at4g12860	*CML2*						
at3g07490	*CML3*						
at3g59440	*CML4*						2
at2g43290	*CML5*	4		1	1		3
at4g03290	*CML6*		2				
at1g05990	*CML7*						
at4g14640	*CML8*						
at3g51920	*CML9*					1	2
at2g41090	*CML10*				1		7
at3g22930	*CML11*						1
at2g41100	*CML12*	2				2	13
at1g12310	*CML13*		2	2	1		5
at1g62820	*CML14*						1
at1g18530	*CML15*						
at3g25600	*CML16*	1	2				
at1g32250	*CML17*						
at3g03000	*CML18*						2
at4g37010	*CML19*	1					
at3g50360	*CML20*				1		
at4g26470	*CML21*	1					
at3g24110	*CML22*						
at1g66400	*CML23*				1	1	2
at5g37770	*CML24*	1		2			2
at1g24620	*CML25*						
at1g73630	*CML26*						
at1g18210	*CML27*	5	5				4
at3g03430	*CML28*						
at5g17480	*CML29*						
at2g15680	*CML30*						2
at2g36180	*CML31*						
at5g17470	*CML32*						
at3g03400	*CML33*						
at3g03410	*CML34*						
at2g41410	*CML35*	3	2		1		10
at3g10190	*CML36*	2					
at5g42380	*CML37*						4
at1g76650	*CML38*	2					2
at1g76640	*CML39*						
at3g01830	*CML40*		2			2	
at3g50770	*CML41*				1		3
at4g20780	*CML42*						1
at5g44460	*CML43*						
at1g21550	*CML44*						2
at3g29000	*CML45*						
at5g39670	*CML46*	3					4
at3g47480	*CML47*						1
at2g27480	*CML48*		2				1
at3g10300	*CML49*	2	2				1
at5g04170	*CML50*	3	2	3			2

Numbers indicate frequency of EST identification (www.tigr.org/tdb/tgi/agi/).

were made from RNA present in distinct tissues or organs, or after treatment of plants with stresses or hormones, some characteristics of *CAM/CML* expression can be inferred. These data are summarized in Tables 2 and 3. ESTs have been found for 42 of the 56 *CAM/CML* genes. The EST data suggest that some of the *CAM/CML* genes may be up regulated by stress and/or hormones because the greatest numbers of ESTs were found in libraries generated from plants subjected to stress or treated with hormones. For example, *CML10*, *CML12* and *CML35* expression levels are likely to be stress and/or hormone induced.

Three of the genes for which no EST expression data are evident (*CML1*, *CML17* and *CML22*) encode proteins that have lost specific features of true CaMs; CML1 and CML22 are predicted to have fewer than 2 functional EF hands and CML17 has only 1.8% methionine. Two gene pairs encoding highly related proteins that fall in the same segmental duplication (CML28/CML29 and CML33/CML34) also have no evidence yet for expression. However, lack of EST identification to date is not strong evidence for nonexpression; sensitive methods of RNA detection will be required to determine if these genes are active.

Zielinski and colleagues have monitored expression of several of the *CAM* and *CML* genes (Ling *et al.*, 1991; Perera & Zielinski, 1992; Gawienowski *et al.*, 1993; Ling & Zielinski, 1993; Zielinski, 2002). For example, *CAM1-CAM4* transcripts are in siliques and leaves; *CAM1* mRNAs are detected in roots. *CML7*, *CML8* and *CML9* expression is detectable in leaves, flowers and developing siliques and *CML10* RNAs are in leaves.

An unusual expression feature of some plant *CAMs* and *CMLs* is that they are rapidly and dramatically up-regulated by mechanical force, such as simple touch stimulation (Braam & Davis, 1990). Furthermore, *CAM2*, *CML12* and *CML24* are up-regulated in expression in response to darkness, cold shock, heat shock, and phytohormones (Braam & Davis, 1990; Braam, 1992; Sistrunk *et al.*, 1994; Antosiewicz *et al.*, 1995; Polisensky & Braam, 1996). Detailed analysis of *CML12* expression has been monitored by direct RNA analyses (Sistrunk *et al.*, 1994), patterns of *CML12::GUS* reporter gene fusion activities *in planta* (Sistrunk *et al.*, 1994) and immunolocalization (Antosiewicz *et al.*, 1995). These data can be summarized to conclude that *CML12* expression correlates with sites predicted to be under mechanical stress, such as branch points and at other sites where cells are undergoing expansion.

