CHAPTER III

RESULT AND DISCUSSION

Identification and phylogenetic analysis of EF-hand-containing proteins

To identify EF-hand-containing proteins, firstly, I functionally searched the *Oryza sativa* L. genome at The Institute for Genomic Research (TIGR) (Yuan *et. al.*, 2005) for Interpro Database Matches by five different methods including HMMPfam, HMMSmart, BlastProDom, ProfileScan and superfamily. Secondly, I searched the rice database using the amino acid sequences of rice CaM1 (Phean-o-pas *et. al.*, 2005) and CBL3 (Kolukisaoglu *et. al.*, 2004) as queries in the programs BLASTp and the protein sequences that were not found by the domain searches were added to the list. In addition, I reviewed literature on reports of EF-hand-containing proteins in rice that have been identified by various methods. All of these protein sequences were again analyzed for EF hands and other domains using InterProScan (Quevillon *et. al.*, 2005). InterProScan is a protein domain identifying tool that combines different protein signature recognition methods from the consortium member databases of the Interpro (Mulder *et. al.*, 2005). As a result, domain searches identified 245 proteins but six sequences did not have an EF hand identifiable by InterProScan using default settings, so they were eliminated from further analysis. BLAST searches have found four more EF-hand-containing proteins and literature review has yielded no additional proteins. Totally, a maximum of 243 putative EF-hand-containing proteins in rice have been identified as shown in Table 3.1. Nearly half of these proteins contain no other identifiable domains predicted by InterProScan. It should be noted that 24 proteins contain a single EF-hand motif that was identified by

Table 3.1 Characteristics of 243 putative EF-hand-containing proteins in rice

TIGR Locus[1]	Chr[2]	Alternative splicing[3]	C aM & CML name	Coding region (bp)[4]	Amino acids[5]	Number of EF hands[6]
LOC_Os01g04330	1	-	**OsCML16**	546	181	4
LOC_Os01g11414	1	-		1755	584	2
LOC_Os01g16240	1	-	**OsCaM1-3**	450	149	4
LOC_Os01g17190	1	-	**OsCaM3**	450	149	4
LOC_Os01g25820	1	-		2718	905	1
LOC_Os01g32120	1	-	**OsCML11**	636	211	4
LOC_Os01g39134	1	1		1191	396	4
LOC_Os01g39134	1	2		849	282	2
LOC_Os01g39134	1	3		735	244	4
LOC_Os01g39770	1	1		873	290	3
LOC_Os01g39770	1	2		657	218	1
LOC_Os01g41510	1	-		657	218	3
LOC_Os01g41990	1	-	**OsCML12**	750	249	4
LOC_Os01g43410	1	-		1287	428	4
LOC_Os01g48680	1	-		2274	757	1
LOC_Os01g51420	1	-		750	249	3
LOC_Os01g53294	1	-		2184	727	3
LOC_Os01g56030	1	-		663	220	2
LOC_Os01g57470	1	-		336	111	1
LOC_Os01g59360	1	-		1548	515	4
LOC_Os01g59530	1	1	**OsCML1**	564	187	4
LOC_Os01g59530	1	2		450	149	4
LOC_Os01g61590	1	-		1656	551	4
LOC_Os01g61880	1	1		2532	843	1
LOC_Os01g61880	1	2		2229	742	1
LOC_Os01g62570	1	-		1905	634	1
LOC_Os01g62740	1	-		633	210	1
LOC_Os01g72080	1	-		615	204	2
LOC_Os01g72100	1	-	**OsCML10**	558	185	4
LOC_Os01g72530	1	-	**OsCML31**	456	151	3
LOC_Os01g72540	1	-	**OsCML23**	591	196	3
LOC_Os01g72550	1	-	**OsCML19**	441	146	3
LOC_Os01g72940	1	-		1854	617	2
LOC_Os02g03020	2	-		444	147	1
LOC_Os02g03410	2	-		1569	522	4
LOC_Os02g06340	2	-		1632	543	1
LOC_Os02g09840	2	-		1581	526	1
LOC_Os02g10470	2	-		687	228	3
LOC_Os02g10740	2	-		1551	516	3
LOC_Os02g12880	2	1		1179	392	3
LOC_Os02g12880	2	2		1152	383	4
LOC_Os02g12880	2	3		903	300	2

Table 3.1 (Continued)

TIGR Locus[1]	Chr[2]	Alternative splicing[3]	CaM & CML name	Coding region (bp)[4]	Amino acids[5]	Number of EF hands[6]
LOC_Os02g14980	2	-		1611	536	1
LOC_Os02g15510	2	-		1245	414	1
LOC_Os02g18880	2	-		849	282	3
LOC_Os02g18930	2	-		828	275	3
LOC_Os02g27820	2	-		1479	492	1
LOC_Os02g27880	2	-		834	277	2
LOC_Os02g27920	2	-		804	267	1
LOC_Os02g27940	2	-		519	172	2
LOC_Os02g37580	2	-		2022	673	1
LOC_Os02g39380	2	-	**OsCML17**	495	164	4
LOC_Os02g39550	2	-		2304	767	2
LOC_Os02g39950	2	-		1386	461	3
LOC_Os02g42200	2	1		5463	1820	1
LOC_Os02g42200	2	2		5151	1716	1
LOC_Os02g43800	2	-		5610	1869	1
LOC_Os02g46090	2	-		1650	549	4
LOC_Os02g48740	2	-		2121	706	1
LOC_Os02g50060	2	-	**OsCML28**	507	168	4
LOC_Os02g50080	2	-		420	139	2
LOC_Os02g50140	2	-		654	217	1
LOC_Os02g50174	2	-		1122	373	2
LOC_Os02g52540	2	-		303	100	2
LOC_Os02g55880	2	-		528	175	2
LOC_Os02g58520	2	-		1638	545	4
LOC_Os03g03660	3	1		1695	564	3
LOC_Os03g03660	3	2		1629	542	3
LOC_Os03g03830	3	-		3207	1068	4
LOC_Os03g14590	3	-		1026	341	1
LOC_Os03g15740	3	-		2547	848	2
LOC_Os03g19720	3	-		522	173	3
LOC_Os03g20370	3	-	**OsCaM1-1**	450	149	4
LOC_Os03g21380	3	-	**OsCML27**	450	149	3
LOC_Os03g21890	3	-		2382	793	1
LOC_Os03g25070	3	-		1803	600	1
LOC_Os03g27790	3	-		468	155	2
LOC_Os03g29770	3	-		477	158	2
LOC_Os03g33570	3	-		309	102	1
LOC_Os03g42840	3	-		678	225	3
LOC_Os03g44180	3	-		345	114	1
LOC_Os03g48270	3	-		1725	574	4
LOC_Os03g50760	3	-		975	324	2
LOC_Os03g53200	3	-	**OsCML4**	465	154	4

Table 3.1 (Continued)

TIGR Locus[1]	Chr[2]	Alternative splicing[3]	CaM & CML name	Coding region (bp)[4]	Amino acids[5]	Number of EF hands[6]
LOC_Os03g54100	3	-		1044	347	1
LOC_Os03g55960	3	-		558	185	2
LOC_Os03g57450	3	-		1800	599	4
LOC_Os03g57510	3	-		1731	576	4
LOC_Os03g58910	3	-		8109	2702	1
LOC_Os03g59390	3	-		1617	538	4
LOC_Os03g59590	3	-		1929	642	2
LOC_Os03g59600	3	-		762	253	1
LOC_Os03g59770	3	-		642	213	2
LOC_Os03g59790	3	-		603	200	2
LOC_Os03g59870	3	-		624	207	2
LOC_Os04g30050	4	-		351	116	1
LOC_Os04g34440	4	-		1866	621	1
LOC_Os04g41540	4	-	**OsCML22**	594	197	4
LOC_Os04g41950	4	-		2262	753	1
LOC_Os04g42430	4	-		1419	472	3
LOC_Os04g43170	4	-		2319	772	1
LOC_Os04g43200	4	-		735	244	1
LOC_Os04g45180	4	-		867	288	1
LOC_Os04g47300	4	-		1602	533	4
LOC_Os04g49510	4	1		1656	551	4
LOC_Os04g49510	4	2		1317	438	1
LOC_Os04g51240	4	-		519	172	1
LOC_Os04g57350	4	-		1629	542	1
LOC_Os04g58480	4	-		1095	364	4
LOC_Os05g03610	5	-		1797	598	1
LOC_Os05g05460	5	-		750	249	3
LOC_Os05g05710	5	-		1467	488	2
LOC_Os05g06890	5	-		1752	583	2
LOC_Os05g06920	5	-		1680	559	2
LOC_Os05g06940	5	-		1737	578	2
LOC_Os05g13580	5	-	**OsCML18**	477	158	4
LOC_Os05g22270	5	-		240	79	1
LOC_Os05g24780	5	-	**OsCML21**	525	174	3
LOC_Os05g26660	5	-		1728	575	1
LOC_Os05g31620	5	-	**OsCML15**	606	201	4
LOC_Os05g38980	5	-		2460	819	1
LOC_Os05g39090	5	-		1644	547	4
LOC_Os05g40930	5	-		1125	374	1
LOC_Os05g41090	5	-		1551	516	3
LOC_Os05g41200	5	-	**OsCML9**	468	155	1
LOC_Os05g41210	5	-	**OsCaM2**	450	149	4

Table 3.1 (Continued)

TIGR Locus[1]	Chr[2]	Alternative splicing[3]	CaM & CML name	Coding region (bp)[4]	Amino acids[5]	Number of EF hands[6]
LOC_Os05g41270	5	-		1569	522	4
LOC_Os05g45210	5	-		2856	951	1
LOC_Os05g45810	5	-		633	210	3
LOC_Os05g50180	5	-	**OsCML14**	522	173	4
LOC_Os05g50810	5	-		1629	542	4
LOC_Os06g07560	6	-	**OsCML30**	711	236	4
LOC_Os06g11030	6	-		300	99	1
LOC_Os06g12690	6	2		750	249	1
LOC_Os06g12690	6	3		747	248	1
LOC_Os06g14324	6	-		678	225	1
LOC_Os06g14350	6	1		678	225	1
LOC_Os06g14350	6	2		552	183	1
LOC_Os06g14370	6	1		729	242	1
LOC_Os06g14370	6	3		519	172	1
LOC_Os06g14370	6	4		477	158	1
LOC_Os06g40200	6	-		1548	515	3
LOC_Os06g40720	6	-		675	224	3
LOC_Os06g46950	6	-		378	125	1
LOC_Os06g47000	6	3		1401	466	1
LOC_Os06g47000	6	1		1767	588	1
LOC_Os06g47330	6	-		1635	544	1
LOC_Os06g47640	6	-	**OsCML29**	513	170	3
LOC_Os06g49790	6	-		1671	556	1
LOC_Os06g50030	6	-		1743	580	1
LOC_Os06g51250	6	-		3567	1188	4
LOC_Os07g01810	7	-		1050	349	1
LOC_Os07g06740	7	1		1707	568	4
LOC_Os07g06740	7	3		639	212	4
LOC_Os07g12240	7	-		453	150	2
LOC_Os07g22710	7	1		1539	512	4
LOC_Os07g22710	7	2		1275	424	1
LOC_Os07g22710	7	3		1056	351	4
LOC_Os07g22710	7	4		717	238	1
LOC_Os07g33110	7	-		1602	533	4
LOC_Os07g38120	7	-		1653	550	4
LOC_Os07g42660	7	-	**OsCML13**	510	169	4
LOC_Os07g42714	7	-		276	91	2
LOC_Os07g42730	7	-		465	154	1
LOC_Os07g42770	7	-		1362	453	1
LOC_Os07g43800	7	-		642	213	1
LOC_Os07g44710	7	-		1785	594	1
LOC_Os07g48130	7	-		2367	788	1

Table 3.1 (Continued)

TIGR Locus[1]	Chr[2]	Alternative splicing[3]	CaM & CML name	Coding region (bp)[4]	Amino acids[5]	Number of EF hands[6]
LOC_Os07g48340	7	-	**OsCML24**	456	151	3
LOC_Os07g48780	7	1	**OsCaM1-2**	450	149	4
LOC_Os07g48780	7	2		342	113	3
LOC_Os08g02420	8	-	**OsCML7**	438	145	2
LOC_Os08g04890	8	-	**OsCML32**	591	196	3
LOC_Os08g07940	8	-		3399	1132	1
LOC_Os08g34070	8	-		663	220	1
LOC_Os08g34340	8	-		528	175	2
LOC_Os08g35210	8	-		3102	1033	1
LOC_Os08g39290	8	-		408	135	2
LOC_Os08g42750	8	-		1698	565	4
LOC_Os08g44390	8	1		762	253	2
LOC_Os08g44390	8	2		651	216	1
LOC_Os08g44390	8	3		528	175	1
LOC_Os08g44660	8	-		600	199	2
LOC_Os09g24580	9	-		486	161	2
LOC_Os09g26660	9	-		3024	1007	1
LOC_Os09g28490	9	-		558	185	2
LOC_Os09g28500	9	-		318	105	2
LOC_Os09g28510	9	-		297	98	2
LOC_Os09g30490	9	-		414	137	2
LOC_Os09g30506	9	-		414	137	2
LOC_Os09g31000	9	-		417	138	1
LOC_Os09g31040	9	-		414	137	2
LOC_Os09g33910	9	-		1734	577	4
LOC_Os10g02680	10	-		270	89	2
LOC_Os10g06790	10	-		699	232	1
LOC_Os10g09850	10	-		1161	386	1
LOC_Os10g13550	10	1		1416	471	1
LOC_Os10g13550	10	2		1329	442	1
LOC_Os10g13550	10	3		912	303	1
LOC_Os10g25010	10	-	**OsCML8**	576	191	4
LOC_Os10g33680	10	-		1626	541	1
LOC_Os10g36710	10	-		1899	632	1
LOC_Os10g39420	10			1605	534	3
LOC_Os10g41510	10	-		642	213	3
LOC_Os11g01270	11	-		2013	670	1
LOC_Os11g01390	11	-	**OsCML25**	594	197	3
LOC_Os11g03980	11	-	**OsCML2**	552	183	4
LOC_Os11g04170	11	-		1626	541	4
LOC_Os11g04480	11	-		927	308	2
LOC_Os11g04560	11	-		588	195	2

Table 3.1 (Continued)

TIGR Locus[1]	Chr[2]	Alternative splicing[3]	CaM & CML name	Coding region (bp)[4]	Amino acids[5]	Number of EF hands[6]
LOC_Os11g04770	11	-		441	146	2
LOC_Os11g04820	11	-		171	56	1
LOC_Os11g07040	11	-		1542	513	4
LOC_Os11g30180	11	-		321	106	1
LOC_Os11g33120	11	-		2811	936	1
LOC_Os11g33360	11	-		354	117	1
LOC_Os11g37550	11	-	**OsCML6**	513	170	4
LOC_Os11g38780	11	-		576	191	2
LOC_Os12g01240	12	-		2001	666	1
LOC_Os12g01400	12	-	**OsCML26**	450	149	3
LOC_Os12g03816	12	-	**OsCML3**	552	183	4
LOC_Os12g03970	12	-		1626	541	4
LOC_Os12g04240	12	1		879	292	2
LOC_Os12g04240	12	2		759	252	2
LOC_Os12g04360	12	-		588	195	2
LOC_Os12g04440	12	-		1908	635	1
LOC_Os12g04560	12	-		309	102	2
LOC_Os12g04580	12	-		309	102	2
LOC_Os12g04680	12	-		243	80	1
LOC_Os12g06100	12	-		1278	425	1
LOC_Os12g06510	12	1		885	294	3
LOC_Os12g06510	12	2		681	226	3
LOC_Os12g07230	12	-		1581	526	4
LOC_Os12g10460	12	-		1467	488	1
LOC_Os12g12730	12	-	**OsCML28**	573	190	4
LOC_Os12g12860	12	-		1692	563	4
LOC_Os12g20080	12	-		693	230	1
LOC_Os12g26010	12	-		975	324	1
LOC_Os12g30150	12	-		1839	612	4
LOC_Os12g35610	12	1		2733	910	1
LOC_Os12g35610	12	2		2679	892	1
LOC_Os12g40510	12	-		678	225	3
LOC_Os12g41110	12	-	**OsCML5**	501	166	4

[1] The Institute of Genomics Research (TIGR) gene identifier number.
[2] Chromosome number in which the gene resides.
[3] Number of alternative splicing within the gene
[4] Length of the coding region in base pairs.
[5] Number of amino acids of the deduced amino acid sequence.
[6] Number of EF hands based on the prediction by InterProScan.

only one prediction program and could be false positives.

Next, sequences of all the proteins identified by the InterProScan as containing an EF-hand motif were aligned using Clustal X (Thompson *et. al.*, 1997). Tree construction using the neighbor-joining method and bootstrap analysis was performed. Figure 3.1 shows the tree outline illustrating the numbers of EF hands predicted by InterProScan for each protein on the right without any gene identifiers. As a result, proteins that do not possess functional domains other the Ca^{2+}-binding EF-hand motifs were found distributed across the tree but most were concentrated in the top half. Conversely, most proteins in the bottom half contain additional domains that give clues to their functions which include transcription factor, ion channel, DNA- or ATP/GTP-binding protein, mitochondrial carrier protein, protein phosphatase and protein kinase. Two known major groups of EF-hand-containing proteins: calcineurin B-like (CBL) (Kolukisaoglu *et. al.*, 2004) and Ca^{2+}-dependent protein kinase (CPK) proteins (Asano *et. al.*, 2005) are separately grouped as shown in Figure 3.1. I observed that most of the proteins containing four EF-hand motifs are either in the CPK group or located at the top of the tree surrounding the typical CaM proteins. With the exception of two, all proteins indicated by "CaM & CML" share at least 25% amino acid identity with OsCaM1 and were selected for further analyses. This list should contain rice proteins that are related to CaM or has functions based on Ca^{2+}-binding mode similar to CaM. Existence of these genes and their deduced amino acid sequences were confirmed using another annotation database, the Rice Annotation Project Database (RAP-DB) (Ohyanagi *et. al.*, 2006).

