

UNIVERSIDADE FEDERAL DOS VALES DO JEQUITINHONHA E MUCURI
INSTITUTO DE CIÊNCIAS AGRÁRIAS
CURSO DE BACHARELADO EM CIÊNCIAS AGRÁRIAS

IDENTIFICAÇÃO DA FAMÍLIA DE PROTEÍNAS CALMODULINA (CaM) DE
Cajanus cajan

Lara de Windson Oliveira Almeida Marchiore

Unai
2018

UNIVERSIDADE FEDERAL DOS VALES DO JEQUITINHONHA E MUCURI
INSTITUTO DE CIÊNCIAS AGRÁRIAS

IDENTIFICAÇÃO DA FAMÍLIA DE PROTEÍNAS CALMODULINA (CaM) DE
Cajanus cajan

Lara de Windson Oliveira Almeida Marchiore

Orientador: **Prof. Dr. Wellington Ferreira Campos**

Trabalho de Conclusão de Curso apresentado ao
Curso de Ciências Agrárias, como parte dos
requisitos exigidos para a conclusão do curso.

Unaí
2018

IDENTIFICAÇÃO DA FAMÍLIA DE PROTEÍNAS CALMODULINA (CaM) DE
Cajanus cajan

Lara de Windson Oliveira Almeida Marchiore

Orientador: **Prof. Dr. Wellington Ferreira Campos**

Trabalho de Conclusão de Curso apresentado ao
Curso de Ciências Agrárias, como parte dos
requisitos exigidos para a conclusão do curso.

APROVADO em ... / ... / ...

Prof^ª Dr^ª RENATA OLIVEIRA BATISTA - UFVJM

Prof. Dr. GUSTAVO LEÃO ROSADO – FACISA

Prof. Dr. WELLINGTON FERREIRA CAMPOS - UFVJM

RESUMO

O cálcio (Ca^{2+}) é um elemento fundamental como mensageiro secundário para mediar muitos processos biológicos e em resposta a estímulos bióticos e abióticos. A proteína calmodulina (CaM) atua como um sensor de Ca^{2+} em eucariotos e regula a cascata de eventos regulatórios e participa do evento de regulação. Apesar da sua importância as proteínas CaM de *Cajanus cajan* ainda não foram identificadas. O objetivo com este trabalho foi identificar a família de proteínas (CaM) no genoma de *C. cajan*. No total foram identificados cinco proteínas CaM em *C. cajan*, sendo que as sequências de aminoácidos das CaM2 e 3 são iguais, caracterizando uma isoforma. Caracteristicamente, as CaM de *C. cajan* são proteínas ácidas, tem 149 aminoácidos e possivelmente estão localizadas no citoplasma ou citosol. Seus genes possuem pelo menos um íntron e dois éxons e estão principalmente localizados nos cromossomos 2 e 7. Além disso, seus genes são expressos em diferentes tecidos e estágios do desenvolvimento vegetal. Em conjunto, os dados evidenciam a presença de cinco proteínas CaM em *C. cajan* e que estas podem estar envolvidas com o crescimento e desenvolvimento vegetal.

Palavras-chave: Cálcio; calmodulina; feijão-guandu.

SUMÁRIO

1. INTRODUÇÃO.....	6
2. OBJETIVOS.....	7
2.1. Objetivo geral	7
2.2. Objetivos específicos	7
3. MATERIAL E MÉTODOS.....	8
3.1. Identificação de genes CaM no genoma de <i>Cajanus cajan</i>	8
3.2. Estrutura e localização física dos genes <i>CcCaM</i>	9
3.3. Caracterização físico-química e localização subcelular das proteínas CcCaM	9
3.4. Análises filogenéticas	9
3.5. Análise de expressão <i>in silico</i>	10
4. RESULTADOS E DISCUSSÃO	10
6. CONCLUSÃO.....	15
7. REFERÊNCIAS	17

1. INTRODUÇÃO

O cálcio (Ca^{2+}) atua como um mensageiro secundário para mediar uma variedade de processos biológicos em resposta a vários estímulos bióticos e abióticos (Sanders et al., 1999; Batistič and Kudla, 2012). O sinal de Ca^{2+} é decodificado e transmitido por proteínas denominadas calmodulina (CaM), pequenas proteínas ácidas, com cerca de 149 aminoácidos. A sua sequência de aminoácidos é altamente conservada e possuem domínios que fazem ligação com o cálcio graças a uma estrutura hélice-volta-hélice conhecida como EF-hand. As calmodulinas ao ligar ao Ca^{2+} podem interagir e modular a atividade de outras proteínas envolvidas em uma grande quantidade de processos biológicos (Hashimoto and Kudla, 2011; Kudla et al., 2010).

