Genetica (2011) 139:255–259
DOI 10.1007/s10709-010-9544-y

The evolutionary history of calreticulin and calnexin genes in green plants

Luiz Eduardo V. Del Bem

Received: 26 May 2010 / Accepted: 27 December 2010 / Published online: 11 January 2011
© Springer Science+Business Media B.V. 2011

Abstract Calreticulin and calnexin are Ca^{2+}-binding chaperones localized in the endoplasmic reticulum of eukaryotes acting in glycoprotein folding quality control and Ca^{2+} homeostasis. The evolutionary histories of calreticulin and calnexin gene families were inferred by comprehensive phylogenetic analyses using 18 completed genomes and ESTs covering the major green plants groups, from green algae to angiosperms. Calreticulin and calnexin possibly share a common origin, and both proteins are present along all green plants lineages. The calreticulin founder gene within green plants duplicated in early tracheophytes leading to two possible groups of orthologs with specialized functions, followed by lineage-specific gene duplications in spermatophytes. Calnexin founder gene in land plants was inherited from basal green algae during evolution in a very conservative copy number. A comprehensive classification in possible groups of orthologs and a catalog of calreticulin and calnexin genes from green plants are provided.

Keywords Calreticulin · Calnexin · Chaperones · Evolution · Green plants

Electronic supplementary material The online version of this article (10.1007/s10709-010-9544-y) contains supplementary material, which is available to authorized users.

L. E. V. Del Bem (✉)
Centro de Biologia Molecular e Engenharia Genética,
Universidade Estadual de Campinas (UNICAMP),
Av. Cândido Rondon 400–Cidade Universitária,
CP 6010, Campinas CEP 13083-875, Brazil
e-mail: lev.del.bem@gmail.com

Introduction

The highly conserved eukaryotic Ca^{2+}-binding proteins calreticulin (CRT) and calnexin (CNX) are the central players in the so-called CRT/CNX cycle of glycoprotein folding quality control. CRT possesses a C-terminal domain with a (K/H)DEL endoplasmic reticulum (ER) retrieval signal (Michalak et al. 2009), while CNX is a ER membrane-bound protein (Jin et al. 2009). CRT/CNX cycle is part of the N-glycan-dependent quality control mechanism that takes place in the ER lumen (Hammond et al. 1994). Glycan processing starts with its transfer to Asn residues in nascent proteins within ER. Several subsequent glycosyl hydrolysis exposes the $Glc_1Man_9GlcNAc_2$ epitope that is then recognized by CRT and CNX that specifically bind monoglucosylated polymannose glycans (Ware et al. 1995; Caramelo and Parodi 2008).

CNX and CRT were first described in plants in 1993 and 1998, respectively (Huang et al. 1993; Crofts and Denecke 1998). Plant's CRTs were further classified into two groups of homologs, CRT1/2 and CRT3, which were initially thought to be resulting from a gene duplication event occurring before the divergence between monocots and eudicots (Persson et al. 2003). A recent work (Jin et al. 2009) suggested that CRT3 group is present in basal land plants and *Arabidopsis* CRT3-specific function on the retention of defective brassinosteroid receptor EFR in the ER, which is a specific function of plant's CRT3 without functional overlapping with CRT1 and 2 (Christensen et al. 2010). EFR accumulation and signaling are impaired in *Arabidopsis crt3* mutant, affecting the immune response to the bacterial epitope elf18 (Saijo et al. 2009) suggesting a role for CRT3 in bacterial pathogen-associated molecular pattern (PAMP), while CRT1 and 2 are possibly involved

in more general chaperone functions (Li et al. 2009; Christensen et al. 2010). Plant's CRTs were also implicated in several physiological processes such as virus defense (Chen et al. 2005), ER calcium buffering (Persson et al. 2001; Christensen et al. 2010), plasmodesma cell–cell transport (Baluska et al. 1999; Laporte et al. 2003), and stress response and tolerance (Jia et al. 2008). In the following sections, I will show detailed phylogenetic analyses of *CRT* and *CNX* genes in Viridiplantae kingdom using 18 completed genomes and ESTs from diverse lineages such as green algae, basal non-vascular and vascular land plants, gymnosperms and angiosperms.