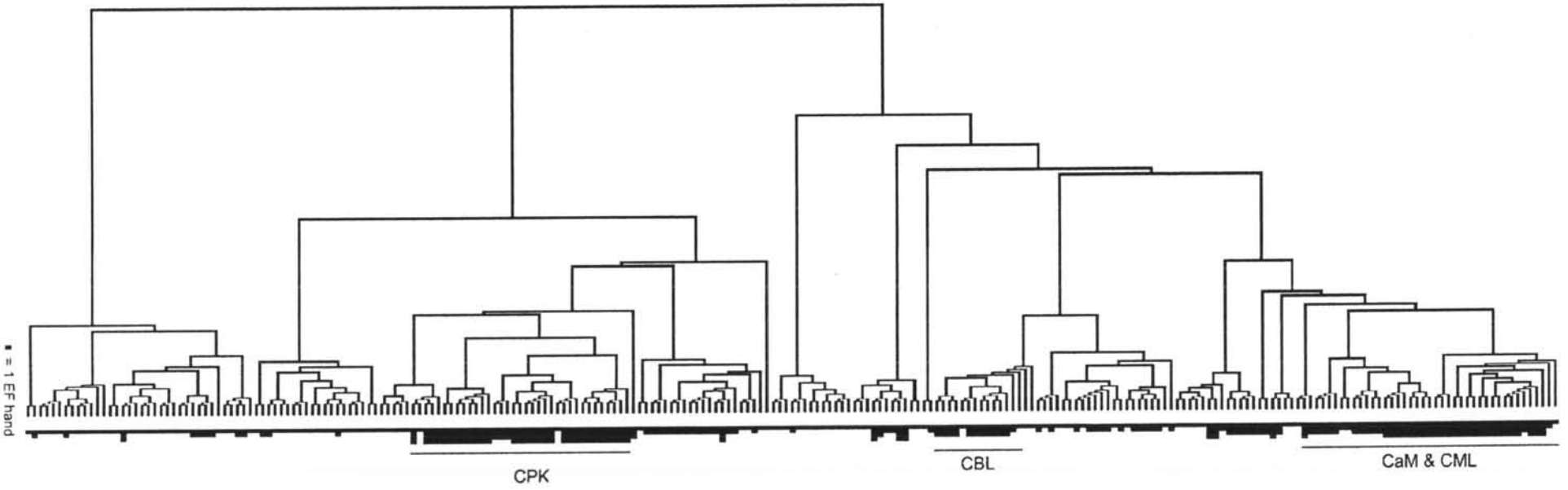


Figure 3.1 Phylogenetic tree showing the overall relatedness of EF-hand-containing proteins in rice. Alignment of full-length protein sequences and phylogenetic analysis were performed as described in the "Materials and Methods" section. The numbers of EF hands predicted by InterProScan for each protein are shown as black blocks on the right with their heights proportional to their numbers of motif. With the exception of two proteins, all proteins indicated by the vertical line labelled "CaM & CML" at the right share more than 25% amino acid identity with OsCaM1 and were selected for further analyses. Positions of CBL and CPK members are also shown along the tree to emphasize their separation.

Rice CaM proteins

The full-length amino acid sequences of the selected proteins were subjected to phylogenetic analysis. Tree construction using the neighbor-joining method and bootstrap analysis performed with ClustalX generated a consensus tree which is depicted in Figure 3.2. This analysis led us to separate these proteins into six groups: 1-6. What defines a "true" CaM and distinguishes it from a CaM-like protein that serves a distinct role *in vivo* is still an open question. Different experimental approaches including biochemical and genetic analyses have been taken to address this question (Buaboocha *et. al.*, 2002). In this study by phylogenetic analysis based on amino acid sequence similarity, five proteins in group 1 that have the highest degrees of amino acid sequence identity (≥ 97%) to known typical CaMs from other plant species were identified. Because of these high degrees of amino acid identity, I classified them as "true" CaMs that probably function as typical CaMs. They were named OsCaM1-1, OsCaM1-2, OsCaM1-3, OsCaM2 and OsCaM3. Their characteristics are summarized in Table 3.2.

OsCam1-1; *OsCam1-2* and *OsCam1-3* encode identical proteins, whereas *OsCam2* and *OsCam3* encode a protein of only two amino acid differences and their sequences share 98.7% identity with those of OsCaM1 proteins. Multiple sequence alignment of the OsCaM amino acid sequences with those of typical CaMs from other species shown in Figure 3.3 indicates their high degree of sequence conservation. It should be mentioned that OsCaM1 amino acid sequences are identical to those of typical CaMs from barley (*H. vulgare*) and wheat (*T. aestivum*) reflecting the close relationships among monocot cereal plants. On average, OsCaM amino acid sequences share about 99%, 90% and 60% identity with those from plants, vertebrates and yeast, respectively.

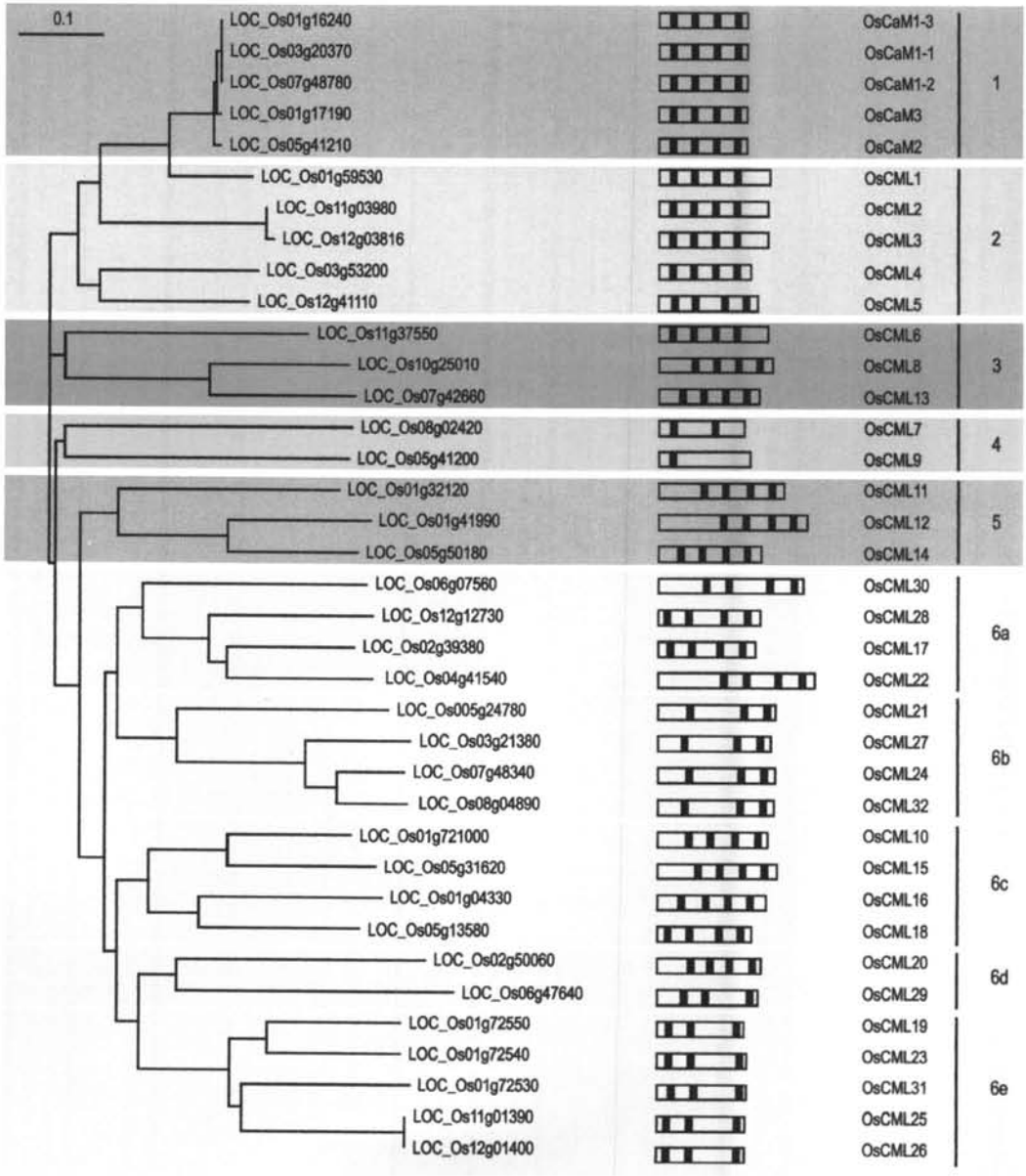


Figure 3.2 Neighbor-joining tree based on amino acid similarities among OsCaM and OsCML proteins. Tree construction using the neighbor-joining method and bootstrap analysis was performed with ClustalX. The TIGR gene identifier numbers are shown and the resulting groupings of CaM and CaM-like proteins designated as 1-6 are indicated on the right. Schematic diagrams of the OsCaM and OsCML open reading frames show their EF hand motif distribution.

Table 3.2 Characteristics of *OsCam* and *OsCML* genes and the encoded proteins

Name	Locus[1]	Chr[2]	cDNA length[3]	Amino Acids[4]	EF hands[5]	% of Met[6]	Identity to OsCaM1(%)[7]	Cys 27[8]	Lys 116[9]	Prenyl-ation[10]	Myristo ylation[11]	References
OsCam1-1	LOC_Os03g20370	3	450	149	4	6.0	100.0	+	+			[10]
OsCam1-2	LOC_Os07g48780	7	450	149	4	6.0	100.0	+	+			
OsCam1-3	LOC_Os01g16240	1	450	149	4	6.0	100.0	+	+			
OsCam2	LOC_Os05g41210	5	450	149	4	6.0	98.7	+	+			[10]
OsCam3	LOC_Os01g17190	1	450	149	4	6.0	98.7	+	+			
OsCML1	LOC_Os01g59530	1	564	187	4	4.3	84.6			+		[8, 9, 10]
OsCML2	LOC_Os11g03980	11	552	183	4	4.9	70.3			+		
OsCML3	LOC_Os12g03816	12	552	183	4	4.9	68.9			+		
OsCML4	LOC_Os03g53200	3	465	154	4	6.5	68.9	+	+			
OsCML5	LOC_Os12g41110	12	501	166	4	4.8	62.2	+	+			
OsCML6	LOC_Os11g37550	11	513	170	4	6.5	53.9	+				
OsCML7	LOC_Os08g02420	8	447	148	2	2.8	47.7		+			
OsCML8	LOC_Os10g25010	10	576	191	4	5.2	47.0					
OsCML9	LOC_Os05g41200	5	468	155	1	3.2	46.1					
OsCML10	LOC_Os01g72100	1	558	185	4	4.3	45.6		+			
OsCML11	LOC_Os01g32120	1	636	211	4	1.4	44.1					
OsCML12	LOC_Os01g41990	1	750	249	4	2.8	43.9					
OsCML13	LOC_Os07g42660	7	510	169	4	5.3	43.6					
OsCML14	LOC_Os05g50180	5	522	173	4	4.6	43.3					
OsCML15	LOC_Os05g31620	5	606	201	4	4.0	40.7					
OsCML16	LOC_Os01g04330	1	546	181	4	3.9	40.5					
OsCML17	LOC_Os02g39380	2	495	164	4	4.9	37.7		+			
OsCML18	LOC_Os05g13580	5	477	158	4	5.7	37.7		+			
OsCML19	LOC_Os01g72550	1	441	146	3	7.5	37.2					
OsCML20	LOC_Os02g50060	2	525	174	4	4.0	35.3				+	
OsCML21	LOC_Os05g24780	5	594	197	3	4.6	35.3					
OsCML22	LOC_Os04g41540	4	753	250	4	3.6	35.2					
OsCML23	LOC_Os01g72540	1	456	151	3	7.9	35.1					
OsCML24	LOC_Os07g48340	7	594	197	3	3.0	33.9					
OsCML25	LOC_Os11g01390	11	450	149	3	6.7	33.6					
OsCML26	LOC_Os12g01400	12	450	149	3	6.7	33.6					
OsCML27	LOC_Os03g21380	3	573	190	3	3.2	33.3					
OsCML28	LOC_Os12g12730	12	519	172	4	4.8	33.1		+			
OsCML29	LOC_Os06g47640	6	513	170	3	4.1	33.1					
OsCML30	LOC_Os06g07560	6	711	236	4	2.1	32.8					
OsCML31	LOC_Os01g72530	1	456	151	3	5.3	31.6					
OsCML32	LOC_Os08g04890	8	591	196	3	2.6	30.2					

[1] The Institute of Genomics Research (TIGR) gene identifier number.
[2] Chromosome number in which the gene resides.
[3] Length of the coding region in base pairs.
[4] Number of amino acids of the deduced amino acid sequence.
[5] Number of EF hands based on the prediction by InterProScan.
[6] Percentage of methionine (M) residues in the deduced amino acid sequence.
[7] Number of identical residues divided by the total number of amino acids that have been aligned expressed in percentage.
[8] Presence of a cysteine equivalent to Cys26 of typical plant CaMs at residue 7(-Y) of the first EF-hand.
[9] Presence of a lysine equivalent to Lys115 of typical plant CaMs.
[10] Presence of a putative prenylation site.
[11] Presence of a putative myristoylation site.

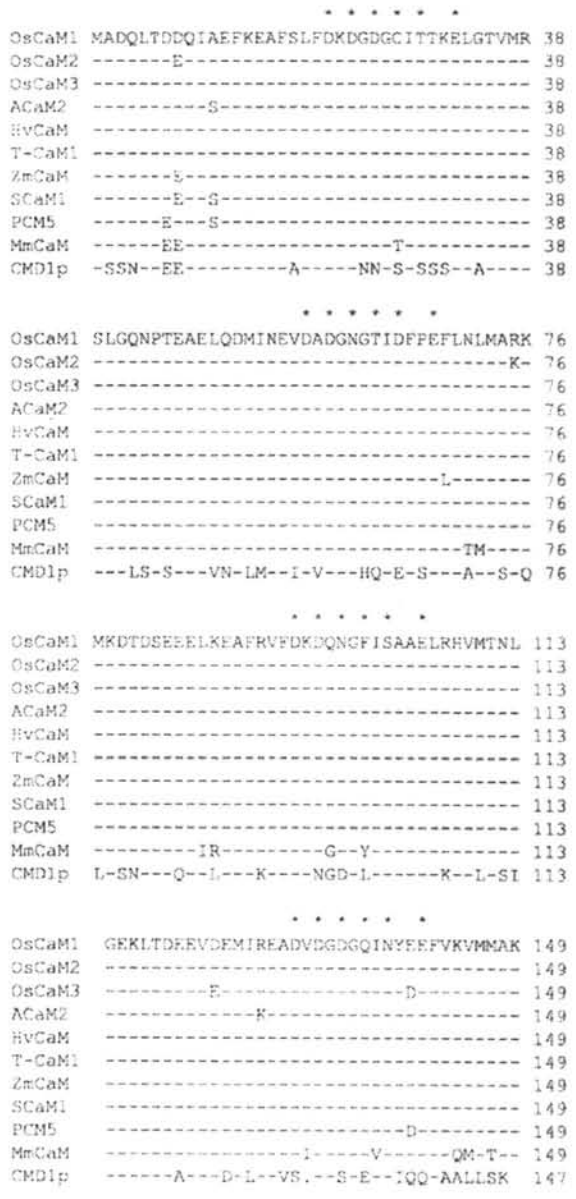


Figure 3.3 OsCaM protein sequence similarity with CaM from other species. Comparison of the deduced amino acid sequences of OsCaM1, 2, and 3 with those of other plants, Mus Musculus CaM (MmCaM), and Saccharomyces cerevisiae CaM (CMD1p). The sequences are compared with OsCaM1 as a standard; identical residues in other sequences are indicated by a dash (-), and a gap introduced for alignment purposes is indicated by a dot (.). Residues serving as Ca^{2+}-binding ligands are marked with asterisks (*).

Hydrophobic residues contributing to hydrophobic interaction in the mechanism of CaM-target protein complex formation which are critical to CaM function are highly conserved. All of the conserved eight methionine (M) and nine phenylalanine (F) residues among plant CaMs are present in all OsCaMs. Conservation of these residues is maintained between plant and vertebrate CaMs, with the exception of the methionine residues at position 145-146 in plants CaMs, which are displaced one residue compared with the vertebrate proteins. Due to its considerable conformational flexibility (Gellman SH, 1991) and being weakly polarized, methionine residues which are estimated to contribute nearly half of the accessible surface area of the hydrophobic patches of CaM allow it to interact with target proteins in a sequence-independent manner (O'Neil and DeGrado, 1990). Sequence conservation related to functionality of plant CaMs also includes lysine (K) at position 116 which is assumed to be trimethylated. All OsCaM proteins possess a lysine residue at this position. Lysine 116 trimethylation is believed to be a posttranslational modification that helps regulate CaM activity.

The presence of multiple CaM isoforms is a defining characteristic of CaMs in plants. Even though the explanation of gene redundancy still cannot be ruled out, accumulating evidence suggests that each of the *Cam* genes may have distinct and significant functions. Previous reports have shown that highly conserved CaM isoforms actually modulate target proteins differently (Karita *et. al.*, 2004). Induced expression of some but not all of the multiple CaM isoforms in a plant tissue in response to certain stimuli has been reported (Phean-o-pas *et. al.*, 2005, Yamakawa *et. al.*, 2001) thus, competition among CaM isoforms for target proteins may be found. It is fascinating that the *OsCam1-1*, *OsCam1-2*, and *OsCam1-3* genes encode identical proteins. How these

protein sequences have been maintained with the natural selection pressure throughout evolution has no clear answer yet but it is likely that each of these genes has physiological significance.

Rice CaM-like (CML) proteins

The remaining proteins from the phylogenetic analysis in Figure 3.2 were named CaM-like or CML according to the classification by McCormack and Braam (McCormack and Braam, 2003). Like CaM, these proteins are composed entirely of EF hands with no other identifiable functional domains. A summary of their characteristics is shown in Table 3.2. They were named according to their percentages of amino acid identity with OsCaM1 which were calculated by dividing the number of identical residues by the total number of residues that had been aligned to emphasize the identical amino acids. These proteins are small proteins consisting of 145 to 250 amino acid residues and sharing amino acid identity between 30.2% to 84.6% with OsCaM1. All the CML proteins in group 2 share more than 60% of amino acid sequence identity with OsCaM1. The CML proteins in groups 3, 4, and 5 have identities with OsCaM1 that average 48.2%, 46.9%, and 43.8%, respectively. By the bootstrapped phylogenetic tree based on amino acid sequence similarity of these proteins, group 6 CML proteins were separated into five subgroups: 6a-6e. These proteins share identities no more than 40.7% with OsCaM1 that average at 35.6% with the exception of OsCML10 (45.6%). All members of groups 6b and 6e contain three EF-hand motifs though with different configurations.