These data accumulated to date are the first indications of the possible sites of function of the *CAM/CML* genes of Arabidopsis. Complete characterization of expression behaviors by direct RNA analyses and patterns of reporter gene activities and the phenotypic consequences of gene knockouts will be invaluable as the next steps in understanding the functional significance of this unexpectedly large family of genes encoding CaMs and CaM-like proteins in Arabidopsis.

Acknowledgements

This material is based upon work was supported by the National Science Foundation (under grant no. IBN9982654 to J.B.) and, in part, by the National Institutes of Health (Biotechnology Training Grant no T32-GM08362 to E.M.).

Supplementary material

The following material is available to download from http://www.blackwellpublishing.com/products/journals/suppmat/NPH/NPH845/NPH845sm.htm

Fig. S1 The alignments used to construct the tree in Fig. 2.

Fig. S2 The alignments used to tabulate amino acid frequency in the Ca^{2+}-binding loops of the CMLs shown in Fig. 4.

Fig. S3 The alignments used to compare CML12 EF hands to CaM.

Fig. S4–S12 The alignments showing intragroup amino acid sequence identities.

References

Ames JB, Ishima R, Tanaka T, Gordon JI, Stryer L, Ikura M. 1997. Molecular mechanics of calcium-myristoyl switches. *Nature* **389**: 198–202.

Antosiewicz DM, Polisensky DH, Braam J. 1995. Cellular localization of the Ca^{2+} binding TCH3 protein of Arabidopsis. *Plant Journal* **8**: 623–636.

Babu YS, Sack JS, Greenhough TJ, Bugg CE, Means AR, Cook WJ. 1985. Three-dimensional structure of calmodulin. *Nature* **315**: 37–40.

Berridge MJ, Bootman MD, Lipp P. 1998. Calcium-a life and death signal. *Nature* **395**: 645–648.

Braam J. 1992. Regulated expression of the calmodulin-related *TCH* genes in cultured *Arabidopsis* cells: Induction by calcium and heat shock. *Proceedings of the National Academy of Sciences, USA* **89**: 3213–3216.

Braam J, Davis RW. 1990. Rain-, wind- and touch-induced expression of calmodulin and calmodulin-related genes in *Arabidopsis. Cell* **60**: 357–364.

Braunewell KH, Gundelfinger ED. 1999. Intracellular neuronal calcium sensor proteins: a family of ER-hand calcium-binding proteins in search of a function. *Cell Tissue Research* **295**: 1–12.

Cates MS, Teodoro ML, Phillips GJ. 2002. Molecular mechanisms of calcium and magnesium binding to parvalbumin. *Biophysics Journal* **82**: 1133–1146.

Chandra A, Upadhyaya KC. 1993. Structure and organization of a novel calmodulin gene of *Arabidopsis thaliana. Cell Molecular Biological Research* **39**: 509–516.

Clapham DE. 1995. Calcium signaling. *Cell* **80**: 259–268.

Day IS, Reddy VS, Ali GS, Reddy ASN. 2002. Analysis of EF-hand-containing proteins in Arabidopsis. *Genome Biology* **3**: 56.1–56.56.24.

Ehlers MD, Augustine GJ. 1999. Calmodulin at the channel gate. *Nature* **399**: 105–108.

van Eldik LJ, Watterson MD. 1998. Calmodulin and calcium signal transduction: an introduction. In: Van Eldik LJ, Watterson MD, eds. *Calmodulin and signal transduction.* New York, USA: Academic Press, 1–15.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology* **300**: 1005–1016.

Gawienowski MC, Szymanski D, Perera IY, Zielinski RE. 1993. Calmodulin isoforms in *Arabidopsis* encoded by multiple divergent mRNAs. *Plant Molecular Biology* **22**: 215–225.

Guo H, Mockler T, Duong H, Lin CT. 2001. SUB1, an Arabidopsis Ca^{2+}-binding protein involved in cryptochrome and phytochrome coaction. *Science* **291**: 487–490.

Haeseleer F, Imanishi Y, Sokal I, Filipek S, Palczewski K. 2002. Calcium-binding proteins: intracellular sensors from the calmodulin superfamily. *Biochemistry and Biophysics Research Communications* **290**: 615–623.

Harding SA, Oh S-H, Roberts DM. 1997. Transgenic tobacco expressing a foreign calmodulin gene shows an enhanced production of active oxygen species. *EMBO Journal* **16**: 1137–1144.

Houdusse A, Cohen C. 1996. Structure of the regulatory domain of scallop myosin at 2A resolution. *Structure* **4**: 21–32.

Ito T, Hirano M, Akama K, Shimura Y, Okada K. 1995. Touch-inducible genes for calmodulin and a calmodulin-related protein are located in tandem on a chromosome of *Arabidopsis thaliana*. *Plant and Cell Physiology* **36**: 1369–1373.