Methods

Comparative sequence analyses

Predicted proteomes for *Arabidopsis thaliana*, version 8.0–http://www.arabidopsis.org; *Arabidopsis lyrata*, version 1.0–http://genomeportal.jgi-psf.org/Araly1/Araly1.home.html; *Populus trichocarpa*, version 1.1–http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html; *Glycine max*, version 1.0–http://www.phytozome.net/soybean.php; *Ricinus communis*, version 0.1–http://castorbean.jcvi.org/downloads.php; *Oryza sativa*, version 5.0–http://rice.plantbiology.msu.edu; *Sorghum bicolor*, version 1.4–http://genome.jgi-psf.org/Sorbi1/Sorbi1.home.html; *Selaginella moellendorffii*, version 1.0–http://genome.jgi-psf.org/Selmo1/Selmo1.home.html; *Physcomitrella patens patens*, version 1.1–http://genome.jgi-psf.org/Phypa1_1/Phypa1_1.home.html; *Volvox carteri*, version 1.0–http://genome.jgi-psf.org/Volca1/Volca1.home.html; *Chlamydomonas reinhardtii*, version 4.0–http://genome.jgi-psf.org/Chlre4/Chlre4.home.html; *Ostreococcus lucimarinus*, version 2.0–http://genome.jgi-psf.org/Ost9901_3/Ost9901_3.home.html; *Ostreococcus tauri*, version 2.0–http://genome.jgi-psf.org/Ostta4/Ostta4.home.html; *Ostreococcus sp.* RCC809, version 2.0–http://genome.jgi-psf.org/OstRCC809_2/OstRCC809_2.home.html; *Micromonas pusilla* CCMP1545, version 2.0–http://genome.jgi-psf.org/MicpuC2/MicpuC2.home.html; *Micromonas sp.* RCC299, version 2.0–http://genome.jgi-psf.org/MicpuN2/MicpuN2.home.html; *Chlorella vulgaris*, version 1.0–http://genome.jgi-psf.org/Chlvu1/Chlvu1.home.html and *Chlorella sp.* NC64A–http://genome.jgi-psf.org/ChlNC64A_1/ChlNC64A_1.home.html were downloaded and pooled together (Viridiplantae 3.0–530.234 sequences). I performed Hidden Markov Model (HMM) searches using HMMER3 software (http://hmmer.janelia.org/) against Viridiplantae 3.0 in order to identify possible CRT and CNX homologs with an e-value threshold of e^{-5}. I used two different alignments as queries in two independent HMM searches, one of them containing *Arabidopsis* CRT1, 2 and 3 protein sequences and the other containing *Arabidopsis* CNX1 and CNX2 protein sequences.

Maize cDNA sequences were obtained from MAGI (http://magi.plantgenomics.iastate.edu/), and ESTs used in this study were downloaded from TIGR Plant Transcript Assemblies (http://plantta.jcvi.org/) and included 202.387 assembled ESTs (unisequences) from *Ceratopteris richardii* (4.492), *Cycas rumphii* (4.335), *Ginkgo biloba* (4.178), *Marchantia polymorpha* (10.721), *Picea abies* (5.204), *Picea glauca* (49.412), *Picea sitchensis* (25.425), *Pinus pinaster* (13.067), *Pinus taeda* (78.873), and *Welwitschia mirabilis* (6.680). ESTs and cDNAs presenting less than 30% of protein query coverage were discarded.