Some important CaM functional features were found existing only in a few CaM-like proteins. The characteristic cysteine (C) at residue 7(-Y) of the first EF hand, a hallmark of higher plant CaM sequences is absent in all CaM-like proteins with the exception of three highly conserved CML proteins, which are OsCML4, OsCML5 and OsCML6. Based on multiple sequence alignment, OsCML4, OsCML5, OsCML7 OsCML10, OsCML17, OsCML18, and OsCML28 are the only CaM-like proteins that contain lysine at a position equivalent to the Lys116 of CaMs. These features may be indicators of proteins that serve similar *in vivo* functions with those of CaMs. OsCML4 and OsCML5 are the only CaM-like proteins that possess both of these signature characteristics. However, another important determinant of CaM function, which is a high percentage of methionine (M) residues, has been found in most of the OsCML proteins. The average percentage of M residues among OsCMLs is 4.6% compared with 6.0% in OsCaMs. Considering the usually low percentage found in other proteins, the Met-rich feature in CMLs is likely an indication of their relatedness to CaMs and possibly similar mechanisms of action i.e. exposure of hydrophobic residues caused by conformational changes upon Ca^{2+} binding. Nonetheless, some newly attained characteristics specific to CMLs probably allow them to fine-tune their Ca^{2+}-regulated activity to more specialized functions.

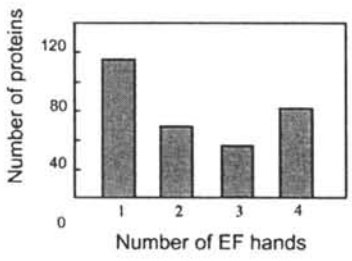
Of these proteins, three OsCMLs contain an extended C-terminal basic domain and a CAAX (C is cysteine, A is aliphatic, and X is a variety of amino acids) motif, a putative prenylation site (CVIL in OsCML1 and CTIL in OsCML2 and 3). OsCML1, also known as OsCaM61 was identified as a novel CaM-like protein by Xiao and colleagues (Xiao *et. al.,* 1999). The CML protein was reported to be membrane-

associated when it is prenylated and localized in the nucleus when it is unprenylated (Dong *et. al.*, 2002). A similar protein called CaM53 previously found in the petunia also contains an extended C-terminal basic domain and a CAAX motif which are required for efficient prenylation (Rodriguez-Concepcion *et. al.*, 1999). Similar subcellular localization of CaM53 depending on its prenylation state was reported. To locate another possible modification, all proteins were analyzed by the computer program, Myristoylator (Bologna *et. al.*, 2004). As a result, OsCML20 was predicted to contain a potential myristoylation sequence. No other potential myristoylated glycines either terminal or internal were found among the rest of the OsCML proteins. In addition, to determine the possible localization of the OsCML proteins, their sequences were analyzed by targetP (Emanuelsson *et. al.*, 2000). OsCML30 was predicted to contain an endoplasmic reticulum signal sequence and OsCML21 was predicted to be an organellar protein. For OsCaMs and other OsCMLs, no targeting sequence was present, thus, they are probably cytosolic or nuclear proteins

Number and structure of EF hand

The number of EF hands in the rice EF-hand-containing proteins varied from 1 to 4. A summary of the number of proteins having 1, 2, 3, or 4 EF hands is shown in Figure 3.4a. It turned out that among the 243 proteins identified, almost all proteins that contain 4 EF hands were included in our study or are CPK proteins. All five OsCaM proteins have two pairs of EF hands with characteristic residues commonly found in plant CaMs. Consensus sequence of the Ca^{2+}-binding site in the EF hands of plant CaMs compared with OsCaM1, OsCaM2, OsCaM3, vertebrate CaM, and CMD1p from yeast is shown in

(a)



(b)

```
                    EF#1                EF#2
                    * * * * * *         * * * * * *
Plant CaMs          DKDGDGCITTKE        DADGNGTIDFPE
OsCaM1,2,3          ------------        ------------
CaMv                ------------        ------------
CMD1p               ------------        ------------

                    EF#3                EF#4
                    * * * * * *         * * * * * *
Plant CaMs          DKDQNGFISAAE        DVDGDGQINYEE
OsCaM1,2            ------------        ------------
OsCaM3              ------------        ----------D-
CaMv                ---G-Y------        -I-----V----
CMD1p               --NGD-L-----        .S--S-E--IQQ
```

(c)

Residue positions on Ca^{2+}-binding loop												
	1	2	3	4	5	6	7	8	9	10	11	12
G		2		85		113			1		6	
A		21		2			1		2	25	18	
V		6					10	26		10	2	
L		6					15	13		7	4	
I							1	74		2		
P										8	1	
F							18			32		
Y							3			7		
W												
C				1			8		1			
M		2					1			1		
S		9		7	7		8		42	2	9	
T		16					10		27	1	3	
K		19		2			8			2	9	
R		21		3			18			6	1	
H		2								3	4	
D	113		81	2	85				25	2	23	3
E		5		1			5		8	3	28	110
N			32	5	21		1		7		3	
Q		4		5			6			2	2	
	+X		+Y		+Z		-Y		-X			-Z
	D	x	D [N]	G	D [N]	G	x	I [V] [L]	S [T] [D]	x	x	E

Figure 3.4 Characteristics of EF hands in rice proteins. (a) Number of EF-hand-containing proteins containing 1, 2, 3 or 4 EF hands. (b) Residues in the EF hands #1-4 of OsCaMs compared with those of typical plant CaMs, vertebrate CaM (CaMv) and Saccharomyces cerevisiae CaM (CMD1p) using a consensus sequence of plant CaMs as a standard; identical residues in other sequences are indicated by a dash (-), and a gap introduced for alignment purposes is indicated by a dot (.). (c) Residues in Ca^{2+}-binding loops in 32 OsCML proteins shown as the frequency at which an amino acid (shown at the left) is found in each position (shown at the top). The amino acids most frequently found are indicated by bold letters and shown below as a consensus sequence along with the positions of residues serving as Ca^{2+}-binding ligands indicated in Cartesian coordinates. Bracketed residues are alternative residues frequently found in each position and "x" is a variety of amino acids. Residues serving as Ca^{2+}-binding ligands are marked with asterisks (*).

Figure 3.4b. Ca^{2+}-coordinating residues among OsCaMs are invariable with those of the plant CaM consensus sequence. Other residues in the Ca^{2+}-binding loop are also conserved with only the exception of aspartate (D) at residue 11 of the fourth EF hand in OsCaM3. Among the twenty EF-hand motifs of OsCaMs, residues 1(+X) and 3(+Y) are exclusively filled with aspartate (D); residues 5(+Z) are aspartate (D) and asparagine (N); and residues 12(-Z) are glutamate (E) which is invariably found in this position of most Ca^{2+}-binding EF hand motifs. This residue may rotate to give bidentate or monodentate metal ion chelation. Glutamate provides two coordination sites, favoring Ca^{2+} over Mg^{2+} coordination (Cates *et. al.,* 2005). Residues 7(-Y) are usually varied; and residues 9(-X) are aspartate (D), asparagine (N), threonine (T), and serine (S) which are all normally found among plant CaMs.

Schematic diagrams of each protein sequence with the predicted EF hands represented by closed boxes are shown in Figure 3.2. Among all the identified OsCaM and OsCML proteins, about three fourths of the EF hands that exist in pairs (59 pairs) are interrupted by 24 amino acids. The rest are positioned at a similar distance relative to each other which is between 25-29 amino acids with the exception of two pairs that are less than 24 amino acids apart. Most OsCML proteins have either two pairs or at least one pair of identifiable EF hands except OsCML9 which has a single EF hand and OsCML7 which appears to have two separate EF hands. OsCML7 and OsCML9 are interesting because of their high amino acid identities with OsCaM1 (47.7% and 46.1%) but they possess only 2 and 1 EF hands; and have relatively low methionine (M) content (2.8% and 3.2%) compared with other OsCML proteins, respectively. In addition, 10 OsCML proteins with one pair of identifiable EF hands have an extra EF hand that does

not pair with any other motif. Pairing of EF-hand motifs in the CaM molecule helps increase its affinity for Ca^{2+}, therefore an unpaired EF hand in these proteins may bind Ca^{2+} with a lower affinity, or may be non-functional.

Ligands for Ca^{2+} coordination in the EF-hand motifs of OsCML proteins are highly conserved. One hundred and thirteen Ca^{2+}-binding sequences were aligned and the frequency at which amino acids were found is tabulated in Figure 3.4c. Most residues in the Ca^{2+}-binding loops are conserved among OsCML proteins, thus suggesting that most of them are functional EF hands. Similar to OsCaMs, residues 1(+X) are exclusively filled with aspartate (D); and residues 3(+Y) and 5(+Z) are usually aspartate (D) or asparagine (N). Even though they are not coordinating residues, glycine (G) at position 6 is absolutely conserved and hydrophobic residues (I, V, or L) are always found at position 8 in all 133 EF hands in OsCaM and OsCML proteins. Residues 12(-Z) are mostly glutamate (E) with the exceptions of an EF hand in OsCML7, OsCML8, and OsCML13 which have aspartate (D) instead. While OsCML8 and OsCML13 have two pairs of EF-hand motifs, OsCML7 possess two separate EF hands with D at residue 12 in the EF-hand motif at the carboxyl terminus. Cates and colleagues (Cates et. al., 2005), previously reported that mutation of E12 to D reduced the affinity of EF hands for Ca^{2+} in parvalbumin by 100-fold and raised the affinity for Mg^{2+} by 10-fold. It is likely that these EF hands bind Mg^{2+}rather than Ca^{2+} but the physiological significance of Mg^{2+}-binding CaM-like activity is still not known.

Cam and _CML_ gene structures and chromosomal distribution

The structures of the *OsCam* and *OsCML* genes were mapped by comparing their full length cDNAs with the corresponding genomic DNA sequences. In cases where no full length cDNA was available, partial cDNA and EST sequences were used. Their results were compared and verified with the annotation at the TIGR database. Out of 37 *OsCam* and *OsCML* genes, 13 genes contain intron(s) in their coding regions in which none of these is found in group 5 and 6 members. It should be mentioned that by TIGR annotation *OsCam1-2* and *OsCML1* genes were shown to have an alternatively spliced mRNA that encodes a slightly different protein with little supporting evidence so they were eliminated from our list. Schematic diagrams depicting intron-exon structures of the intron-containing genes are shown in Figure 3.5. All *OsCam* genes contain a single intron which interrupts their coding regions within the codon encoding Gly26, a typical rearrangement of all plant *Cam* genes. Interestingly, all of the intron-containing *OsCML* genes are also interrupted by an intron at the same location as *OsCam* genes. The conservation of this intron position indicates their close relationships which is consistent with the fact that these genes encode members of the CML proteins groups 1-4, closely-related CaM-like proteins to OsCaMs. *OsCML1*, *OsCML2*, and *OsCML3* genes contain an additional intron that resides at the codon corresponding to the last residue of genes encoding conventional CaMs. These proteins have an extended C-terminal basic domain and a putative prenylation site. The position of these introns reflects the separation of functional domains within these proteins and suggests that the sequences encoding their carboxyl extensions arose later in the evolution by the fusion of existing *Cam* genes to the additional exons. Similarly, *OsCML8* and *OsCML13* which encode group 3 proteins

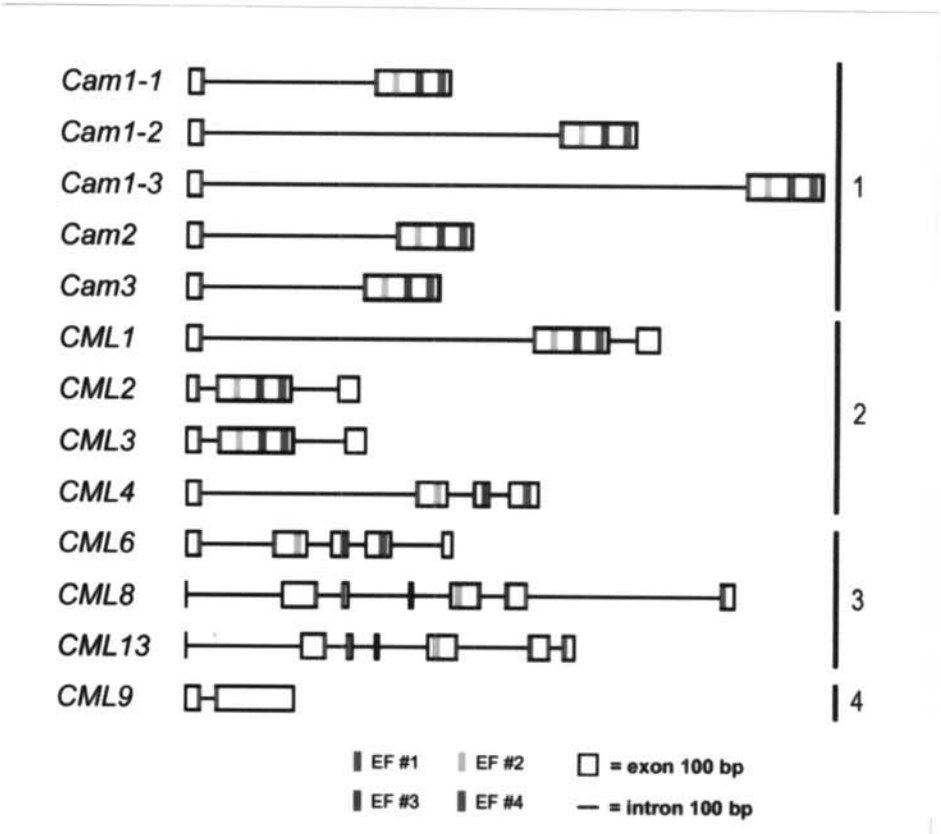


Figure 3.5 Schematic representation of the *OsCam* and *OsCML* genes. Boxes represent exons and lines represent introns. EF-hand motif #1, #2, #3, and #4 are represented by green, yellow, blue and red stripes at their positions, respectively. Groupings of the genes are shown on the right.

have the same gene structure which is the same intron number (6) and location. The gene duplication event that led to the existence of *OsCML8* and *OsCML13* is also supported by the high degree of amino acid identity (60%) between OsCML8 and OsCML13. In these proteins, one of the six introns locates within the sequence encoding the third EF-hand motif, a location comparable to Gly26 of the first EF-hand motif. This intron is probably the remnant of a duplication event that originally gave rise to two EF-hand pairs in these proteins. Interestingly, OsCML8 and OsCML13 are two out of only three OsCMLs that contain aspartate (D) at residues 12(-Z). These observations suggest that the mutation of E12 to D in OsCML8 and OsCML13 probably occurred before the duplication event that led to their existence.

The chromosomal location of each gene was determined from the annotation at the TIGR database. *OsCam* and *OsCML* genes were found distributed across 11 chromosomes of rice as shown in Figure 3.6 with chromosome 1 having the most numbers (10) of genes. *OsCam1-1* was mapped in chromosome 3, *OsCam1-2* in chromosome 7; *OsCam1-3*, and *OsCam3* in chromosome 1; and *OsCam2* in chromosome 5. Their nucleotide sequences share between 86-90 % identities which are lower than their amino acid identities (98-100%). Multiple *OsCam* genes encoding nearly identical proteins have been maintained through natural selection suggesting the functional significance of each gene. *OsCam1-1* and *OsCam1-2* which encode identical proteins were mapped to the duplicated regions of chromosome 3 and 7, respectively. *OsCam1-1* and *OsCam2* were also located within duplicated genome segments of their respective chromosomes. These observations suggest that these pairs of genes are derived from segmental duplication. In addition, there are many pairs/groups of *OsCML* genes which

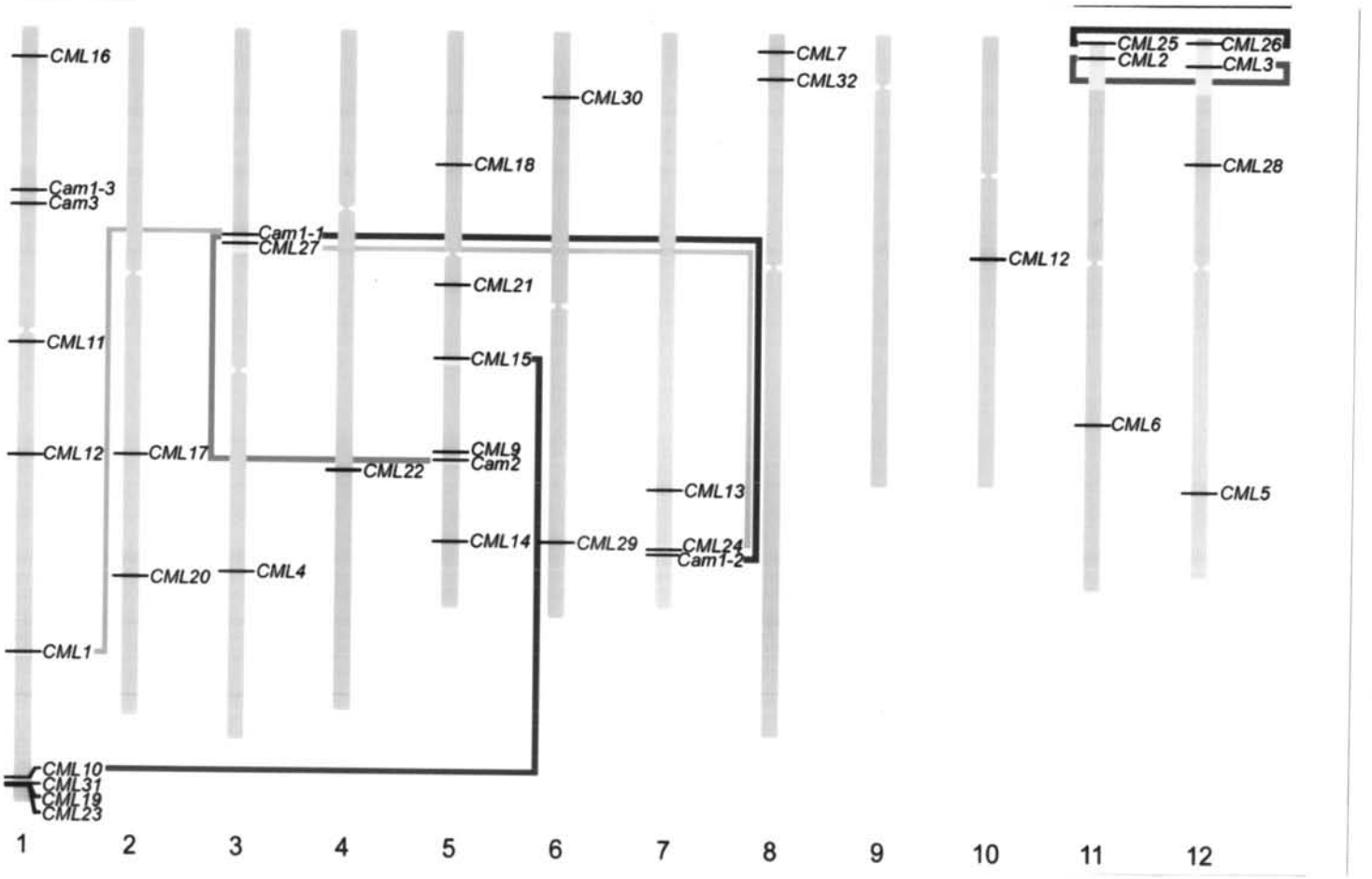


Figure 3.6 Chromosomal distribution of the *OsCam* and *OsCML* genes in the rice genome. The chromosome numbers are shown at the bottom; horizontal lines represent the respective genes; and the centromeric regions appear constricted. Regions of the predicted segmental duplications are indicated by grey sections in the chromosomes and lines connecting the affected loci.

encode proteins that share a high degree of amino acid identity (≥ 60%). OsCML2/OsCML3 (98.9% identical) and OsCML25/OsCML26 (100% identical) are the most closely related pairs. *OsCML2* and *OsCML3* encode potential Ca^{2+}-binding proteins in group 2 with an absolute conservation of the C-terminal sequences that contain a prenylation site (CTIL). *OsCML2* and *OsCML25*; and *OsCML3* and *OsCML26* were mapped to the recently duplicated regions of chromosomes 11 and 12, respectively. Therefore, *OsCML2/OsCML3*; and *OsCML25/OsCML26* may have arisen through the segmental duplication event. Other pairs/groups of closely related CaM-like genes that are likely to be derived from gene duplication events are *OsCML1/OsCam1-1*; *OsCML10/OsCML15*; *OsCML24/OsCML27*; and *OsCML19/OsCML23/OsCML31*. All members in each pair or group have the same number and positions of EF-hand motifs. The positions of predicted segmental duplication according to the analyses by TIGR are illustrated along with the chromosomal locations of the affected genes in Figure 3.6. Conversely, *OsCML19*, *OsCML23* and *OsCML31* are arranged in tandem orientation on chromosome 1 suggesting that they were derived from tandem duplication. Interestingly, *OsCML27* is adjacent to *OsCam1-1* on chromosome 3 and its duplicated gene, *OsCML24*, resides in tandem with *OsCam1-2* (OsCaM1-1 and OsCaM1-2 are 100% identical). Therefore, a local duplication followed by a segmental duplication possibly occurred.