O genoma de *Arabidopsis* abriga sete genes que codificam quatro isoformas de proteínas CaM (McCormack and Braam, 2003). No arroz, há cinco genes que codificam três isoformas de CaM, enquanto em tomate, tabaco, batata e soja há, respectivamente, 6, 7, 4 e 5 genes CaM (Sang et al., 1995; Boonburapong and Buaboocha, 2007; Zhao et al., 2013).

Diversos membros de proteínas CaM em diferentes espécies estão envolvidas com respostas aos estresses abióticos, por exemplo, a superexpressão do gene CaM4 de *Glycine max* aumentou a tolerância de plantas de soja ao estresse salino (Rao et al., 2014). Assim, a elucidação dos mecanismos que envolvem as proteínas CaM nas respostas aos distintos tipos de estresse tem gerado grande interesse. Entretanto, esses mecanismos e seus genes CaM são desconhecidos em *Cajanus cajan*.

C. cajan, popularmente conhecida como feijão-guandú, é uma espécie tolerante ao calor, à seca e baixa fertilidade do solo, tornando-a uma opção para os sistemas agrícolas onde há pouco capital ou adoção de tecnologias. Além disso, apresenta alto teor de proteínas e carboidratos, possui um sistema radicular cuja raiz principal pode alcançar 3 metros de profundidade, apresenta grande potencial para resistir ao estresse hídrico tem certa rusticidade e multifuncionalidade (Odeny, 2007; Khoury et al., 2015).

O município de Unaí é coberto pelo Bioma Cerrado que cobre aproximadamente 207 milhões de hectares e apresenta importância no cenário agrícola brasileiro. Bioma brasileiro que exibe expressiva biodiversidade explorada na produção de alimentos, forragem, madeira, fibras, fármacos e plantas ornamentais. Até a década de 1970 era uma região que tinha como base de sua economia a criação de gado, cultivo de arroz, produção de carvão vegetal e madeira. Nas últimas décadas, o desenvolvimento agrícola tornou a região fonte de interesse econômico e de desenvolvimento (EMBRAPA CERRADOS, 2005). A produção estimada para a safra 2017/18 indica um volume de 225,6 milhões de toneladas. Apesar desse resultado ser 5,1% menor que o da última safra, a expectativa é que a produção expresse o comportamento normal de safras anteriores (CONAB, 2018).

Em função das inúmeras mudanças climáticas das últimas décadas, entender os mecanismos de respostas das plantas aos estímulos ambientais pode minimizar os impactos negativos na produção agrícola. Especialmente na cidade de Unaí onde as características edafoclimáticas desafiam a resposta adaptativa das plantas.

2. OBJETIVOS

2.1. Objetivo geral

Objetivou-se com a realização deste trabalho identificar a família de proteínas Calmodulina (CaM) no genoma de *Cajanus cajan*.

2.2. Objetivos específicos

- Identificar e caracterizar os genes CaM de *C. cajan* (CcCaM);
- Caracterizar bioquimicamente as CcCaM;
- Predizer a localização subcelular das CcCaM;
- Apresentar as relações filogenéticas das CcCaM;
- Caracterizar a expressão dos genes CcCaM nos distintos tecidos e estágios de desenvolvimento de plantas de *C. cajan*.

3. MATERIAL E MÉTODOS

3.1. Identificação de genes CaM no genoma de *Cajanus cajan*

As sequências de aminoácidos das proteínas CaM de *Arabidopsis thaliana* (AtCaM), *Glycine max* (GmCaM) e *Oryza sativa* (OsCaM) foram coletadas nos respectivos bancos de dados: “The Arabidopsis Information Resource” (TAIR, <http://www.arabidopsis.org>) (Lamesch et al., 2012), Phytozome V12 (<http://www.phytozome.net>) (Goodstein et al., 2012) e “Rice Genome Annotation Project” (<http://rice.plantbiology.msu.edu/>) (Kawahara et al., 2013). Em seguida, as sequências foram usadas como *queries* no BLASTp (Altschul et al., 1997) para procurar as sequências de proteínas CaM de *Cajanus cajan* (CcCaM) no banco de dados “Legume Information System” (LIS, <http://legumeinfo.org>) (DASH et al., 2016). Os parâmetros estatísticos usados no BLASTp foram: alvo, proteoma; Expect value threshold, 1e-5; e Matriz de comparação, BLOSUM62.

Os números de identificação de domínios EF-hand (SM00054, IPR011992, IPR018247, IPR003299, IPR002048, IPR007330, IPR001751, PF13499, PF13405, PF13803, PF13833, PF13202, PF00036, PD00012, PS00018, PS50222, PS00303 e SSP47473) e palavras-chave (‘Calmodulin’ e ‘EF-hand’) foram usados no banco de dados LIS para uma procura complementar por proteínas CcCaM. As sequências coletadas foram analisadas por similaridade para remover as sequências redundantes.