Khan AR, Johnson KA, Braam J, James MNG. 1997. Comparative modeling of the three-dimensional structure of the calmodulin-related TCH2 protein from Arabidopsis. *Proteins* **27**: 144–153.

Köhler D, Neuhaus G. 2000. Characterization of calmodulin binding to cyclic nucleotide-gated ion channels from *Arabidopsis thaliana*. *FEBS Letters* **471**: 133–136.

Ling V, Perera IY, Zielinski RE. 1991. Primary structures of *Arabidopsis* calmodulin isoforms deduced from the sequences of cDNA clones. *Plant Physiology* **96**: 1196–1202.

Ling V, Zielinski RE. 1993. Isolation of an Arabidopsis cDNA sequence encoding a 22kDa calcium-binding protein (CaBP-22) related to calmodulin. *Plant Molecular Biology* **22**: 207–214.

Luan S, Kudla J, Rodriguez-Concepcion M, Yalovsky S, Gruissem W. 2002. Calmodulins and calcineurin B-like proteins: calcium sensors for specific signal response coupling in plants. *Plant Cell* **14**: S389–S400.

Nelson MR, Chazin WJ. 1998a. An interaction-based analysis of calcium-induced conformational changes in Ca^{2+} sensor proteins. *Protein Science* **7**: 270–282.

Nelson MR, Chazin WJ. 1998b. Calmodulin as a calcium sensor. In: Van Eldik LJ, Watterson MD, eds. *Calmodulin and signal transduction*. New York, USA: Academic Press, 17–64.

O'Neil KT, DeGrado WF. 1990. How calmodulin binds its targets: Sequence independent recognition of amphiphilic alpha-helices. *Trends in Biochemistry and Science* **15**: 59–64.

Ohya Y, Botstein D. 1994. Diverse essential functions revealed by complementing yeast calmodulin mutants. *Science* **263**: 963–966.

Perera IY, Zielinski RE. 1992. Structure and expression of the *Arabidopsis* CaM-3 calmodulin gene. *Plant Molecular Biology* **19**: 649–664.

Polisensky DH, Braam J. 1996. Cold-shock regulation of the Arabidopsis *TCH* genes and the effects of modulating intracellular calcium levels. *Plant Physiology* **111**: 1271–1279.

Roberts DM, Harmon AC. 1992. Calcium-modulated proteins: targets of intracellular calcium signals in higher plants. *Annual Review of Plant Physiological Plant Molecular Biology* **43**: 375–414.

Rose GD, Geselowitz AR, Lesser GJ, Lee RH, Zehfus MH. 1985. Hydrophobicity of amino acid residues in globular proteins. *Science* **229**: 834–838.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biological Evology* **4**: 406–425.

Sistrunk ML, Antosiewicz DM, Purugganan MM, Braam J. 1994. *Arabidopsis TCH3* encodes a novel Ca^{2+} binding protein and shows environmentally induced and tissue-specific regulation. *Plant Cell* **6**: 1553–1565.

Snedden WA, Fromm H. 2001. Calmodulin as a versatile calcium signal transducer in plants. *New Phytologist* **151**: 35–66.

Strynadka NCJ, James MNG. 1989. Crystal structure of the helix-loop-helix calcium-binding proteins. *Annual Review of Biochemistry* **58**: 951–998.

The Arabidopsis Genome Initiative. 2000. Analysis of the genomic sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **24**: 4876–4882.

Toutenhoofd SL, Strehler EE. 2000. The calmodulin multigene family as a unique case of genetic redundancy: multiple levels of regulation to provide spatial and temporal control of calmodulin pools? *Cell Calcium* **28**: 83–96.

Vogel HJ. 1994. Calmodulin: a versatile calcium mediator protein. *Biochemistry and Cell Biology* **72**: 357–376.

Vogel HJ, Zhang M. 1995. Protein engineering and NMR studies of calmodulin. *Molecular Cell Biochemistry* **149/150**: 3–15.

Wriggers W, Mehler E, Pitici F, Weinstein H, Schulten K. 1998. Structure and dynamics of calmodulin in solution. *Biophysics Journal* **74**: 1622–1639.

Zielinski RE. 1998. Calmodulin and calmodulin-binding proteins in plants. *Annual Review of Plant Physiological and Plant Molecular Biology* **49**: 697–725.

Zielinski RE. 2002. Characterization of three new members of the *Arabidopsis thaliana* calmodulin gene family: conserved and highly diverged members of the gene family functionally complement a yeast calmodulin null. *Planta* **214**: 446–455.