Comparative analysis of rice and Arabidopsis *Cam* and *CML* genes

The full-length amino acid sequences of rice CaMs and CMLs and Arabidopsis CaMs and CMLs were subjected to phylogenetic analysis. Tree construction using the

neighbor-joining method and bootstrap analysis was performed with ClustalX. In Arabidopsis by the neighbor joining tree based on amino acid similarities, McCormack and Braam (McCormack and Braam, 2003) divided CaMs and CMLs into 9 groups. I found that several rice CaMs and CMLs shared high levels of similarity with Arabidopsis CaMs and CMLs and displayed relationships among the family members similar to those previously reported in Arabidopsis as shown in Figure 3.7. All of OsCaM proteins in Arabidopsis and rice are highly conserved (sharing 96.6%-99.3% identity). Interestingly, both Arabidopsis and rice have three *OsCam* genes that encode identical proteins (*ACaM2*, *3*, *5* and *OsCam1-1*, *1-2*, *1-3*). Rice CMLs groups 2, 3, 4, and 5 proteins were closely related to Arabidopsis CMLs group 2, 5, 3, and 4, respectively. The more divergent rice CMLs groups 6a to 6e are also distributed among members of Arabidopsis CML groups 6, 7, 8, 6, and 9, respectively. Apparently, groups 1 from both species are embedded in groups 2. These resulted from the arbitrary separation of groups 1 (CaMs) even though group 2 members share very high degrees of identity (at least 50%) with group 1 proteins. Because what defines a "true" CaM and distinguishes it from a CaM-like protein that serves a distinct role *in vivo* is still unknown, therefore at the moment, only members that share extremely high degrees of identity (>97%) were grouped together to emphasize that they were considered and are possible "true" CaMs.

Based on amino acid sequence alignments (data not shown), many of OsCMLs have putative homologues in Arabidopsis. In group 2, OsCML4 which shares a high level of identity with AtCML8 and AtCML11 has the same number (3) and locations of introns except that AtCML11 lacks the first intron. Similarly, *AtCML19* and *AtCML20* which share a high level of identity with their homologues (*OsCML8* and *OsCML13* in

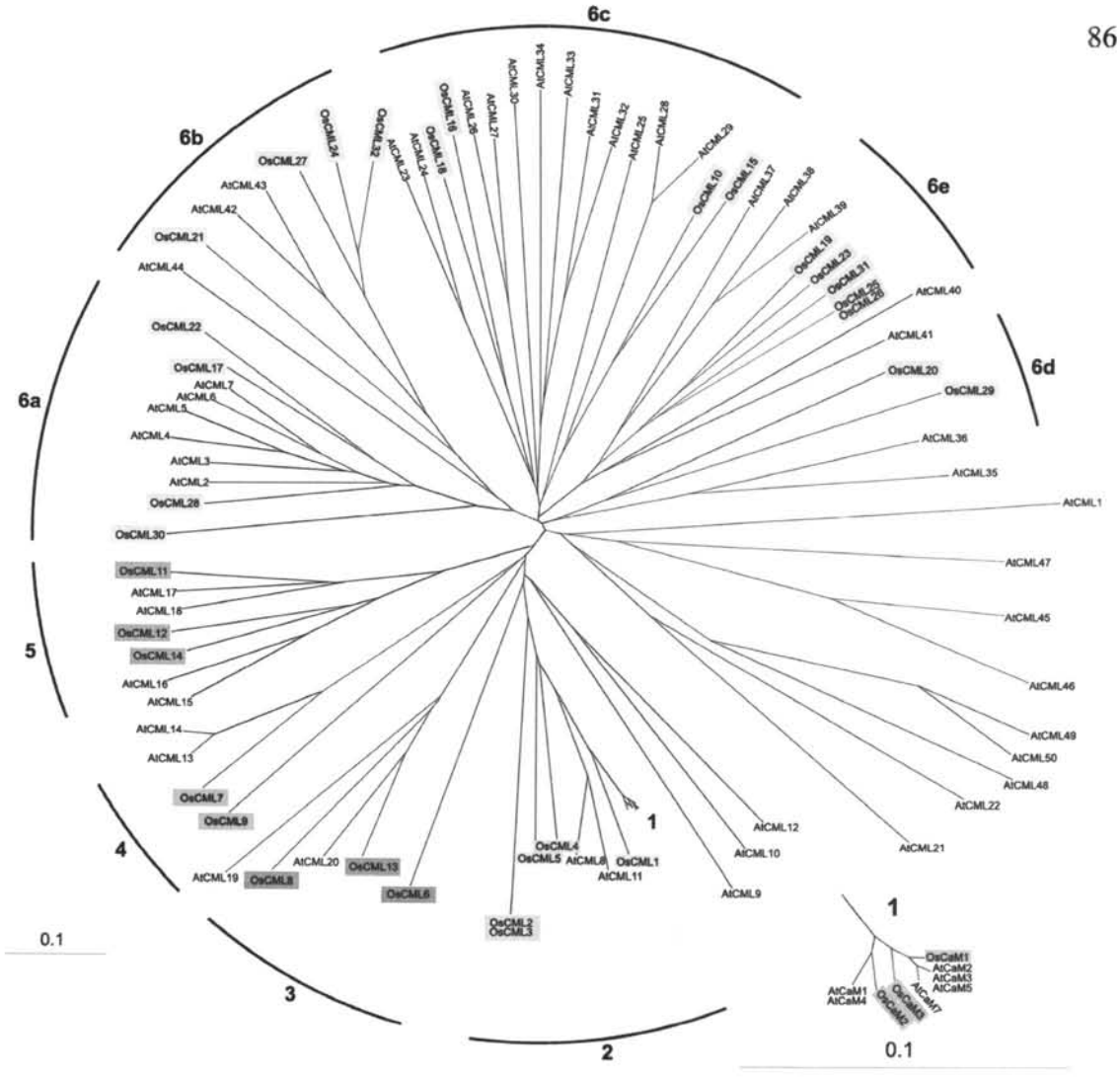


Figure 3.7 Phylogenetic relationships among rice and Arabidopsis CaM and CML proteins. Tree construction using the neighbor-joining method and bootstrap analysis was performed with ClustalX based on the amino acid similarities among the proteins. Rice protein names are highlighted with colours representing each group as used in Figure 3.2 for clarity and groupings of OsCaM and OsCML proteins are indicated accordingly. OsCaM (group 1) and AtCaM portion of the tree was expanded and shown in the bottom right corner.

group 3) have a similar gene structure which is the conservation of five out of the six introns present in their rice counterparts. Interestingly, AtCML19/20 and OsCML8/13 proteins have aspartate (D) at residues 12(-Z) in one of their EF hands, though not on the same hand. AtCML13 and AtCML14, which were thought to have a common progenitor, have very high level of identity (74.3% and 70.9%) with group 4 OsCML7 and all have the mutation of E12 to D in an EF hand corresponding to the third EF hand position. However, OsCML7 has lost an EF hand corresponding to the second position while a second E12 to D mutation was found in AtCML13 and AtCML14. Therefore, similar to AtCML13 and AtCML14, OsCML7 has only one EF hand with canonical amino acids which may result in an impaired ability to bind Ca^{2+}. In OsCML group 5, OsCML11 is closely similar to AtCML17 and AtCML18 and, interestingly all have a relatively low percentage of methionine (M) compared with other CML proteins that share similar levels of identity with CaMs. OsCML11 has only 1.4% methionine content which suggests that its mode of action upon Ca^{2+} binding is probably different from the hydrophobic surface exposure upon conformational changes of CaM.

Previous reports identified 250 EF-hand-containing proteins from the Arabidopsis genome (Day *et. al.,* 2002). Seven loci were defined as *Cam* genes and 50 additional genes were *CML* genes (Hrabak *et. al.,* 2003). Here, I identified 243 EF-hand-containing proteins, five *Cam* genes and 32 *CML* genes. Fewer members of rice CMLs were identified and several Arabidopsis CMLs did not fall into any group of the rice proteins probably because rice OsCML proteins I included in these analyses had at least 25% identity with typical CaMs compared to 16% in Arabidopsis by McCormack and Braam (2003). I noticed that all of the Arabidopsis proteins that did not fall into any group of the

rice proteins shared only 16-30 % identity with typical CaMs. Therefore, both plants appear to have more or less similar numbers of EF-hand-containing and CaM-like proteins. Both also have similar numbers of *CPK* (34 in Arabidopsis and 29 in rice) and *CBL* genes (10 in both Arabidopsis and rice) (Kolukisaoglu *et. al.*, 2004), (Perera and Zielinski, 1992). However, one strikingly different characteristic that I observed is the three OsCML proteins (OsCML1, OsCML2, and OsCML3) that have the carboxyl-terminal CAAX motif for prenylation but none was found in CMLs from Arabidopsis (McCormack and Braam, 2003). It would be interesting to know what functions these rice proteins possess and how the prenylation state affects their activity.

Cam and *CML* expression

Because the presence of cDNA or EST clones indicates expression of the corresponding genes, I performed searches against the cDNA/EST rice databases. The searches revealed that majority of the *OsCam* and *OsCML* genes have corresponding cDNA or EST clones. I have identified all the EST clones for each of the *OsCam* and *OsCML* genes. Characteristics of their expression can be inferred according to which libraries the EST clones were derived from. A summary of the numbers of EST clones found in different organs is presented in Table 3.3. Based on the availability of their EST clones, most *OsCam* and *OsCML* genes are expressed. Some *OsCML* genes are highly expressed in specific organs compared with other genes such as *OsCML13* and *OsCML18* in floral tissues. No cDNA or EST clone is available for *OsCML6*, *OsCML19*, *OsCML23*, and *OsCML25*. However, it is not conclusive that these genes do not express relying solely on the absence of their EST clones. Nonetheless, the availability of EST clones for

Table 3.3 ESTs showing *OsCam* and *OsCML* expression in different tissues.

Cam/CML name	Number of ESTs identified							*Cam/CML* name	Number of ESTs identified						
	leaf	root	panicle	seed	callus	others	Total		leaf	root	panicle	seed	callus	others	Total
OsCam1-1	106	17	49	6	14	89	281	*OsCML15*	5	1	8	-	3	37	54
OsCam1-2	34	13	17	7	11	38	120	*OsCML16*	20	8	12	2	1	43	86
OsCam1-3	21	5	9	-	9	15	59	*OsCML17*	1	-	10	-	-	6	17
OsCam2	35	7	25	5	17	36	125	*OsCML18*	11	2	88	-	1	17	119
OsCam3	57	5	30	6	23	29	150	*OsCML19*	-	-	-	-	-	-	0
OsCML1	30	-	4	1	2	6	43	*OsCML20*	-	-	7	-	-	2	9
OsCML2	-	-	-	-	-	1	1	*OsCML21*	-	2	-	-	-	-	2
OsCML3	61	-	26	6	6	22	121	*OsCML22*	4	-	35	-	2	9	50
OsCML4	20	5	2	2	2	7	38	*OsCML23*	-	-	-	-	-	-	0
OsCML5	-	-	-	-	1	2	3	*OsCML24*	10	4	1	-	-	16	31
OsCML6	-	-	-	-	-	-	0	*OsCML25*	-	-	-	-	-	-	0
OsCML7	18	-	6	4	3	33	64	*OsCML26*	1	-	1	-	-	1	3
OsCML8	13	1	4	-	1	4	23	*OsCML27*	21	2	7	3	5	30	68
OsCML9	1	-	-	-	-	3	4	*OsCML28*	-	-	1	-	-	2	3
OsCML10	16	7	19	-	1	30	73	*OsCML29*	6	2	1	-	-	4	13
OsCML11	7	-	1	-	2	-	10	*OsCML30*	10	13	7	2	2	21	55
OsCML12	3	1	1	-	-	1	6	*OsCML31*	11	5	5	1	5	19	46
OsCML13	13	4	109	-	4	12	142	*OsCML32*	-	-	2	-	-	2	4
OsCML14	-	-	-	-	2	10	12								

the rest of the *OsCam* and *OsCML* genes indicate that they are expressed and indeed are functional genes.

Because five *OsCam* genes encode only three different proteins, whether or not they have different physiological functions is an interesting question. Here, I experimentally determined whether the expression of each of the *OsCam* genes is restricted to specific organs. Total RNA was isolated from the leaves, roots, flowers, immature seeds and calli of rice plants as shown in Figure 3.8 and used to perform reverse transcription and PCR amplification reactions. Primers selected by computer analysis of the cDNA and EST sequences corresponding to these genes are given in the "Materials and Methods" section. A control RT-PCR reaction without adding reverse transcriptase was done in parallel with each experimental reaction to ensure that the product obtained could be attributed to the product of the reverse transcriptase reaction. Figure 3.9 shows that bands of the expected sizes based on each of the gene sequences (698, 526, 551, 201, 520 and 591 base pairs for *OsCam1-2*, *OsCam1-2*, *OsCam1-3*, *OsCam2*, *OsCam3*, and *OsCML1* respectively) were detected in all organs or tissues examined including the leaves and roots of 2-week old seedlings, mature leaves, flowers, immature seeds and calli. No band was detected in the control RT-PCR reactions. It should be noted that the RT-PCR conditions used in this study did not allow quantitative determination, therefore comparison of the expression levels among different organs or different genes can not be made. Nevertheless, it can be concluded that all of *OsCam* genes were expressed in all organs that I examined.

The expression of closely related *Cam* genes in a single organ was not surprising. Several similar occurrences in other plant species have been reported. In tobacco, all 13

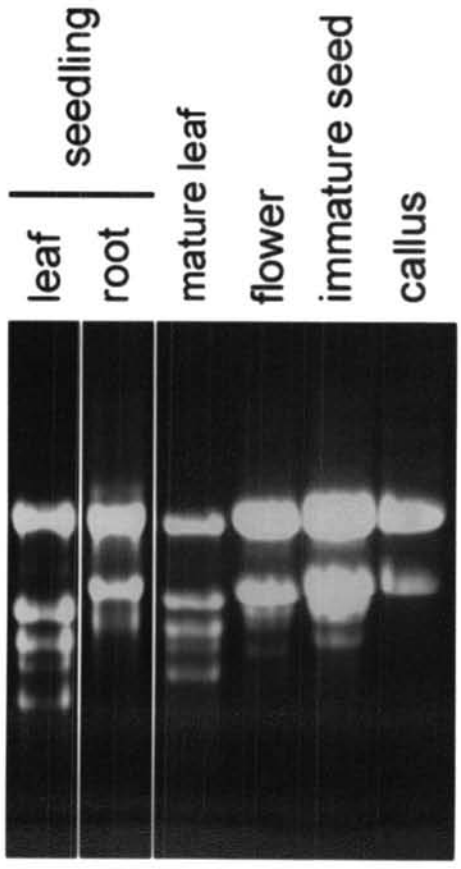


Figure 3.8 The total RNA extracted from different organs of rice plants electrophoresed on a 1.5% formaldehyde agarose gel and visualized by ethidium bromide staining.