As sequências não redundantes foram submetidas à análise de domínios de proteína usando os Modelos Hidden Markov dos programas SMART (Simple Modular Architecture Research Tool, <http://smart.embl.de>) (Letunic and Bork, 2017), Pfam (<http://pfam.xfam.org>) (Finn et al., 2016) e InterPro (<http://www.ebi.ac.uk/interpro/>) (Finn et al., 2017). Em seguida, as sequências foram comparadas com as proteínas CaM2 de *A. thaliana* (AtCaM2) e *G. max* (GmCaM2) usando o programa ClustalW (THOMPSON et al., 1994). As sequências de *C. cajan* que tinham aproximadamente 149 aminoácidos, apenas dois pares de *motifs* EF-hand e apresentaram mais de 78% de identidade com as sequências de AtCaM2 e GmCaM2 foram consideradas como proteínas CcCaM verdadeiras e nomeadas de acordo com a sua similaridade com proteínas CaM de *A. thaliana*. Por fim, as

sequências de nucleotídeos e demais informações sobre os genes *CcCaM* foram obtidas no banco de dados LIS.

3.2. Estrutura e localização física dos genes *CcCaM*

A estrutura (regiões de UTR, éxons e íntrons) dos genes *CcCaM* foi deduzida usando a ferramenta GSDS (Gene structure Display Server; <http://gsds.cbi.pku.edu.cn/>) (Hu et al., 2015), via alinhamento dos respectivos mRNAs com as sequências genômicas correspondentes. As informações presentes no banco de dados LIS foram usadas para determinar a localização dos genes *CcCaM* nos cromossomos de *C. cajan*.

3.3. Caracterização físico-química e localização subcelular das proteínas CcCaM

Parâmetros bioquímicos tais como ponto isoelétrico (pI) e peso molecular (kDa), foram feitos usando a ferramenta *ExPASy ProtParam* (<http://web.expasy.org/protparam/>) (Gasteiger et al., 2003). As ferramentas *Wolf PSORT* (http://www.genscript.com/psort/wolf_psort.html) (Horton et al., 2007) foi usada para prever a localização subcelular das proteínas CcCaM.

3.4. Análises filogenéticas

Sequências de proteínas CaM de diferentes espécies de monocotiledôneas, dicotiledôneas e de *C. cajan* foram alinhadas usando os programas ClustalW (Thompson et al., 1994). O arquivo contendo os alinhamentos foi usado para a construção da árvore filogenética no programa MEGA7 (Kumar et al., 2016). Os seguintes parâmetros estatísticos foram usados para a construção das árvores: (a) Reconstrução filogenética; (b) Maximum Likelihood; (c) Método Bootstrap; (d) 1000 replicatas; (e) Aminoácidos; (f) Jones-Taylor-Thornton (JTT); (g) Deleção parcial; (h) 95%; (i) Nearest-Neighbor-Interchange (NNI); (j) Muito Forte (Mohanta et al., 2017).

3.5. Análise de expressão *in silico*

A análise de expressão gênica utilizou os dados de RNA-seq disponibilizados no banco de dados “Legume Information System” (LIS, <http://legumeinfo.org>) (DASH et al., 2016). Os dados foram obtidos por Pazhamala et al. (2017), onde podem ser encontrados maiores detalhes experimentais.

4. RESULTADOS E DISCUSSÃO

O conjunto gênico constituído de 48.680 genes (Varshney et al., 2011) foi utilizado para procurar os genes CaM de *C. cajan*, usando as sequências de aminoácidos das proteínas AtCaM2 e GmCaM2, como também os números de identificação dos domínios EF-hand e as palavras chave “calmodulin” e “EF-hand”. No total foram coletadas 108 sequências não redundantes. Após análise de domínios foram selecionadas 79 sequências que continham apenas domínios EF-hand. Destas, apenas cinco sequências apresentavam mais de 78% de identidade com as proteínas AtCaM2 e GmCaM2, continham 149 aminoácidos e 2 pares de EF hand e, portanto, foram consideradas como calmodulinas verdadeiras em *C. cajan*. As calmodulinas de *C. cajan* foram nomeadas de acordo com a sua porcentagem de identidade com AtCaM2 e GmCaM2.

As CaM de *C. Cajan* apresentam 4 domínios EF-hand, são proteínas ácidas (pI 4,01 a 4,12), pequenas (16,80 a 17,06 KDa), sendo que as CaM1, 2 e 3 tem 149 aminoácidos; as CaM4,5 tem 150 aminoácidos e têm mais de 78% de similaridade com as CaM2 de *Arabidopsis* e *G. max*. Estas também ricas e metionina (Tabela 1). Todas estas características bioquímicas das proteínas CcCaM são semelhantes aquelas já encontradas para *Arabidopsis* (McCormack e Braam, 2003) e arroz (Boonburapong e Buaboocha, 2007).