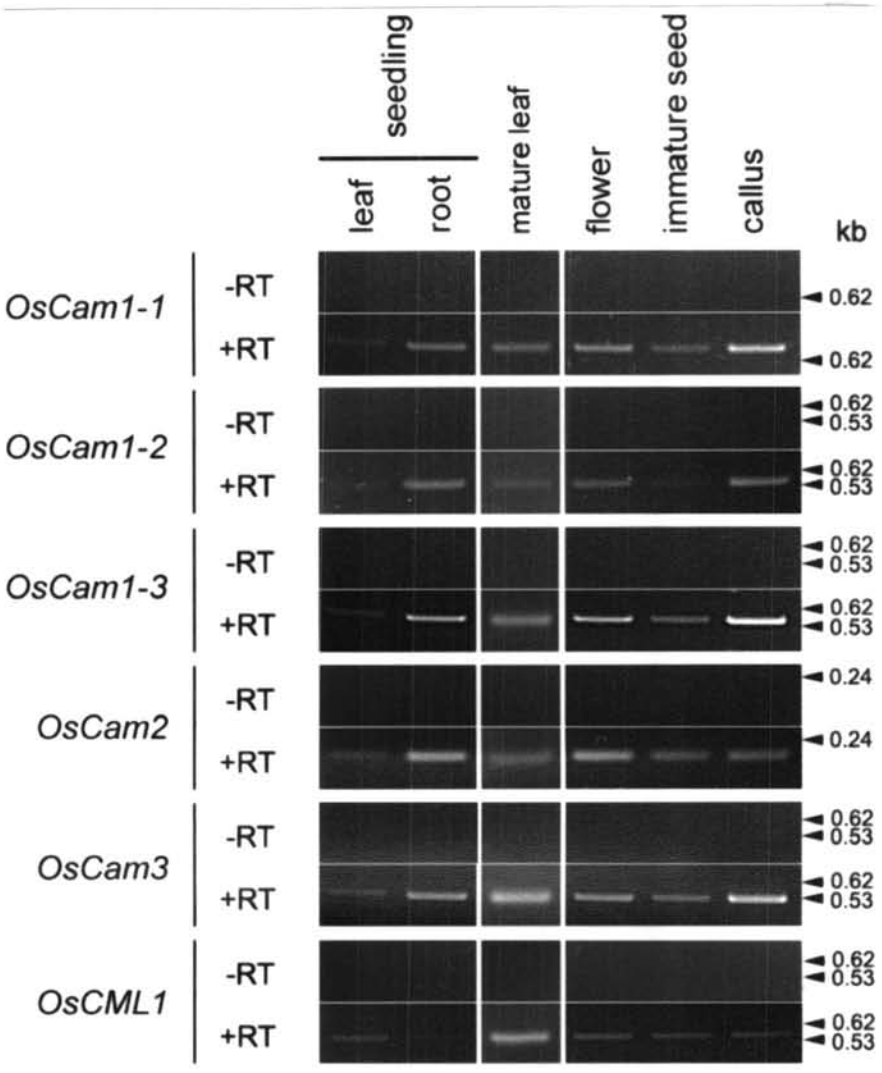


Figure 3.9 Expression patterns of the *OsCam* and *OsCML1* genes. The total RNA isolated from organs indicated was used in RT-PCR assays either without (-RT) or with (+RT) the addition of M-MLV reverse transcriptase. The cDNAs were amplified by PCR using gene-specific primers in the "Materials and Methods" section. The products derived from 250 ng of total RNA were separated in agarose gels and visualized by ethidium bromide staining. The sizes of DNA markers in base pairs are shown on the right.

Cam closely related genes were expressed in almost all organs examined with a few exceptions, notably *NtCam13*, which was exclusively expressed in the root (Yamakawa *et. al.*, 2001). However, *NtCam13* encodes a protein of less than 80% identity to typical plant CaMs. Similarly, *ACam1-ACam5* genes which encode nearly identical proteins were all expressed in the leaves and siliques of Arabidopsis (Gawienowski *et. al.,* 1993, Poohvaiah *et. al.,* 1993). While *Cam* expression is ubiquitous among different cells, protein concentrations may vary in specific cell types. Immunolocalization studies have shown that root cap cells and meristematic zones have increased CaM accumulation (Takezawa *et. al.,* 1995). In addition, levels of steady state transcripts of *Cam* genes have been reported to be modulated at different developmental stages (Choi *et. al.,* 1996, Van der Luit *et. al.,* 1999) and in response to external stimuli such as salinity, wind, cold, wounding and pathogen attack (Yamakawa *et. al.*, 2001, Delumeau *et. al.*,2002 , Duval *et. al.,* 2002, Kawasaki *et. al.,* 2001). *OsCam1-1* was shown to be rapidly and strongly increased in leaves under osmotic stress (Phean-o-pas *et. al.,* 2005). Modulation of expression in specific organs of a CaM isoform possibly serves its roles in a timely fashion.

Cloning of *Cam* and *CML* genes

The cDNA prepared from leaves of *Oryza sativa* L. (KDML105) seedlings was used as template for PCR amplification of the *Cam* and *CML* genes. Primers as given in the "Materials and Methods" section were selected by computer analysis of the cDNA and EST sequences of these genes. PCR amplification using these primers generated products of the coding sequence of *OsCam1-1* (*OsCam1-1i-CD*), the 3'UTR of *OsCam1-*

1 (*OsCam1-1i-3'*) and the coding sequence of *OsCML1* (*OsCML1i-CD*) which are of approximately 0.5, 0.2 and 0.6 kb, respectively as shown by agarose gel electrophoresis in Figure 3.10. Each DNA fragment was purified using the QIAquick gel extraction kit.

The purified PCR products were ligated into pGEM®-T vector by ligase overnight and transformed into the competent *E. coli* DH5α cells. The transformants were selected by blue-white colony screening on ampicillin agar plates containing X-gal and IPTG. White colonies were randomly selected and cultured in 1 ml LB broth containing 100 μg/ml of ampicillin at 37 °C overnight and the cultures were subjected to plasmid extraction. To verify the insertion of PCR products into pGEM®-T, the recombinant plasmids containing *OsCam1-1i-CD, OsCam1-1i-3'* and *OsCML1i-CD* were digested with *Nco*I and *Spe*I at 37 °C overnight. Subsequently, these reactions were analyzed by 1.5% agarose gel electrophoresis. The results showed that the inserted DNA fragments of *OsCam1-1i-CD, OsCam1-1i-3'* and *OsCML1i-CD* were approximately 0.5, 0.2 and 0.6 kb in length, respectively as shown in Figure 3.11 – 3.13.

To confirm whether the inserted fragments were indeed *Cam* and *CML* genes, the recombinant plasmids were subjected to DNA sequencing. DNA sequences of *OsCam1-1i-CD, OsCam1-1i-3'* and *OsCML1i-CD* are shown in Figure 3.14. The resulting *Cam* and *CML* DNA sequences of *Oryza sativa* cv. KDML105, which is an indica subspecies were compared with those of *Oryza sativa* cv. Nipponbare, which is a japonica subspecies using the EMBOSS Pairwise Alignment Algorithms. The results revealed that the *OsCam1-1i-CD* sequence shared 99.9% identity with the coding sequence of *OsCam1-1* from the japonica rice (*OsCam1-1j*) as shown in Figure 3.15. The only difference in nucleotide is at position 408 which guanine base (G) in *OsCam1-1i-CD* (A)

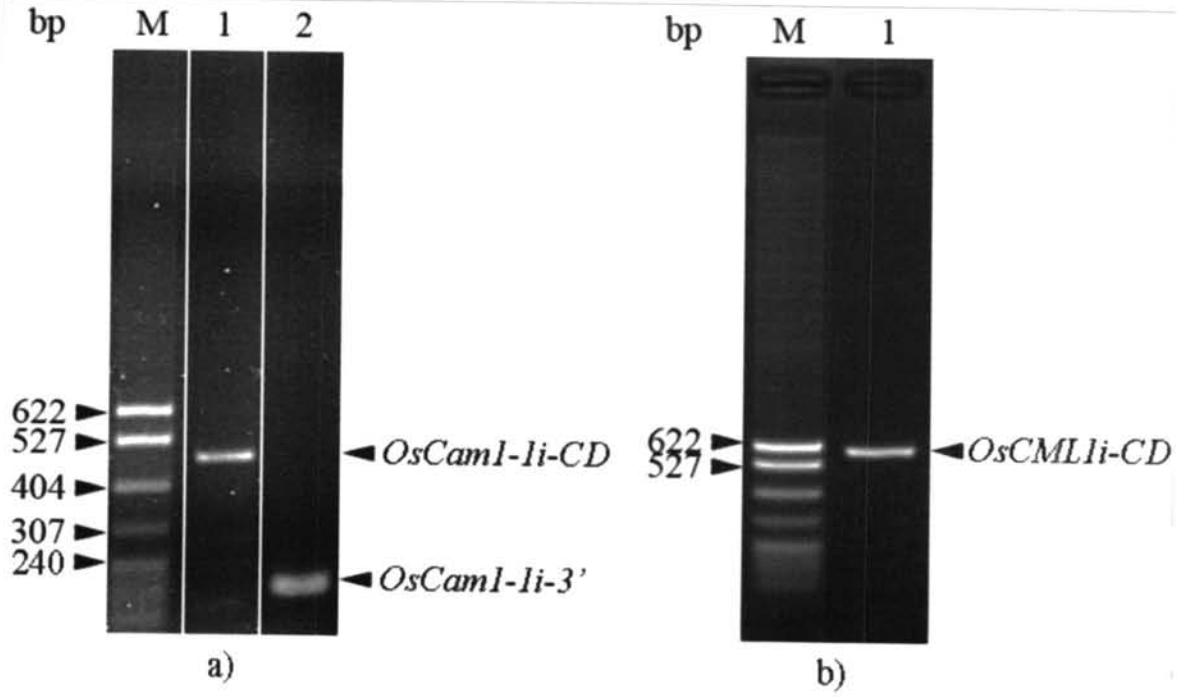


Figure 3.10 Agarose gel electrophoresis of the amplified coding sequence of *OsCam1-1 (OsCam1-1i-CD)*, 3'UTR of *OsCam1-1 (OsCam1-1i-3')* and coding sequence of *OsCML1 (OsCML1i-CD)*. The PCR products were separated on 2% agarose gels and visualized by ethidium bromide staining.

a) Lane M DNA Marker: pBR322/*Msp*I

Lane 1 PCR product of *OsCam1-1i-CD*

Lane 2 PCR product of *OsCam1-1i-3'*

b) Lane M DNA Marker: pBR322/*Msp*I

Lane 1 PCR product of *OsCML1i-CD*

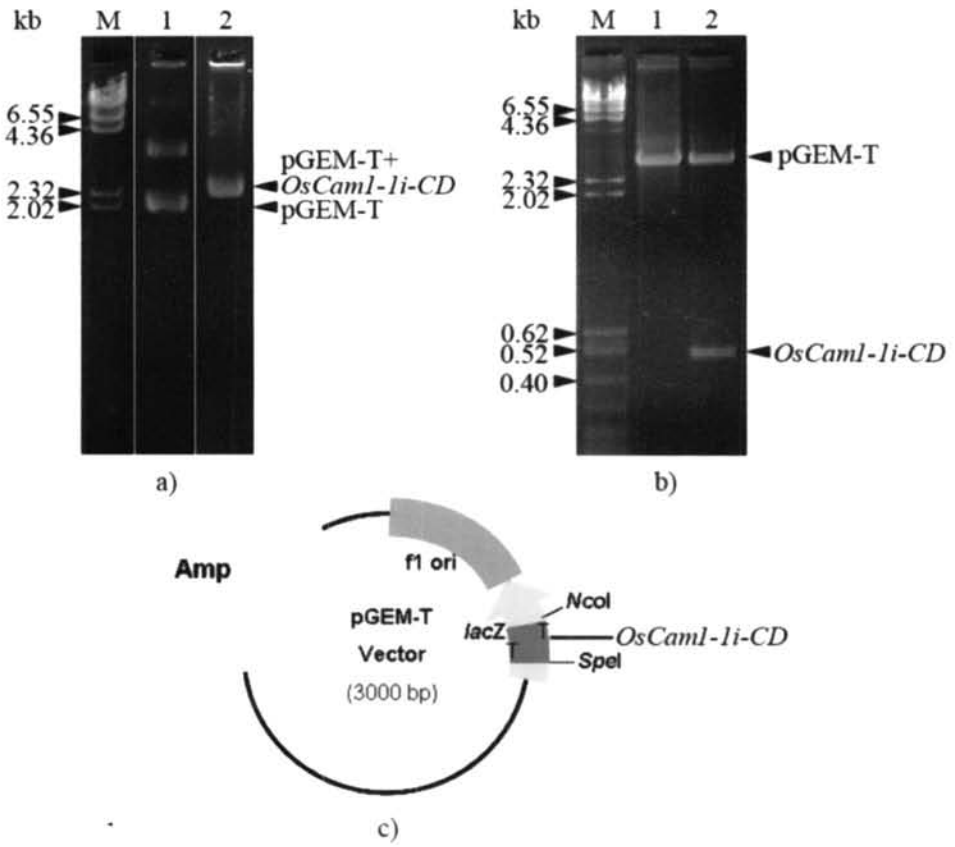


Figure 3.11 Agarose gel electrophoresis of the recombinant *OsCam1-1i-CD* gene inserted in pGEM®-T vector (pGEM-T+*OsCam1-1i-CD*). The DNA was separated on 1.5% agarose gels and visualized by ethidium bromide staining.

a) Lane M DNA Marker: Lamda/*Hind* III

Lane 1 undigested pGEM-T

Lane 2 undigested pGEM-T+ *OsCam1-1i-CD*

b) Lane M DNA Marker: pBR322/*Msp*I

Lane 1 pGEM-T digested with *Nco*I and *Spe*I

Lane 2 pGEM-T+ *OsCam1-1i-CD* digested with *Nco*I and *Spe*I

c) Plasmid map of the recombinant *OsCam1-1i-CD* gene in pGEM®-T

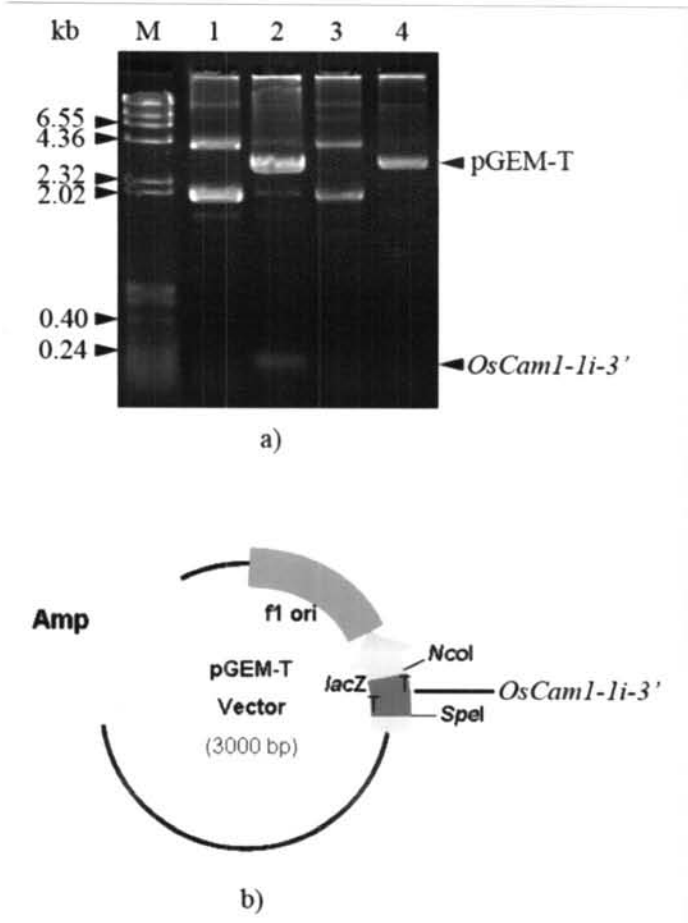


Figure 3.12 Agarose gel electrophoresis of the recombinant *OsCam1-1i-3'* gene inserted in pGEM®-T vector (pGEM-T+*OsCam1-1i-3'*). The DNA was separated on a 1.5% agarose gel and visualized by ethidium bromide staining.

a) Lane M DNA Marker: Lamda/*Hind* III

Lane 1 undigested pGEM-T+ *OsCam1-1i-3'*

Lane 2 pGEM-T+ *OsCam1-1i-3'* digested with *Nco*I and *Spe*I

Lane 3 undigested pGEM-T

Lane 4 pGEM-T digested with *Nco*I and *Spe*I

b) Plasmid map of the recombinant *OsCam1-1i-3'* gene in pGEM®-T

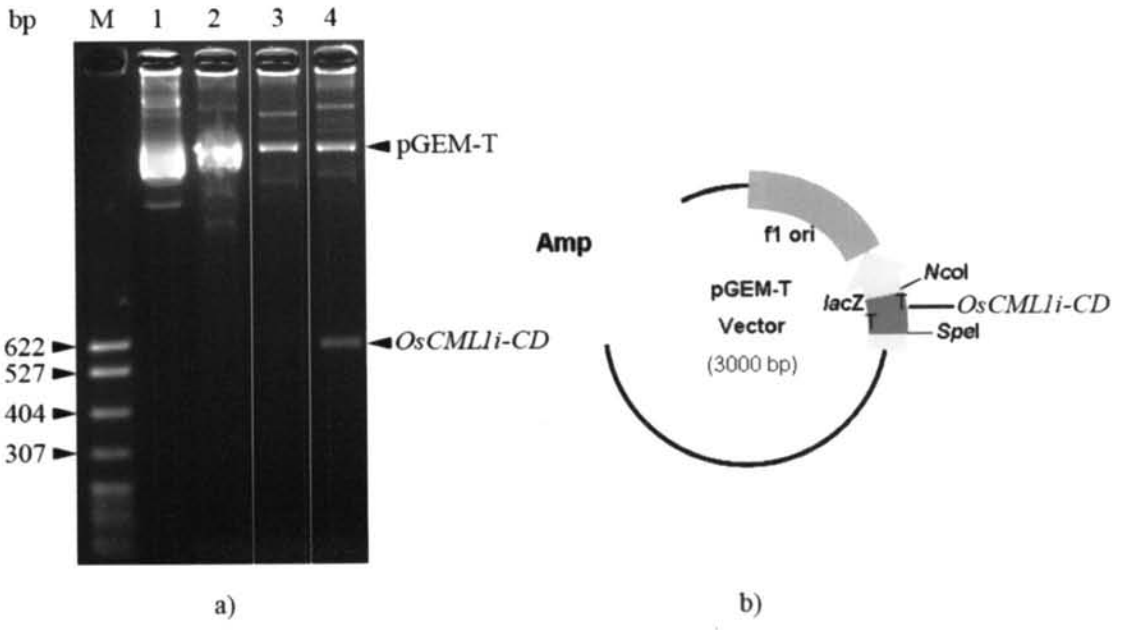


Figure 3.13 Agarose gel electrophoresis of the recombinant *OsCML1i-CD* gene inserted in pGEM®-T vector (pGEM-T+*OsCML1i-CD*). The DNA was separated on a 1.5% agarose gel and visualized by ethidium bromide staining.

a) Lane M DNA Marker: pBR322/*Msp*I

Lane 1 undigested pGEM-T

Lane 2 undigested pGEM-T+ *OsCML1i-CD*

Lane 3 pGEM-T digested with *Nco*I and *Spe*I

Lane 4 pGEM-T+ *OsCML1i-CD* digested with *Nco*I and *Spe*I

b) Plasmid map of the recombinant *OsCML1i-CD* gene in pGEM®-T

>OsCam1-1i-CD

GAAGCCAGGCTAAGCCCAGCGGC**ATG**GCGGACCAGCTCACCGACGACCAGATCGCCGAGTTCAAGGAGGCC
TTCAGCCTCTTCGACAAGGACGGCGATGGTTGCATCACAACCAAGGAGCTGGGAACCGTGATGCGTTCGCT
GGGGCAGAACCCAACGGAGGCCGAGCTCCAGGACATGATCAACGAGGTCGACGCGGACGGCAACGGCACCA
TCGACTTCCCGGAGTTCCTCAACCTGATGGCACGCAAGATGAAGGACACCGACTCGGAGGAGGAGCTCAAG
GAGGCGTTCAGGGTGTTCGACAAAGACCAGAACGGCTTCATCTCCGCCGCCGAGCTCCGCCACGTCATGAC
CAACCTCGGCGAGAAGCTGACCGACGAGGAGGTCGACGAGATGATCCGCGAGGCCGACGTCGACGGTGACG
GCCAGATCAACTACGAGGAGTTCGTCAAGGTCATGATGGCCAAG**TGA**GGCACCACTTCCCCTGCCGATGAT
G

>OsCam1-1i-3'

CTGCCGATGATGGCATAGTACCCTGGGAGGAGGAAACCGTGCATTGCCGTATTAGTAAGGGGATGCAAACA
CTGGTTTCAGTCGTCTTCCCTGATGAAGAAAACCGAACCGTACTAGTTGTAGTTGCTGAACATTTTTCTAT
CTCTCCAGTCTCTCCTGTGTGCCATGGAACTTCTTGCTTGATTTTTCTGTGTGAATCTGTTAAGGCTTGC

>OsCML1i-CD

GCTTTGCTCGCCTTCTCGAAGCTTCTGCTGCC**ATG**GCGGACCAGCTCTCCGAAGAGCAGATTGTAGAGTTC
AGGGAGGCCTTCAGCCTCTTCGACAAGGACGGCGACGGTTCTATCACCACCAAGGAGCTAGGAACCGTGAT
GCGAAGTCTGGGGCAGAACCCAACAGAGGCGGAGCTGCAGGACATGATCAGCGAGGTGGACGCGGACAGCA
ACGGCAACATCGAATTCAAGGAGTTCCTGGGCCTGATGGCGCGCAAGCTGAGGGACAAGGACTCCGAGGAG
GAGCTGAAGGAGGCGTTCCGCGTCTTCGACAAGGACCAGAACGGCTTCATCTCCGCCGCCGAGCTCCGCCA
CGTGATGGCCAACATCGGGGAGCGGCTCACCGACGAGGAGGTCGGCGAGATGATCAGCGAGGCCGACGTCG
ACGGCGACGGGCAGATCAACTACGAGGAGTTCGTCAAGTGCATGATGGCCAAGAAGAGGAGGAAGAGGATA
GAGGAGAAGAGGGAGCACGACGGCGGCAGCAGGACGAAGAGTGCAGGGCCCTCCGCCGCGCCGGCGAGCAA
GCGTGGCCAGAAGTGCGTGATCC

Figure 3.14 DNA sequences of the coding region of *OsCam1-1 (OsCam1-1i-CD)*, 3'UTR of *OsCam1-1 (OsCam1-1i-3')* and the coding sequence of *OsCML1 (OsCML1i-CD)* gene. The underlined letters represent primer binding sites and the bold letters indicate start and stop codons.

```
OsCam1-1j         1 CATTCTCTCCGCGACGGTCTCGTCTTCCCCACCCCTCGCCTCCTCGCGCG   50
OsCam1-1i-CD      1                                                      0

OsCam1-1j        51 CTCGGTGAGAGAAGCGAAGAAGAAGAAGAAGAGGAGGAGGAAGAAGCCAG  100
OsCam1-1i-CD      1                                                --------   8

OsCam1-1j       101 GCTAAGCCCAGCGGCATGGCGGACCAGCTCACCGACGACCAGATCGCCGA  150
OsCam1-1i-CD      9 ----------------------------------------------------  58

OsCam1-1j       151 GTTCAAGGAGGCCTTCAGCCTCTTCGACAAGGACGGCGATGGTTGCATCA  200
OsCam1-1i-CD     59 ----------------------------------------------------  108

OsCam1-1j       201 CAACCAAGGAGCTGGGAACCGTGATGCGTTCGCTGGGGCAGAACCCAACG  250
OsCam1-1i-CD    109 ----------------------------------------------------  158

OsCam1-1j       251 GAGGCCGAGCTCCAGGACATGATCAACGAGGTCGACGCGGACGGCAACGG  300
OsCam1-1i-CD    159 ----------------------------------------------------  208

OsCam1-1j       301 CACCATCGACTTCCCGGAGTTCCTCAACCTGATGGCACGCAAGATGAAGG  350
OsCam1-1i-CD    209 ----------------------------------------------------  258

OsCam1-1j       351 ACACCGACTCGGAGGAGGAGCTCAAGGAGGCGTTCAGGGTGTTCGACAAA  400
OsCam1-1i-CD    259 ----------------------------------------------------  308

OsCam1-1j       401 GACCAGAACGGCTTCATCTCCGCCGCCGAGCTCCGCCACGTCATGACCAA  450
OsCam1-1i-CD    309 ----------------------------------------------------  358

OsCam1-1j       451 CCTCGGCGAGAAGCTGACCGACGAGGAGGTCGACGAGATGATCCGCGAAG  500
OsCam1-1i-CD    359 ----------------------------------------------G-----  408

OsCam1-1j       501 CCGACGTCGACGGTGACGGCCAGATCAACTACGAGGAGTTCGTCAAGGTC  550
OsCam1-1i-CD    409 ----------------------------------------------------  458

OsCam1-1j       551 ATGATGGCCAAGTGAGGCACCACTTCCCCTGCCGATGATGGCATAGTACC  600
OsCam1-1i-CD    459 ----------------------------------------              498

OsCam1-1j       601 CTGGGAGGAGGAAACCGTGCATTGCCGTATTAGTAAGGGGATGCAAACAC  650
OsCam1-1i-CD    499                                                    498
```

Figure 3.15 Nucleotide sequence alignment of the coding sequence of *OsCam1-1* *OsCam1-1i-CD* (*Oryza sativa* indica) with the *OsCam1-1j (Oryza sativa* japonica) used EMBOSS Pairwise Alignment Algorithms. Identical residues in other sequences are indicated by a dash (-), the underlined letters represent primer binding sites and the bold letters indicate start and stop codons.

is found in place of adenine base in *OsCam1-1j*. Alignment of the deduced amino acid sequences from *OsCam1-1i-CD* and *OsCam1-1j* showed that they are identical (Figure 13.16). In the 3'UTR, the only difference in nucleotide is at position 174 from the stop codon which thymine base (T) is found in *OsCam1-1j* in place of guanine base (G) in *OsCam1-1I* as shown in Figure 3.17. Whereas *OsCML1i-CD* coding region sequence has 8 nucleotide differences from the coding sequence of *OsCML1* from the japonica rice (*OsCML1j*) as shown in Figure 3.18. Alignment of the deduced amino acid sequences from *OsCML1i-CD and OsCML1j* showed that they have 4 amino acid differences as shown in Figure 3.19. The result from sequence alignments of these *Oryza sativa* genes from both subspecies indicated that they are highly conserved especially the *OsCam1-1* gene. Because of the central role of CaM in eukaryotic biology, it is reflected in its conservation (McCormack and Braam, 2003). CaM sequences are strongly conserved across all species: for example, all known vertebrate CaMs are identical in amino acid sequence and share 91% amino acid identity to plants. Plant sequences share 61% identity with those from yeast and sequence conservation among plant and algal species ranges from 84 to 100% (Zielinski, 1998). In Arabidopsis, typical CaM members including CaM1 to CaM7 are highly similar to animal CaMs and to each other (>95% identical in amino acid sequence) (Luan *et al.*, 2002). In soybean, there are five CaM isoforms (SCaM1 to -5). SCaM1, -2 and -3 are highly conserved compared to other plant CaM isoforms including Arabidopsis CaM isoforms (Lee *et al.*, 1995).

```
                                    ---Loop----
                          --helixE--         -helixF--         ---
OsCaM1-1j           1 MADQLTDDQIAEFKEAFSLFDKDGDGCITTKELGTVMRSLGQNPTEAELQ      50
OsCaM1-1i-CD        1 --------------------------------------------------      50

                        ---Loop----                            ---Loop-
                   helixE--         -helixF--          --helixE--
OsCaM1-1j          51 DMINEVDADGNGTIDFPEFLNLMARKMKDTDSEEELKEAFRVFDKDQNGF     100
OsCaM1-1i-CD       51 --------------------------------------------------     100

                      ----                          ---Loop----
                         -helixF--         ---helixE--         -helixF--
OsCaM1-1j         101 ISAAELRHVMTNLGEKLTDEEVDEMIREADVDGDGQINYEEFVKVMMAK     149
OsCaM1-1i-CD      101 --------------------------------------------------     149
```

Figure 3.16 Amino acid sequence alignment of OsCaM1-1i-CD (*Oryza sativa* indica) with OsCaM1-1j *(Oryza sativa* japonica) used EMBOSS Pairwise Alignment Algorithms. Identical residues in other sequences are indicated by a dash (-), the bold letters indicate the Ca^{2+}-binding residues. The positions of helix E, loop and helix F are indicated above their sequences.

```
OsCam1-1j       551 ATGATGGCCAAGTGAGGCACCACTTCCCCTGCCGATGATGGCATAGTACC   600
OsCam1-1i-3'      1                                 ----------------------    22

OsCam1-1j       601 CTGGGAGGAGGAAACCGTGCATTGCCGTATTAGTAAGGGGATGCAAACAC   650
OsCam1-1i-3'     23 --------------------------------------------------    72

OsCam1-1j       651 TGGTTTCAGTCGTCTTCCCTGATGAAGAAAACCGAACCGTACTAGTTGTA   700
OsCam1-1i-3'     73 --------------------------------------------------   122

OsCam1-1j       701 GTTGCTGAACATTTTTCTATCTCTCCAGTCTCTCCGGTGTGCCATGGAAC   750
OsCam1-1i-3'    123 ----------------------------------T---------------   172

OsCam1-1j       751 TTCTTGCTTGATTTTTCTGTGTGAATCTGTTAAGGCTTGCTCTGATCTCT   800
OsCam1-1i-3'    173 ----------------------------------------           212

OsCam1-1j       801 CCGAA    805
OsCam1-1i-3'    213          212
```

Figure 3.17 Nucleotide sequence alignment of *OsCam1-1i-3'* (*Oryza sativa* indica) with the *OsCam1-1j* (*Oryza sativa* japonica) used EMBOSS Pairwise Alignment Algorithms. Identical residues in other sequences are indicated by a dash (-), the underlined letters represent primer binding sites and the bold letters indicate stop codon

```
OsCML1j           1 GACACAGCCCGCGCACCTCCACAGCATTAGCCATCAACGACCAGCATCTC     50
OsCML1i-CD        1                                                       0

OsCML1j          51 AGCTTTGCTCGCCTTCTCGAAGCTTCTGCTGCCATGGCGGACCAGCTCTC    100
OsCML1i-CD        1 ---------------------------------------------------    49

OsCML1j         101 CGAAGAGCAGATTGGAGAGTTCAGGGAGGCCTTCAGCCTCTTCGACAAGG    150
OsCML1i-CD       50 --------------T-----------------------------------    99

OsCML1j         151 ACGGCGACGGTTCTATCACCACCAAGGAGCTAGGAACCGTGATGCGAAGT    200
OsCML1i-CD      100 ---------------------------------------------------   149

OsCML1j         201 CTAGGGCAGAACCCAACGGAGGCGGAGCTGCAGGACATGATCAGCGAGGT    250
OsCML1i-CD      150 --G--------------A-------------------------------   199

OsCML1j         251 GGACACGGACAGCAACGGCAACATCGAATTCAAGGAGTTCCTGGGCCTGA    300
OsCML1i-CD      200 ----G-----------------------------------------------   249

OsCML1j         301 TGGCGCGCAAGCTGAGGGACAAGGACTCCGAGGAGGAGCTGAAGGAGGCA    350
OsCML1i-CD      250 ---------------------------------------------------   299

OsCML1j         351 TTCCGCGTCTTCGACAAGGACCAGAACGGTTTCATCTCTGCCACCGAGCT    400
OsCML1i-CD      300 ----------------------------C--------C---G-------   349

OsCML1j         401 CCGCCACGTGATGGCCAACATCGGGGAGCGGCTCACCGACGAGGAGGTCG    450
OsCML1i-CD      350 ---------------------------------------------------   399

OsCML1j         451 GCGAGATGATCAGCGAGGCCGACGTCGACGGCGACGGGCAGATCAACTAC    500
OsCML1i-CD      400 ---------------------------------------------------   449

OsCML1j         501 GAGGAGTTCGTCAAGTGCATGATGGCCAAGAAGAGGAGGAAGAGGATAGA    550
OsCML1i-CD      450 ---------------------------------------------------   499

OsCML1j         551 GGAGAAGAGGGACCACGACGGCGGCAGCAGGACGAAGAGTGCAGGGCCCT    600
OsCML1i-CD      500 ------------G--------------------------------------   549

OsCML1j         601 CCGCCGCGCCGGCGAGCAAGCGTGGCCAGAAGTGCGTGATCCTGTAATAA    650
OsCML1i-CD      550 ----------------------------------------            591

OsCML1j         651 TTGAGCCAGCACTGAGATTCTCATGAGTCAATGAGCTACACGAATGATGT     700
OsCML1i-CD      592                                                     591
```

Figure 3.18 Nucleotide sequence alignment of *OsCML1i-CD* (*Oryza sativa* indica) with *OsCML1j* (*Oryza sativa* japonica) used EMBOSS Pairwise Alignment Algorithms. Identical residues in other sequences are indicated by a dash (-), the underlined letters represent primer binding sites and the bold letters indicate start and stop codons.

```
                                    ---Loop---
                          --helixE--        -helixF-         --
OsCML1j             1 MADQLSEEQIGEFREAFSLFDKDGDGSITTKELGTVMRSLGQNPTEAELQ     50
OsCML1i-CD          1 ----------V----------------------------------------     50

                        ---Loop---                              ---Loop
                   helixE--        -helixF-          --helixE--
OsCML1j            51 DMISEVDTDSNGNIEFKEFLGLMARKLRDKDSEEELKEAFRVFDKDQNGF    100
OsCML1i-CD         51 -------A----------------------------------------    100

                      ---                          ---Loop---
                         -helixF-         --helixE--        -helixF-
OsCML1j           101 ISATELRHVMANIGERLTDEEVGEMISEADVDGDGQINYEEFVKCMMAKK    150
OsCML1i-CD        101 ---A-----------------------------------------------    150


OsCML1j           151 RRKRIEEKRDHDGGSRTKSAGPSAAPASKRGQKCVI    186
OsCML1i-CD        151 ---------E-------------------------    186
```

Figure 3.19 Amino acid sequence alignment of OsCML1i-CD (*Oryza sativa* indica) with OsCML1j *(Oryza sativa* japonica) used EMBOSS Pairwise Alignment Algorithms. Identical residues in other sequences are indicated by a dash (-), the bold letters indicate the Ca^{2+}-binding residues. The positions of helix E, loop and helix F are indicated above their sequences.

Expression pattern of *Cam* and *CML* gene under salt stress in KDML105

Rice seedlings were hydroponically grown in WP No.2 Solution. Three-week old seedlings were transferred to the growth medium containing 150 mM NaCl and harvested at different time points (0, 0.5, 1, 2 and 4 hours) in liquid nitrogen and stored at -80 °C. Total RNA was isolated from leaves and twenty micrograms of total RNA were fractionated in formaldehyde agarose gels and transferred to charged nylon membranes. RNA gel blot analysis of the five *Cam* and one *CML* mRNAs was carried out to detect their expression in leaves and one of these genes would be selected for further studies on their expression under salt stress and exogenous abscisic acid application.

Even though the coding regions of these genes are highly conserved, the 5' and 3' untranslated regions (UTR) are diverged therefore the probes made from these sequences can be used to differentiate their respective mRNAs. From sequence comparison, while their nucleotide sequences share more than 85% identity in the coding regions, the identity among the sequences in the 3' UTR is relatively low (less than 40%). Therefore, DNA fragments made from the 3' UTR by PCR amplification were used to prepare isoform-specific probes. Figure 3.20 showed that the PCR product sizes (3'UTR) of *OsCam1-1, OsCam1-2, OsCam1-1, OsCam2, OsCam3*and *OsCML1* were similar to their expected sizes of 212, 183, 218, 201, 145 and 238 bp, respectively. The PCR products were incorporated with [α-^{32}P]dCTP and used for hybridization which was carried out at 40°C for 16 hours. The blots were washed in 2X SSPE, 0.1% SDS twice and then once in 1X SSPE, 0.1% SDS at room temperature. For *OsCam2* and *OsCML1*, the blots were further washed in 0.1X SSPE, 0.1% SDS at 45°C. Positive hybridizing bands were detected by autoradiography. It should be noted that these probes were made of PCR

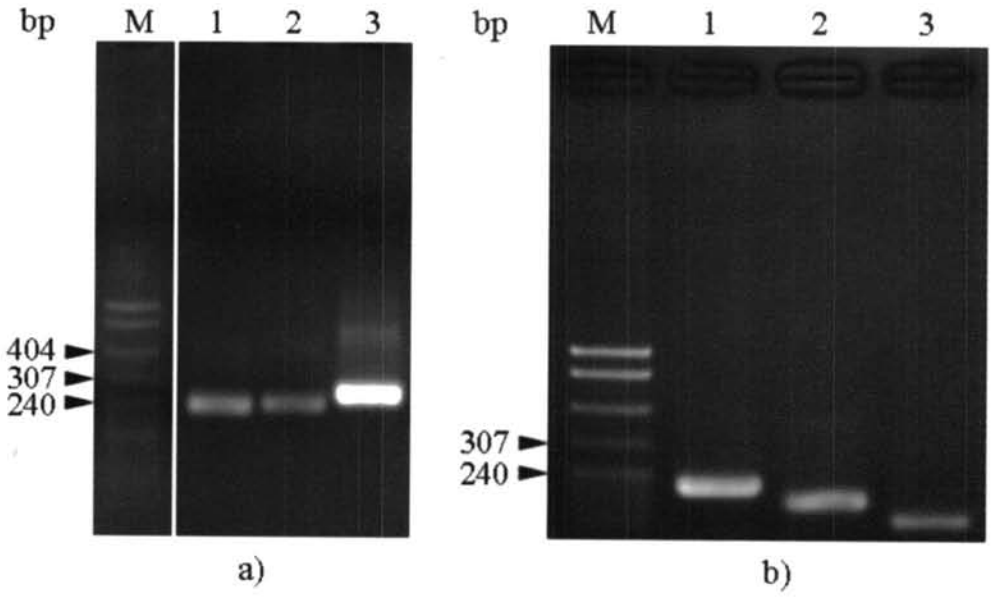


Figure 3.20 Agarose gel electrophoresis of the amplified fragments of the *Cam and CML* 3'UTR for use of making isoform specific probes. The PCR products was separated on a 1.8% agarose gel and visualized by ethidium bromide staining.

a) Lane M DNA Marker: pBR322/*Msp*I

Lane 1 3'UTR of *OsCam2*

Lane 2 3'UTR of *OsCam1-1*

Lane 3 3'UTR of *OsCML1*

b) Lane M DNA Marker : pBR322/*Msp*I

Lane 1 3'UTR of *OsCam1-3*

Lane 2 3'UTR of *OsCam1-2*

Lane 3 3'UTR of *OsCam3*

products derived from the genes of the japonica rice. However, the previous result as well as the sequence comparison with some available cDNA clones from the indica rice has indicated that their 3'UTR sequences are nearly identical to those of the indica genes, therefore DNA fragments prepared from these genes could effectively be used for hybridization with those from the indica rice.