Os genes CcCaM tem entre 542 e 2401 pares de base (pb), seus transcritos tem entre 450 e 453 pb, e estão localizados nos cromossomos 2 e 7, exceto um que não foi possível determinar sua localização (Tabela 2). A análise de estrutura gênica aponta que os genes CcCaM tem entre 1 e 3 íntrons e 2 e 4 éxons (Figura 1).

Tabela 1: Características físico-químicas das proteínas CaMs de *C. cajan*.

Gene	Locus ^a	EF hand ^b	AA ^c	PM ^d	pI ^e	M% ^f	AtCaM2% ^g	GmCaM2% ^h
CcCaM1	C.cajan_18083	4	149	16,80	4,09	6	95	95
CcCaM2	C.cajan_05680	4	149	16,86	4,12	6	98	98
CcCaM3	C.cajan_40611	4	149	16,86	4,12	6	98	98
CcCaM4	C.cajan_07108	4	150	17,06	4,07	6	78	78
CcCaM5	C.cajan_20364	4	150	16,92	4,01	6	80	79

^a Locus anotado no banco de dados LIS; ^b Numero de domínios Ef hand baseado no InterProScan; ^c Número total de aminoácidos; ^d PM, Peso Molecular, KDa; ^e pI, Ponto isoeletrico; ^f M% = Porcentagem de aminoácidos Metionina, ^g Porcentagem de identidade com a CaM2 de *Arabidopsis* ^h Porcentagem de identidade com a CaM2 de *G. max*

Tabela 2: Caracterização dos genes CaM de *C. cajan*

Gene	Locus ^a	Cromossomo	Localização (pb) ^b	Gene (pb)	CDS (pb)
CcCaM1	C.cajan_18083	7	8293076..8293617	542	450
CcCaM2	C.cajan_05680	2	12686980..12688341	1362	450
CcCaM3	C.cajan_40611	?	Scaffold133838:23090..25311	2222	450
CcCaM4	C.cajan_07108	2	27661710..27663051	1342	453
CcCaM5	C.cajan_20364	7	12479238..12481638	2401	453

^a Locus anotado no banco de dados LIS; ^b Localização no cromossomo, pb = pares de base.

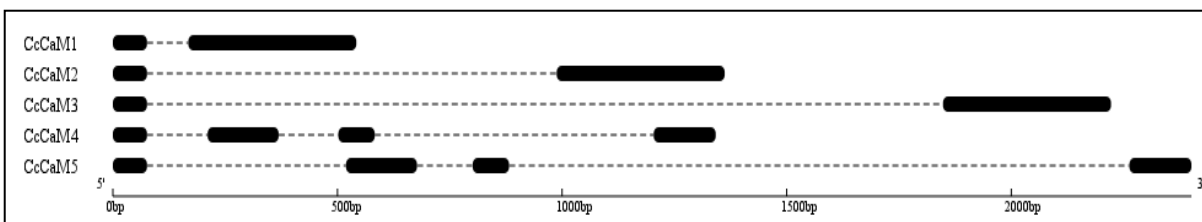


Figura 1: Estrutura dos genes CcCaM. Quadrado preto = éxon; Linha pontilhada = íntron.

De acordo com o alinhamento, as sequências de aminoácidos das calmodulinas de *C. cajan* e de diferentes espécies são conservadas (Figura 2). O alinhamento ainda indica que as CaM2 e 3 de *C. cajan* são iguais em sequência de aminoácidos, entretanto, possuem locus e genes distintos, caracterizando uma isoforma (Figura 1 e 2, Tabela 2). Em outras espécies tais como *Arabidopsis* e *G. max* possuem, respectivamente, sete e cinco genes, sendo que em *Arabidopsis* estes genes codificam três isoformas distintas, enquanto em *G. max* há a codificação de quatro isoformas (McCormack e Braam, 2003; Lee et al., 1995). Por fim, a predição da localização subcelular indica que as CaM de *C. cajan* estão presentes no núcleo (CcCaM 1, 2 e 3) e no citoplasma (CcCaM 4 e 5) (Tabela 3).

A análise filogenética revela que as CcCaM2 e 3 formam um clado juntamente com as CaM 2,3, 5, 6 e 7 de *Arabidopsis* e as CaM 1 e 3 de *G. max*. Já as CcCaM 4 e 5 estão agrupadas com as CaM 4 e 5 de *G. max*, enquanto a CcCaM 1 agrupou-se com CaM de arroz, milho e tomate (Figura 3). Estes dados sugerem a possibilidade de que diferentes grupos de CaM tenham distintos papéis nos processos biológicos vegetais. Neste sentido, foi analisado o perfil de expressão dos genes CcCaM em diferentes tecidos e estágios de desenvolvimento (Figura 4).

A expressão dos genes CcCaM1 a CcCaM5 é bem variada entre os tecidos e estágios de desenvolvimento de plantas de *C. cajan* (Figura 4). Especificamente, a quantidade de transcritos dos genes CcCaM2 e 3 é superior as demais calmodulinas. Além disso, observou-se uma variação temporal na expressão dos genes CcCaM3, cujos transcritos começam a aumentar no tecido radicular a partir do estágio vegetativo até o estágio de senescência, sugerindo que este genes esteja relacionado com o desenvolvimento e crescimento vegetal. Ademais, a expressão deste mesmo transcrito no nódulo é alta no estágio vegetativo e começa a reduzir a partir do estágio reprodutivo e estabiliza no estágio de senescência. Recentemente, foi demonstrado o envolvimento da calmodulina-like 38 na regulação negativa do crescimento radicular de plantas de *Arabidopsis* (Campos et al 2017). Diante disso, levantamos a hipótese que a CcCaM3 participe da regulação negativa do crescimento radicular de *C. cajan*. Desta forma, experimentos com genética reversa permitirão desvendar o real papel desta calmodulina.

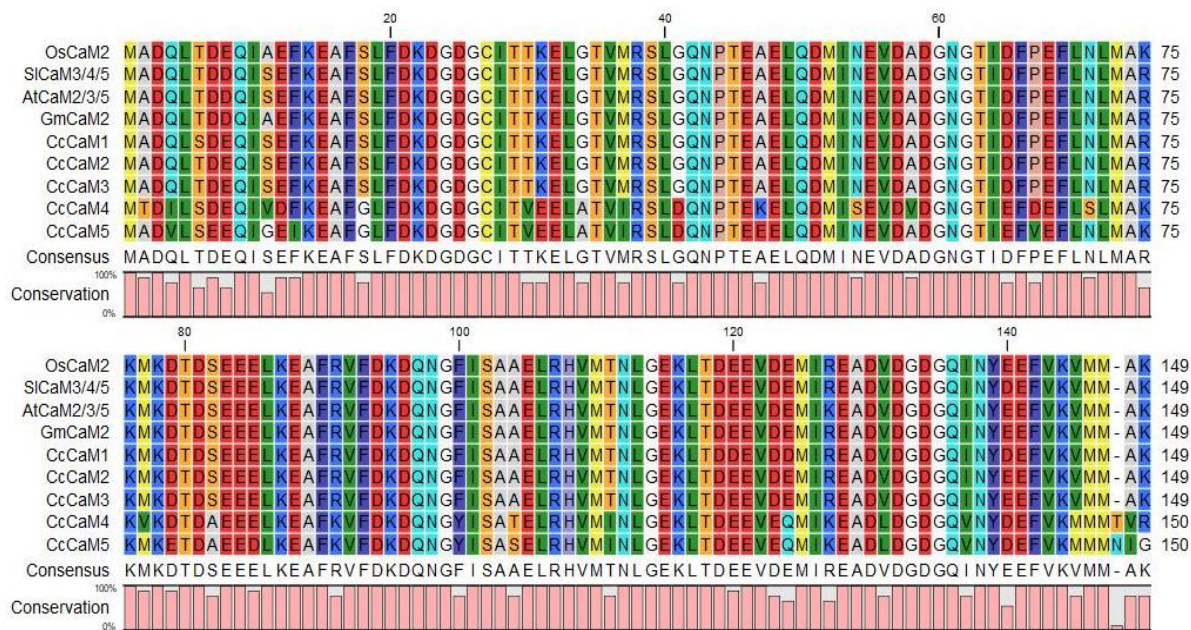


Figura 2: Alinhamento de sequência de aminoácidos das calmodulinas de *O. sativa* (OsCaM2), *Solanum lycopersicum* (SlCaM3/4/5), *A. thaliana* (AtCaM2/3/5), *G. max* (GmCaM2) e *C. cajan* (CcCaM1 a 5).

Tabela 3: Predição da localização subcelular das proteínas CaM de *C. cajan*.