The expression pattern of *Cam* and *CML* encoding genes under salt stress is shown in Figure 3.21. The level of steady state mRNA corresponding to *OsCam1-1* gene strongly increased under salt stress as early as 30 min and peaked at 1 hour after treatment. On the contrary, the level of *OsCam1-2, OsCam1-3, OsCam2* and *OsCam3* mRNAs appeared relatively unchanged throughout the period monitored under salt stress, however *OsCam1-3* and *OsCam3* displayed a very low level of expression. In Figure 3.21, RNA gel blot analysis also shows a low level of *OsCML1* mRNA which transiently increased at 1 hour. Modulation of gene expression in response to stress signals may reflect the function of the corresponding gene product. These results clearly indicate the induction of *OsCam1-1* gene and its possible role in mediating responses to salt stress.

In several plant species, *Cam* genes encoding different CaM isoforms have been reported to be differentially expressed in different developmental stages or in response to external stimuli such as salinity, wind, cold, wounding and pathogen attack (van der Luit *et al.*, 1999; Delumeau *et al.*, 2000; Yamakawa *et al.*, 2001; Duval *et al.*, 2002). Consistent with this result, the expression of *OsCam1-1* in two rice varieties, Pokkali and IR29 has been reported to increase during the initial phase of salt stress determined by microarray technique (Kawasaki *et al.*, 2001). In addition, NaCl treatment also increased the *OsCML1* mRNA level (1 hour) but its increase exhibited a more transient modulation.

The expression pattern of genes encoding OsCaM1-1, OsCam1-2, OsCam1-3, OsCam2 and OsCam3 was evidently different under salt stress treatment. The stimuli that strongly increased the level of steady state mRNA corresponding to *OsCam1-1* did not affect that of other *OsCam* genes. Previous reports have shown that highly conserved CaM isoforms may actually modulate their target proteins differently. In Arabidopsis, CaM2 was shown to stimulate NAD kinase more effectively than CaM4 and CaM6 (Liao *et al.*, 1996). By amino acid sequence comparison, OsCaM1-1 is more similar to AtCaM2, while OsCaM2 is more similar to AtCaM4. Glu for Asp substitution at position 7 similar to OsCaM2 may actually affect target binding because of their longer side chain as hypothesized by Duval *et al.* (2002).

In addition, competition among CaM isoforms for target proteins may be present therefore different transcriptional regulation can significantly affect their ration of CaM isoforms meaning that more "active" CaM isoform may exist in a cell that responds to a particular stimulus. By this interpretation, OsCaM1-1 possibly functions in mediate stress response under salinity by interacting with a subset of target proteins that might not act as *vivo* targets of other OsCaM proteins. Because stress-induced modulation of gene expression often reflects the function of the corresponding gene product in signaling, the results observed in this study suggest that OsCaM1-1 isoform probably functions in Ca^{2+}-mediated response to salt stress in rice.

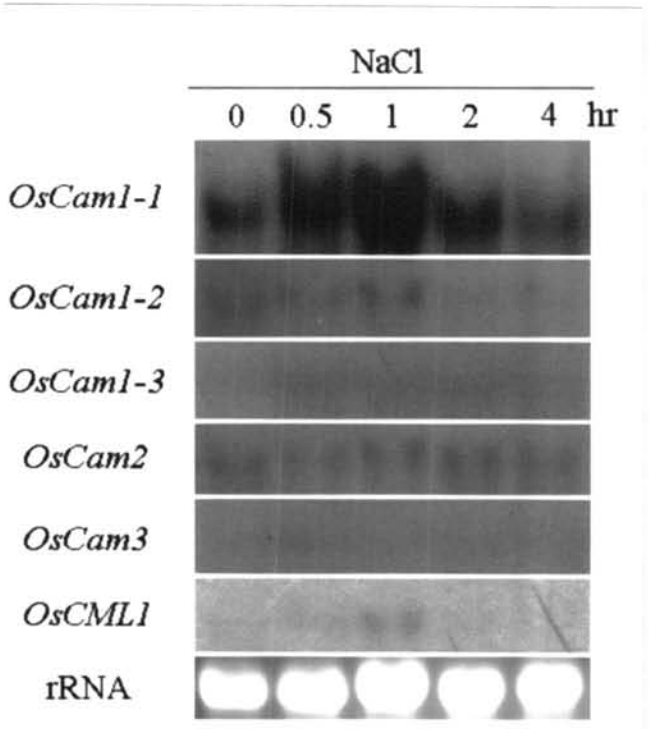


Figure 3.21 Expression patterns of CaM-encoding genes under salt stress signals. Twenty micrograms of total RNA isolated from leaves harvested at time points indicated after salt stress treatment were loaded into each lane in formaldehyde agarose gels, transferred to charged nylon membranes, and hybridized with the respective 3'-untranslated regions of the *Cam* and *CML* mRNAs at using the same hybridization conditions as above. Positive hybridizing bands were detected by autoradiography. rRNA on an agarose gel visualized by ethidium bromide staining are used as a loading control.

Genomic DNA blot analysis of *OsCam1-1*

To determine the copy number of *OsCam1-1*, rice (cv. KDML105) genomic DNA was isolated from 3-week old seedling leaves and digested with *Eco*RI, *Eco*RV, *BamH*I, *Bst*EII or *Hin*cII. Ten micrograms of the digested genomic DNA were fractionated in an agarose gel and transferred to a charged nylon membrane. The DNA fragments made from the 3' UTR of *OsCam1-1* (*OsCam1-1i-3'*) by PCR amplification were used to prepare the isoform-specific probe. DNA fragments were incorporated with [α-^{32}P]dCTP and used for hybridization which was carried out at 40°C for 16 hours. The blot was washed in 2X SSPE, 0.1% SDS twice and then once in 1X SSPE, 0.1% SDS at room temperature. Positive hybridizing bands were detected by autoradiography. Figure 3.22 revealed single positive hybridizing fragments of approximately 7.2, 4.4, 5.1, and 0.9 kb in size when the genomic DNA was digested with *Eco*RI, *Bam*HI, *Bst*EII, and *Hin*cII, respectively. But the genomic DNA digested with *Eco*RV displayed two bands corresponding to fragments of approximately 4.8 and 2.2 kb as shown in Figure 3.22 because the genomic DNA might be incompletely digested or the probe may hybridize with other *OsCam* genes. Although the 3' UTR sequences among different *OsCam* genes are more diverged than their coding regions, they share some degree of identity (~40%). Nonetheless, these data support the conclusion that rice contains a single copy of *OsCam1-1* which is in agreement with the TIGR database that predicts one *OsCam1-1* gene on chromosome 3 as shown in Figure 3.6.

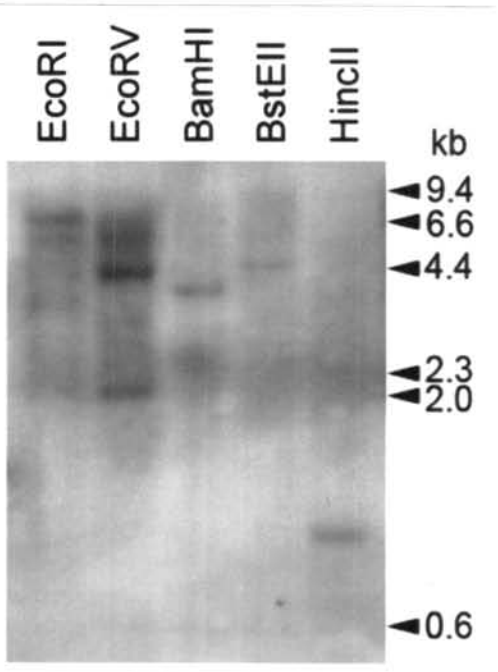


Figure 3.22 Genomic DNA blot analysis of *OsCam1-1* gene. Rice genomic DNA was digested with restriction enzymes and hybridized with the respective 3'-untranslated regions of the *OsCam1-1* (*OsCam1-1i-3'*) at using the same hybridization conditions as above. Positive hybridizing bands were detected by autoradiography.

rRNA can be used as a loading control

The plasmid containing Elongation Factor 1α (Accession number: AU091878) was digested with *Bst*EII at 37 °C overnight. Subsequently, this reaction was separated by 1.8% agarose gel electrophoresis as shown in Figure 3.23. The desired DNA fragment of about 0.5 kb was purified using the QIAquick gel extraction kit and incorporated with [α-^{32}P]dCTP to be used as probe in this experiment.

Total RNA from roots of FL530 was isolated and separated by formaldehyde agarose gel electrophoresis, blotted onto positively charged nylon membrane and hybridized with the denatured ^{32}P-oligolabeled Elongation Factor 1α probe. The resulting autoradiographs showing levels of Elongation Factor 1α mRNAs were compared with their corresponding agarose gels displaying rRNA visualized by ethidium bromide staining (Figure 3.24). The results showed that the relative band intensity of rRNA and Elongation Factor 1α displayed a similar pattern. Therefore, the amounts of rRNA can correctly be used for loading control for further experiments.

Expression pattern of *OsCam1-1* under salt stress in KDML105 and FL530

Rice seedlings of KDML105 and FL530 were hydroponically grown in WP No.2 Solution. Three-week old seedlings were transferred to the growth medium containing 85 mM NaCl and harvested at different time points (0, 0.5, 1, 2 and 4 hours) in liquid nitrogen and stored at -80 °C. Twenty micrograms of total RNA isolated from leaves or roots were fractionated in formaldehyde agarose gels and transferred to charged nylon membranes. Expression patterns of *OsCam1-1* under salt stress in KDML105 and FL530 were examined by RNA gel blot analysis. The *OsCam1-1i-3'* were used to prepare an

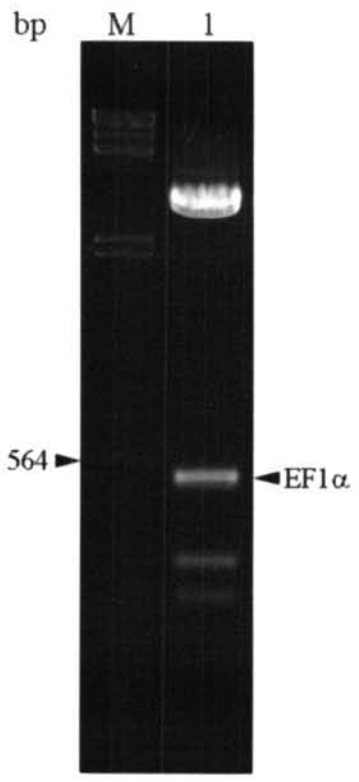


Figure 3.23 Agarose gel electrophoresis of the digested elongation factor 1α plasmid. The DNA fragments were separated on a 1.8% agarose gel and visualized by ethidium bromide staining.

Lane M DNA Marker: pBR322/*Msp*I

Lane 1 elongation factor 1α digested with *Bst*EII

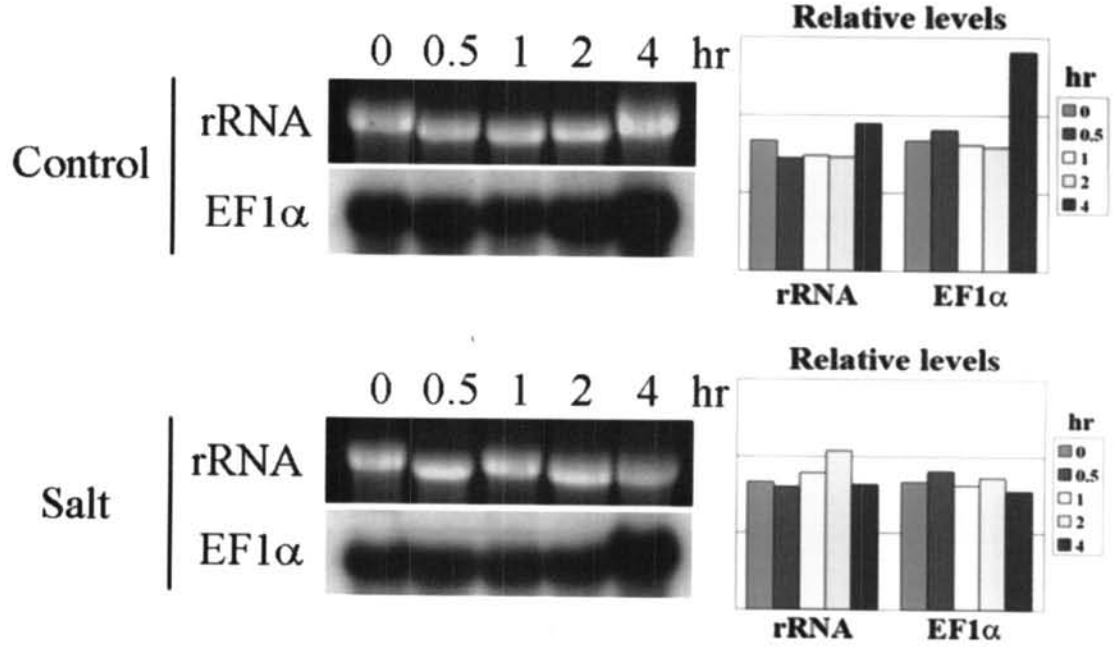


Figure 3.24 rRNA can be used as a loading control. Each lane was loaded with total RNA isolated from seedling roots of FL530 that were treated with or without 85 mM NaCl for 0, 0.5, 1, 2, and 4 hours. RNA was analyzed by gel-blot hybridization with denatured ^{32}P-oligolabeled Elongation Factor 1α (EF) probe. The autoradiograph showing expression levels of Elongation Factor 1α mRNAs was compared with the corresponding agarose gel displaying rRNA visualized by ethidium bromide staining.

isoform specific probe. DNA fragments were incorporated with [α-^{32}P]dCTP and used for hybridization which was carried out at 40°C for 16 hours. The blots were washed in 2X SSPE, 0.1% SDS twice and then once in 1X SSPE, 0.1% SDS at room temperature. Positive hybridizing bands were detected by autoradiography.

Figure 3.25 showed that salt stress increased the mRNA level of *OsCam1-1* in KDML105 and FL530 in leaves. Both showed the maximal mRNA levels at 1-2 hours, compared with controls, and the mRNA level slightly decreased after 4 hours of treatment (Figure 3.25A). Even though the induction can be observed in both lines, the expression level of *OsCam1-1* gene at 0.5 hours in FL530 leaves is slightly higher than that in KDML105 suggesting that the induction of *OsCam1-1* gene in FL530 leaves probably occurred earlier than that of KDML105. Expression of a calmodulin gene has been reported in a salt sensitive tomato and a salt tolerant species. It has been shown that expression of the *Cam* gene increased during the 15-45 minutes after salt stress treatment, however, the level of mRNA in the salt tolerant species is earlier than that in the salt sensitive tomato (Delumeau *et al.*, 2000). This result suggested that this *Cam* gene may contribute to salt tolerance.