Gene	Locus ^a	Localização subcelular ^b
CcCaM1	C.cajan_18083	nucl: 4, mito: 4, extr: 3, cyto: 2
CcCaM2	C.cajan_05680	nucl: 4, mito: 4, extr: 3, cyto: 2
CcCaM3	C.cajan_40611	nucl: 4, mito: 4, extr: 3, cyto: 2
CcCaM4	C.cajan_07108	cyto: 5.5, chlo: 4, cyto_nucl: 3.8, extr: 2, cysk_nucl: 1,3
CcCaM5	C.cajan_20364	cyto: 5.5, chlo: 5, cyto_nucl: 4, extr: 2

^a Locus anotado no banco de dados LIS

^b Predição da localização subcelular com a ferramenta PSORT

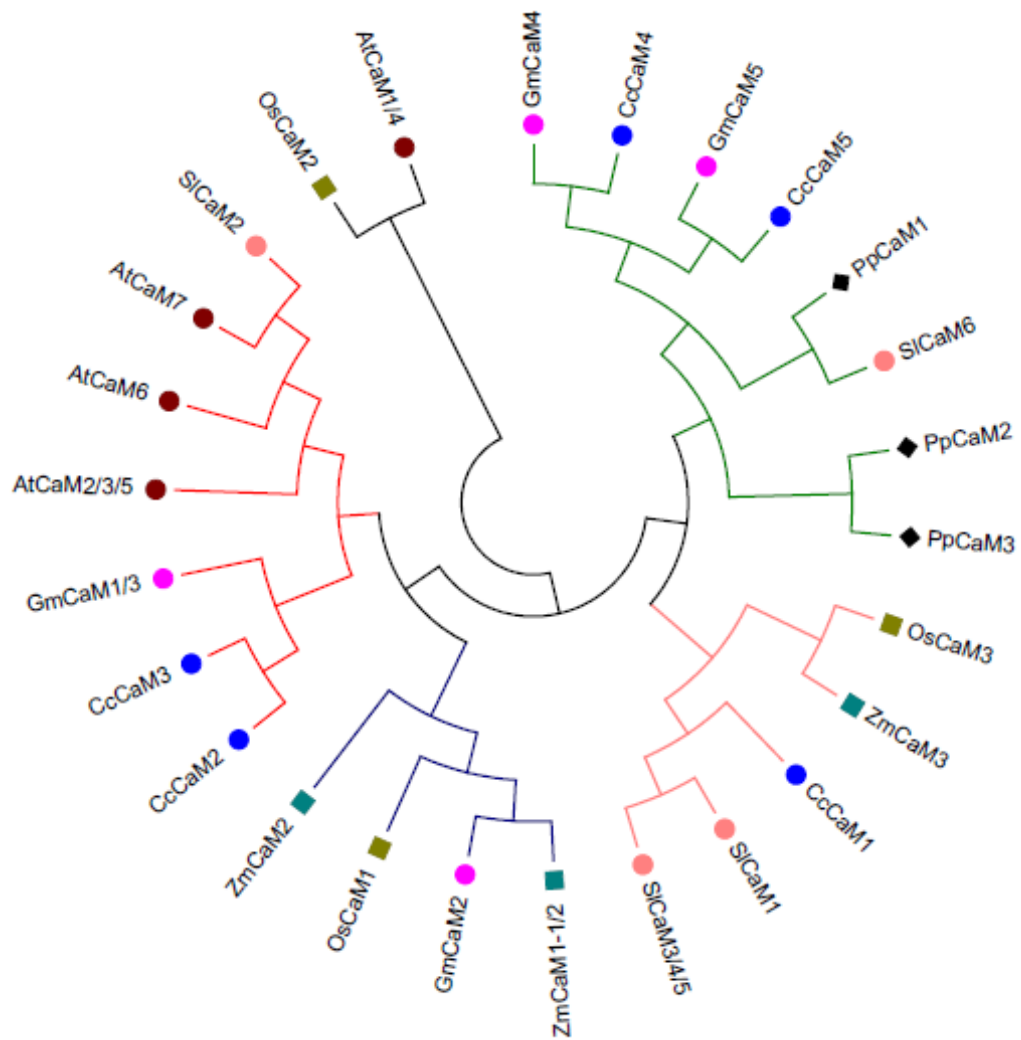


Figura 3: Árvore filogenética de proteínas CaM de *Zea mays* (ZmCaM), *Physcomitrella patens* (PpCaM), *O. sativa* (OsCaM), *S. lycopersicum* (SlCaM), *A. thaliana* (AtCaM), *G. max* (GmCaM) e *C. cajan* (CcCaM).