In contrast, the expression of *OsCam1-1* gene in both lines in roots was not induced by salt stress during 4 hours of treatment compared with controls (Figure 3.25B). Although roots of KDML105 (Figure 3.25B) showed slightly increasing mRNA levels from 0 to 4 hours in salt stress condition, they showed no significant difference from the control, indicating that salt stress did not induce the expression of mRNA. However, the apparent mRNA increase might be due to some effect from transferring the rice plants to fresh media. Kawasaki *et al.* (2001) compared gene expression pattern of

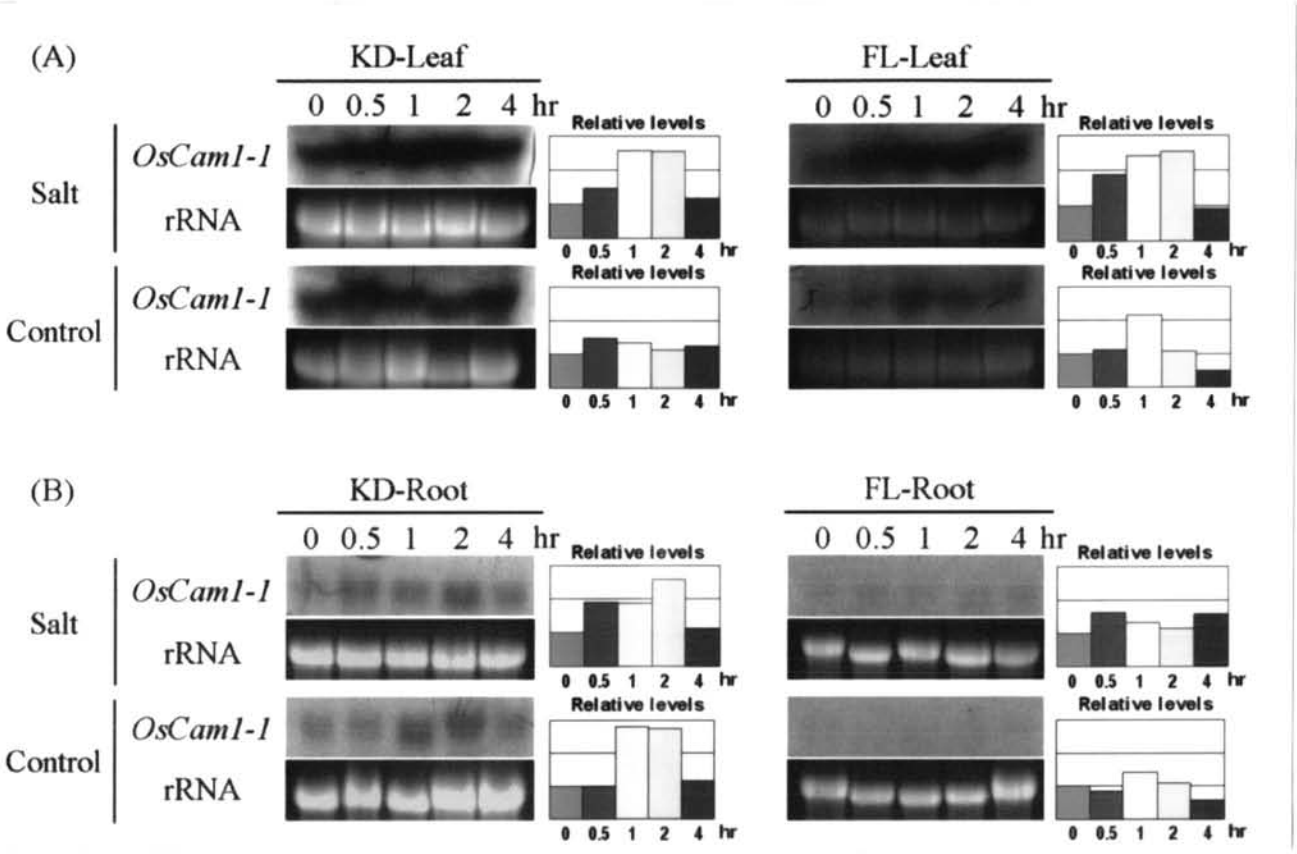


Figure 3.25 RNA blot analysis and quantitative comparison of *OsCam1-1* gene under salt stress in KDML105 and FL530. Each lane was loaded with total RNA isolated from seedling leaves and roots that were treated with or without 85 mM NaCl for 0, 0.5, 1, 2, and 4 hours. RNA was analyzed by gel-blot hybridization with denatured ^{32}P-oligolabeled *OsCam1-1i-3'* probe. An ethidium bromide-stained gel of each analysis is shown under its corresponding autoradiography.

two rice cultivars: Pokkali, a salt-tolerant cultivar and IR29, a salt-sensitive cultivar using microarray technique. They showed that expression of *OsCam1-1* in the salt sensitive rice was induced earlier than that in the salt tolerant cultivar but the induction was prolonged in the salt tolerant cultivar. However, concentration of NaCl used in Kawasaki's experiment was 150 mM. In these studies, 85 mM NaCl was used. Even though this concentration was enough for induction in leaves, it might be too low for the induction of *OsCam1-1* gene in roots.

Expression pattern of *OsCam1-1* under salt stress and with exogenous ABA application in KDML105 and FL530

Rice seedlings of KDML105 and FL530 were hydroponically grown in WP No.2 Solution. To test the effect of abscisic acid (ABA), 3-week old seedlings were sprayed with 100 μM ABA or buffer of the ABA solution (0.5% (v/v) Triton X-100, 2.5% (v/v) ethanol) as control and left for 2 hours before salt stress treatment was applied. Leaves and roots were harvested at different time points (0, 0.5, 1, 2 and 4 hours) in liquid nitrogen and stored at -80 °C. Twenty micrograms of total RNA isolated from leaves or roots were fractionated in formaldehyde agarose gels and transferred to charged nylon membranes. Expression patterns of *OsCam1-1* under salt stress with exogenous ABA application in KDML105 and FL530 were examined by RNA gel blot analysis. The *OsCam1-1i-3'* fragments were incorporated with [α-^{32}P]dCTP and used for hybridization which was carried out at 40°C for 16 hours. The blots were washed in 2X SSPE, 0.1% SDS twice and then once in 1X SSPE, 0.1% SDS at room temperature. Positive hybridizing bands were detected by autoradiography.

Expression pattern of *OsCam1-1* in leaves

Figure 3.26A showed the amounts of *OsCam1-1* mRNA from KDML105 and FL530 in leaves after spraying with exogenous abscisic acid (ABA). Buffer of the ABA solution containing 0.5 %(v/v) of Triton X-100, 2.5 %(v/v) ethanol was used as a control for spraying. The higher levels of *OsCam1-1* mRNA (KDML105) under salt stress were observed after 1 hour and then decreased after 2 hours in both conditions: with ABA and with buffer (Figure 3.26A). Additionally, under control condition (without salt stress), the *OsCam1-1* mRNA level increased at 0.5 hours and then decreased after 1 hour in both conditions: with ABA and with buffer (Figure 3.26B). Results from the control (without salt stress) suggest that spraying itself had some effect on the expression of *OsCam1-1* mRNA. Spraying might present a mechanical stimulus that probably caused increased levels of *OsCam* mRNA of KDML105. Touch is known to increase the mRNA expression of calmodulin genes in plants. One example was from Arabidopsis in which genes of at least three different calmodulin isoforms were induced by touch (Braam, 1990). Hence, in the experiments under salt stress treatment, spraying also caused increased levels of mRNA in leaves of KDML105 in addition to those from salt stress. However, with ABA, the *OsCam1-1* mRNA levels were noticeably lower than those with buffer. Therefore, ABA probably counters the effect of *OsCam1-1* induction by spraying as well as by salt stress.

For FL530, under control condition (without salt stress), the *OsCam1-1* mRNA level did not alter in both conditions: with ABA and with buffer (Figure 3.26B). These results indicate that spraying did not have an effect on the expression of *OsCam1-1* gene in FL530 leaves. In addition, in the absence of salt stress, ABA did not appear to have an

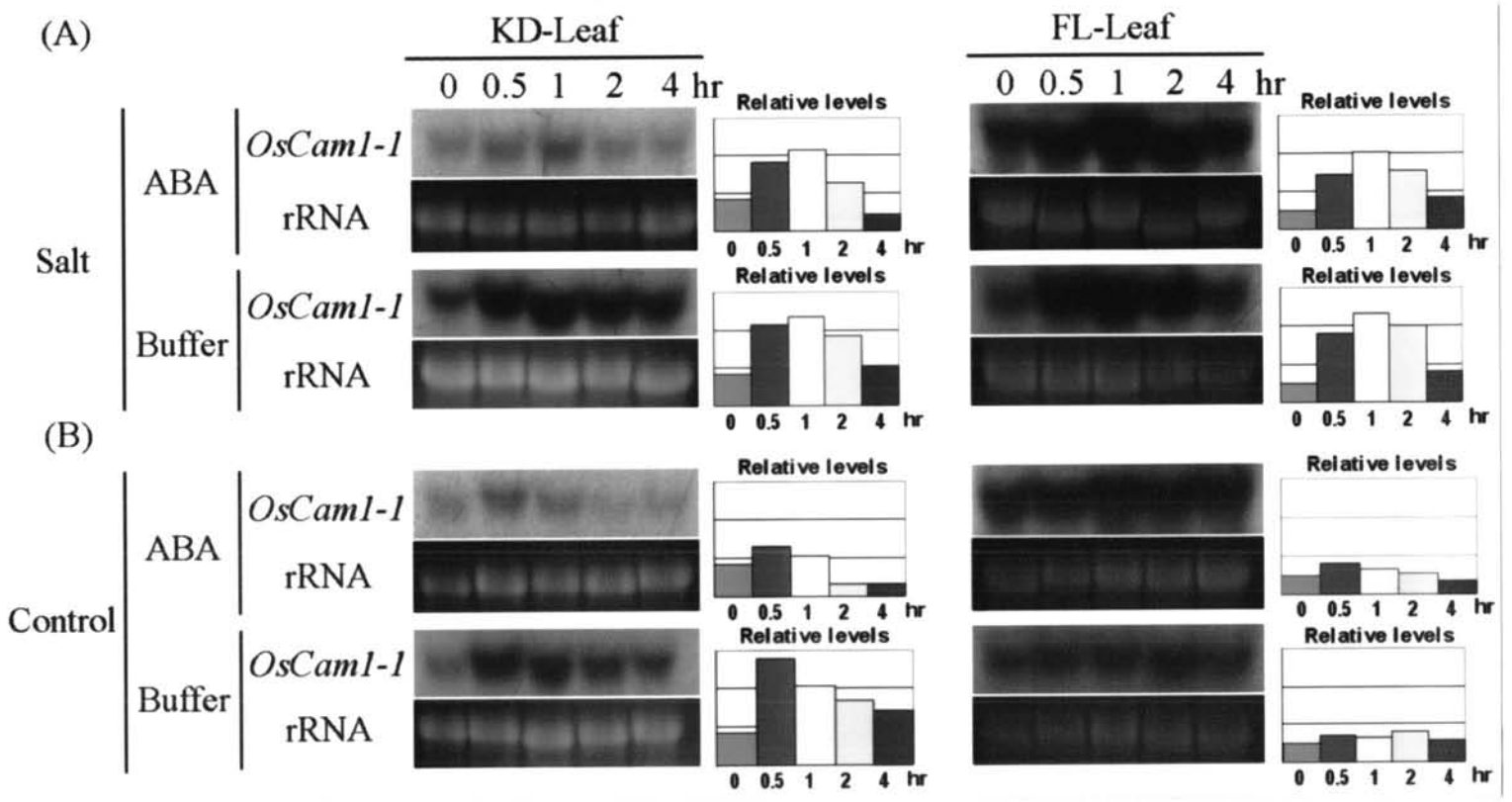


Figure 3.26 RNA blot analysis and quantitative comparison of *OsCam1-1* gene under salt stress and with exogenous ABA application in KDML105 and FL530 leaves. Each lane was loaded with total RNA isolated from ABA-sprayed seedling leaves of FL530 that were treated with or without 85 mM NaCl for 0, 0.5, 1, 2, and 4 hours and. RNA was analyzed by gel-blot hybridization with denatured ^{32}P-oligolabeled *OsCam1-1i-3'* probe. An ethidium bromide-stained gel of each analysis is shown under its corresponding autoradiography.

effect on the induction of *OsCam1-1* gene expression. Under salt stress, levels of *OsCam1-1* mRNA in both conditions: with ABA and with buffer, increased as expected from the effect of salt stress. Therefore, ABA and spraying did not have an effect on the induction of *OsCam1-1* gene in FL530 leaves.

Salinity has been known to enhance ABA level in many plants including rice (Shinosaki and Yamaguchi, 2000). Over the years evidence has been gathered which indicate that ABA induces physiological changes which can predispose plants to tolerate salt. Transpiration through stomatal pores is a crucial response of the plant under osmotic stress regulated by ABA. The closure of stomatal pores in aerial tissues is an important mechanism by which higher plants regulate their water balance (Bartels and Sunkar, 2005). ABA affects the water status by reducing transpiration, via effects on stomatal regulation and by possibly increasing water flux into roots (Christopher *et al.*, 1987). In this study, induction of *OsCam1-1* expression was decreased by the pre-treatment with exogenous ABA in leaves of KDML105 but was not affected in leaves of FL530. It appeared that exogenous ABA application reduced the induction effect by salt stress which was received later in leaves of KDML105 by predisposing the plants to tolerate salt stress. However, induction of *OsCam1-1* expression by salt stress still occurred in leaves of FL530 even after the ABA pretreatment suggesting that expression of this gene is more fine-tuned in FL530 than that of KDML105. Whether this is a cause or a consequence of its tolerance remains to be seen.

Expression pattern of *OsCam1-1* in roots

Figure 3.27 showed the amounts of mRNA of *OsCam1-1* from KDML105 and FL530 in roots after spraying with exogenous abscisic acid (ABA). Buffer of the ABA solution containing 0.5 %(v/v) of Triton X-100, 2.5 %(v/v) ethanol was used as a control for spraying. In KDML105, the high levels of *OsCam1-1* mRNA under salt stress with ABA and without ABA were observed at 0 hour and at 0.5 hours, respectively and they began to decrease after 1 hour after treatment as shown in Figure 3.27A. Correspondingly, under control condition (without salt stress), the level of *OsCam1-1* mRNA was noticeably high at 0 hour and then decreased after 1 hour in condition with ABA, while the level of mRNA without ABA increased at 0.5 hours and then decreased after 1 hour as shown in Figure 3.27B. The increased levels of mRNA by buffer treatment are probably due to the effect of spraying. Nonetheless, the highly increased levels of mRNA at 0 hour in both salt and control condition indicate that ABA has a direct effect on the expression pattern of *OsCam1-1* gene in KDML105 roots.

For FL530 under salt stress, the increased level of *OsCam1-1* mRNA with ABA treatment was observed at 0 hour and then it decreased after 1 hour while the mRNA level did not change during 4 hours with buffer treatment (Figure 3.27 A). Similarly, the mRNA did not change in control without ABA treatment indicating that spraying has no effect of *OsCam1-1* mRNA level of FL530 root. However, under control condition with ABA treatment, the *OsCam1-1* mRNA was noticeably high at 0 hour and then decreased shortly afterwards (Figure 3.27B). These results indicate that exogenous ABA application increases the *OsCam1-1* mRNA level in roots of FL530 both under salt stress and control conditions. Interestingly, the induction occurred to a lesser extend in roots without salt

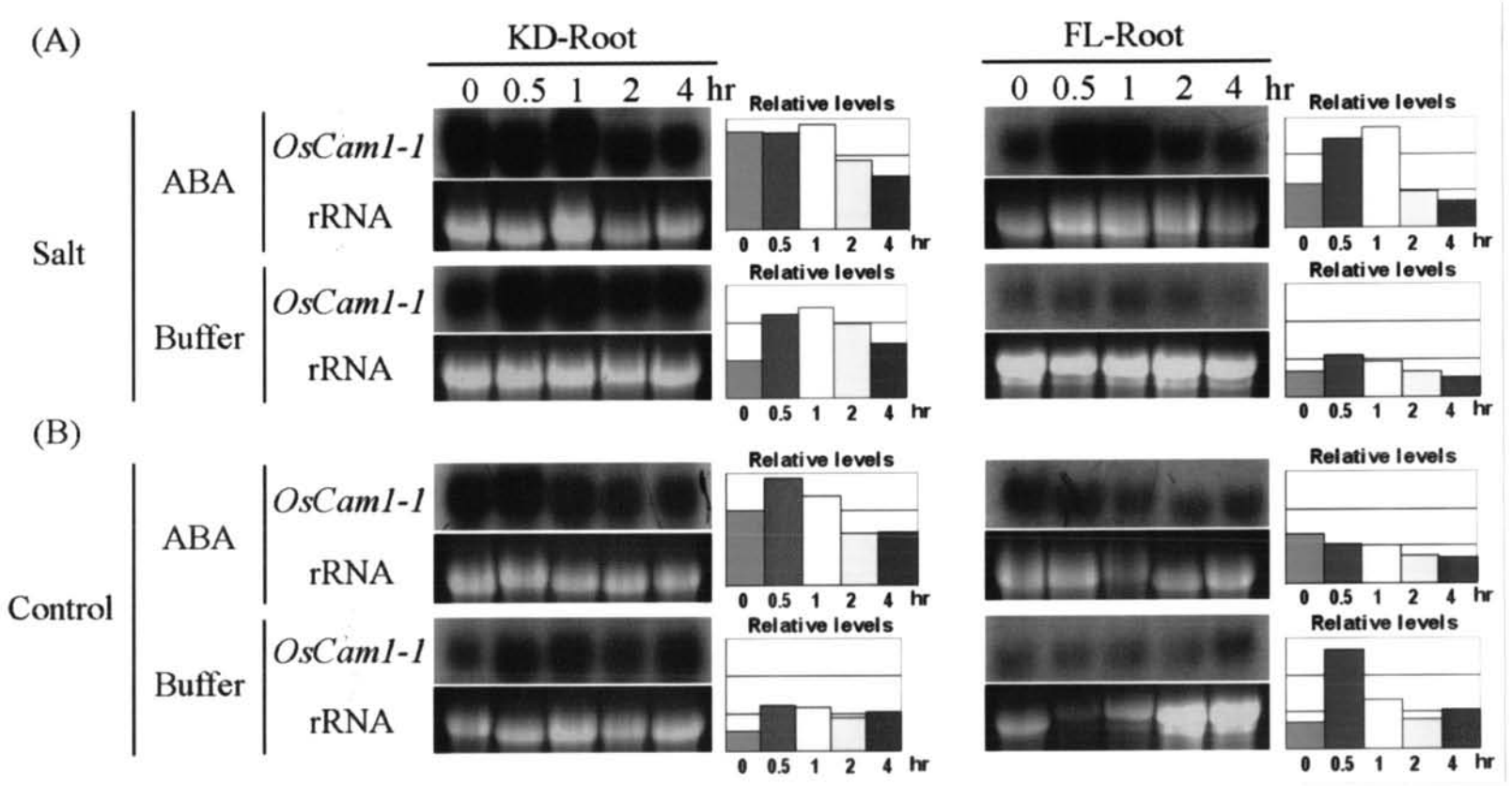


Figure 3.27 RNA blot analysis and quantitative comparison of *OsCam1-1* gene under salt stress and with exogenous ABA application in KDML105 and FL530 roots. Each lane was loaded with total RNA isolated from ABA-sprayed seedling roots of FL530 that were treated with or without 85 mM NaCl for 0, 0.5, 1, 2, and 4 hours and. RNA was analyzed by gel-blot hybridization with denatured ^{32}P-oligolabeled *OsCam1-1i-3'* probe. An ethidium bromide-stained gel of each analysis is shown under its corresponding autoradiography.

stress treatment and was prolonged under salt stress.

It has previously been shown that exogenously applied ABA induced many aspects of changes in various root morphological features in rice seedling, including tip swelling, root hair formation and lateral root production (Chen *et al.*, 2006). In addition, as indicated by the increased exudation volume and K^+ content in xylem sap, ABA enhanced root cell vitality and increased water permeability. These responses of roots to exogenous ABA application have been shown depending on an increase of cytosolic Ca^{2+} and calmodulin. In this study, the exogenous ABA-induced expression *Cam1-1* gene suggests that *OsCam1-1* may serve as a Ca^{2+} sensor to activate downstream components leading to similar responses in roots. It should be noted that mechanical stimulus created by spraying had a direct effect on *OsCam1-1* gene induction in KDML105 which complicated the interpretation of it results. This effect may be reduced by applying exogenous ABA in growing media rather than by spraying on leaves. Nonetheless, it was clear that ABA application increased *OsCam1-1* induction in roots of FL530 which effect of spraying was not found and its induction was prolonged when the plant experienced salt stress. This result suggests that *OsCam1-1* gene has an important role in salt stress response mediated by ABA in rice plants.