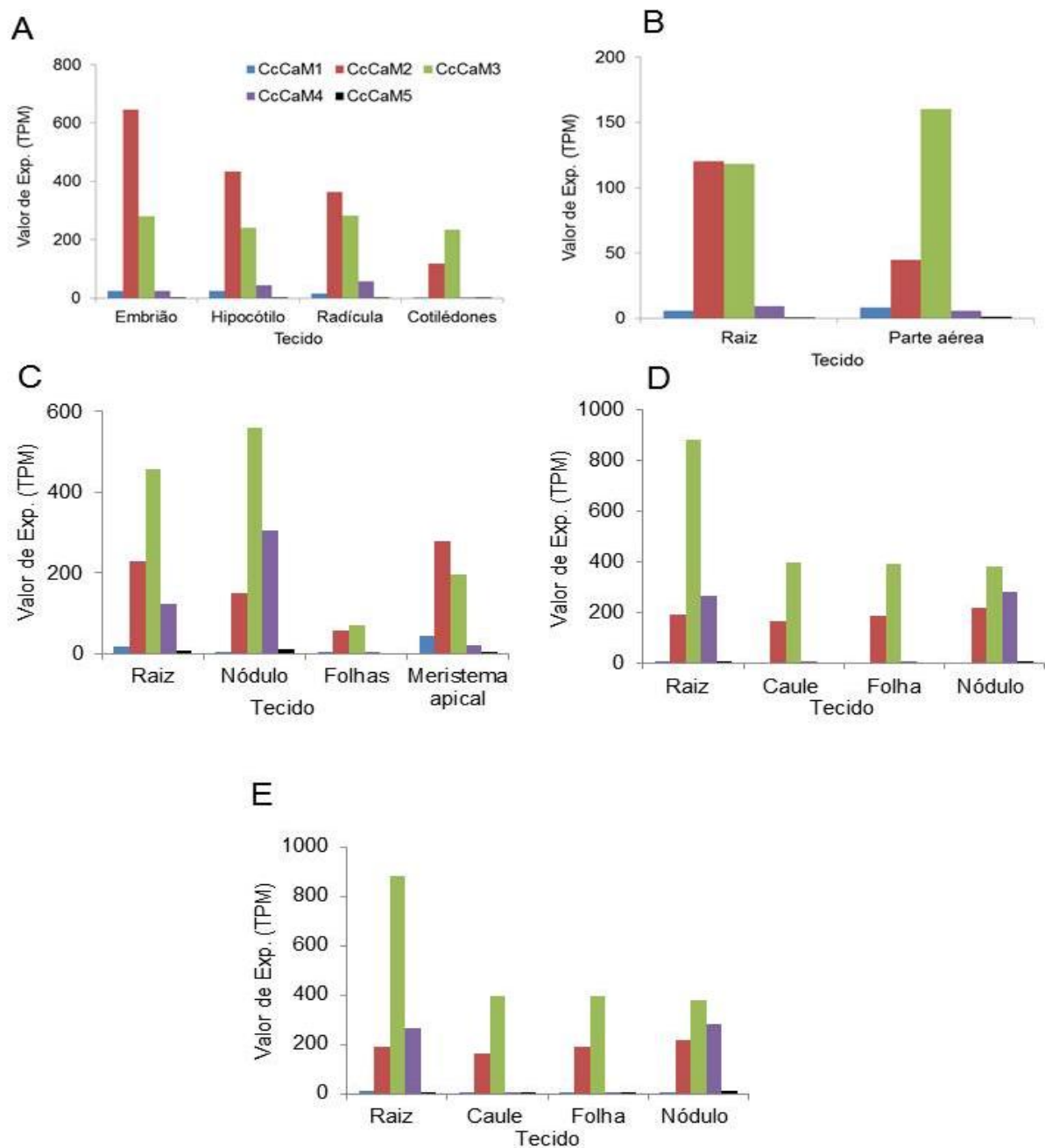


Figura 4: Perfil de expressão dos genes que codificam CaM em *C. cajan* (CcCaM1 a CcCaM5) em diferentes tecidos nos estádios de (A) germinação; (B) plântula; (C) vegetativo, (D) reprodutivo (E) e de senescência em plantas de *C. cajan*. TPM = Transcritos por milhão.

6. CONCLUSÃO

- A família de proteínas CaM em *C. cajan* é constituída por 5 genes que codificam 4 isoformas;
- Os genes CcCaM são expressos em todos os estádios fenológicos da planta, mas seus papéis ainda deverão ser estudados.

7. REFERÊNCIAS

- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J.** (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389–3402.
- Batistič, O. and Kudla, J.** (2012). Analysis of calcium signaling pathways in plants. *Biochim. Biophys. Acta - Gen. Subj.* **1820**: 1283–1293.
- Boonburapong, B. and Buaboocha, T.** (2007). Genome-wide identification and analyses of the rice calmodulin and related potential calcium sensor proteins. *BMC Plant Biol.* **7**.
- Companhia Nacional de Abastecimento. Acompanhamento da safra brasileira de grãos. v.5-safra 2017/2018 - n.5 - Quinto levantamento - Fevereiro 2018, p 1-140.
- Dash, S. et al.** (2016). Legume information system (LegumeInfo.org): A key component of a set of federated data resources for the legume family. *Nucleic Acids Res.* **44**: D1181–D1188.
- Edgar, R.C.** (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**: 1792–1797.
- Edgar, R.C. and Batzoglou, S.** (2006). Multiple sequence alignment. *Curr. Opin. Struct. Biol.* **16**: 368–373.
- Finn, R.D. et al.** (2017). InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.* **45**: D190–D199.
- Finn, R.D. et al.** (2016). The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Res.* **44**: D279–D285.
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R.D., and Bairoch, A.** (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**: 3784–3788.
- Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., and Rokhsar, D.S.** (2012). Phytozome: A comparative platform for green plant genomics. *Nucleic Acids Res.* **40**: 1178–1186.

- Hashimoto, K. and Kudla, J.** (2011). Calcium decoding mechanisms in plants. *Biochimie* **93**: 2054–2059.
- Horton, P., Park, K.J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C.J., and Nakai, K.** (2007). WoLF PSORT: Protein localization predictor. *Nucleic Acids Res.* **35**: 585–587.
- Hu, B., Jin, J., Guo, A.Y., Zhang, H., Luo, J., and Gao, G.** (2015). GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics* **31**: 1296–1297.
- Kawahara, Y. et al.** (2013). Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* **6**: 4.
- Khoury, C.K., Castañeda-Alvarez, N.P., Achicanoy, H.A., Sosa, C.C., Bernau, V., Kassa, M.T., Norton, S.L., van der Maesen, L.J.G., Upadhyaya, H.D., Ramírez-Villegas, J., Jarvis, A., and Struik, P.C.** (2015). Crop wild relatives of pigeonpea [*Cajanus cajan* (L.) Millsp.]: Distributions, ex situ conservation status, and potential genetic resources for abiotic stress tolerance. *Biol. Conserv.* **184**: 259–270.
- Kudla, J., Batistic, O., and Hashimoto, K.** (2010). Calcium Signals: The Lead Currency of Plant Information Processing. *Plant Cell* **22**: 541–563.
- Kumar, S., Stecher, G., and Tamura, K.** (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* **33**: 1870–1874.
- Lamesch, P. et al.** (2012). The Arabidopsis Information Resource (TAIR): Improved gene annotation and new tools. *Nucleic Acids Res.* **40**: 1202–1210.
- Letunic, I. and Bork, P.** (2017). 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res.*: 1–4.
- McCormack, E. and Braam, J.** (2003). Calmodulins and related potential calcium sensors of Arabidopsis. *New Phytol.* **159**: 585–598.
- Mohanta, T.K., Kumar, P., and Bae, H.** (2017). Genomics and evolutionary aspect of calcium signaling event in calmodulin and calmodulin-like proteins in plants. *BMC Plant Biol.* **17**: 38.
- Odeny, D.A.** (2007). The potential of pigeonpea (*Cajanus cajan* (L.) Millsp.) in Africa. *Nat. Resour. Forum* **31**: 297–305.
- Pazhamala, L.T., Purohit, S., Saxena, R.K., Garg, V., Krishnamurthy, L., Verdier, J.,**

- and Varshney, R.K.** (2017). Gene expression atlas of pigeonpea and its application to gain insights into genes associated with pollen fertility implicated in seed formation. *J. Exp. Bot.* **68**: 2037–2054.
- Rao, S.S., El-Habbak, M.H., Havens, W.M., Singh, A., Zheng, D., Vaughn, L., Haudenschild, J.S., Hartman, G.L., Korban, S.S., and Ghabrial, S.A.** (2014). Overexpression of GmCaM4 in soybean enhances resistance to pathogens and tolerance to salt stress. *Mol. Plant Pathol.* **15**: 145–160.
- Sanders, D., Brownlee, C., and Harper, J.** (1999). Communicating with calcium. *Plant Cell* **11**: 691–706.
- Sang Hyoung Lee, Jong Cheol Kim, Mal Soon Lee, Won Do Heo, Hae Young Seo, Hae Won Yoon, Jong Chan Hong, Sang Yeol Lee, Jeong Dong Bahk, Hwang, I., and Moo Je Cho** (1995). Identification of a novel divergent calmodulin isoform from soybean which has differential ability to activate calmodulin-dependent enzymes. *J. Biol. Chem.* **270**: 21806–21812.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J.** (1994). ClustalW: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res. Acids Res* **22**: 4673–4680.
- Varshney, R.K. et al.** (2011). Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. *Nat. Biotechnol.* **30**: 83.
- Zhao, Y., Liu, W., Xu, Y.-P., Cao, J.-Y., Braam, J., and Cai, X.-Z.** (2013). Genome-wide identification and functional analyses of calmodulin genes in Solanaceousspecies. *BMC Plant Biol.* **13**: 1–15.