Phylogenetic analyses

All significantly similar sequences found by HMM were automatically recovered using an in-house algorithm (Del Bem and Vincentz 2010) and manually checked. Sequences were aligned using MAFFT 6.717b (Katoh and Toh 2008) under L–INS–i parameters, and all gaps were removed. Three phylogenetic methods were used to infer the presented trees. Neighbor joining (NJ; Saitou and Nei 1987) using PAM 001 matrix to calculate the genetic distances (Dayhoff et al. 1978) and maximum parsimony (MP; Eck and Dayhoff 1966) under default parameters, both conducted in MEGA 4.0 software (Tamura et al. 2007). Bayesian analyses using the Markov chain Monte Carlo technique were performed using MrBayes3 software (Ronquist and Huelsenbeck 2003) under WAG model of protein substitution (Whelan and Goldman 2001). The Bayesian log-likelihood scores were found to stabilize after 10.000 generations. Therefore, I discarded the initial 10.000 generation trees and sampled one out of every 100 generations from the remaining 0.99 million generations (9,900 trees) to calculate posterior probabilities of each branch. The alignments used contained 329 and 338 amino acids of CNXs and CRTs homologs, respectively. The consensus trees presented in Fig. 1a, b and Supplemental Figs. 1 and 2 were constructed with the three different analyses by *consense* software from PHYLIP suite (Felsenstein 1989) using the majority-rule consensus tree method. The resulting consensus trees were drawn with MEGA4 tree display tool (Tamura et al. 2007). PoGOs were defined as described in Del Bem and Vincentz (2010).

Results and discussion

CRT and CNX possess two conserved domains: the calcium-binding P-domain 'InterPro IPR009033' (http://www.ebi.ac.uk/interpro/IEntry?ac=IPR009033) and the lectin-like N-Domain 'InterPro IPR018124' (http://www.

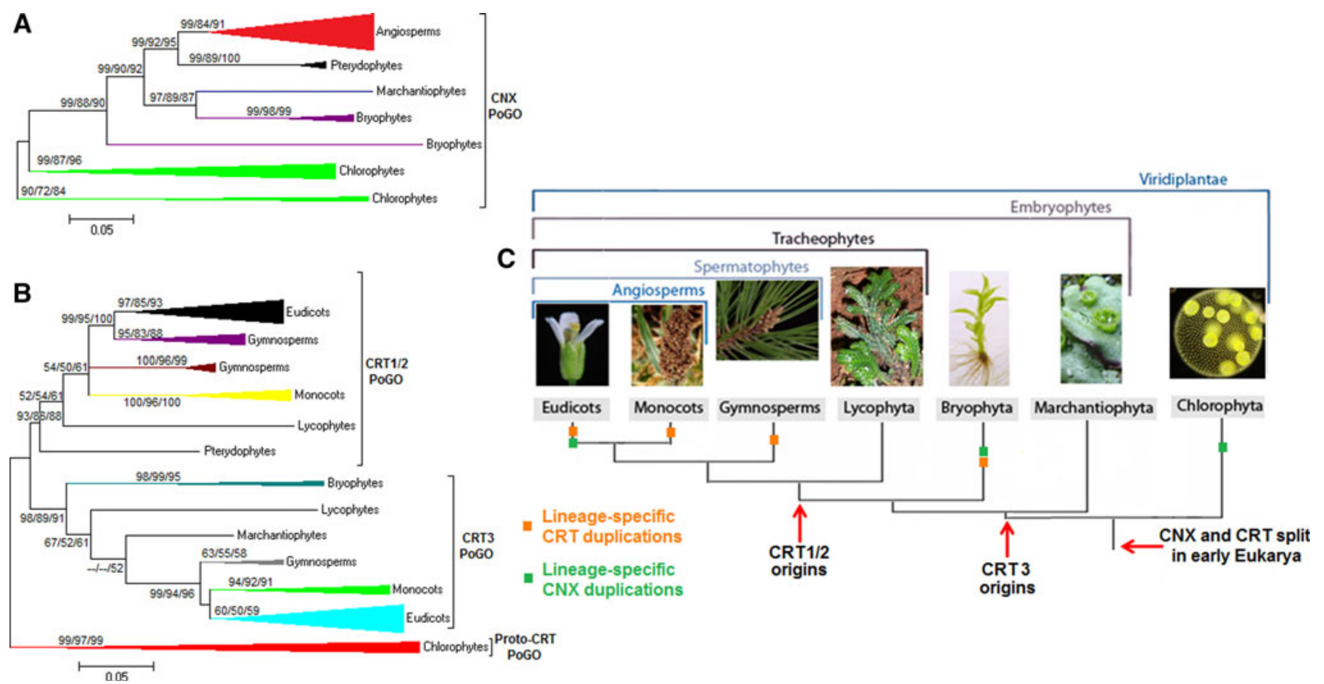

Fig. 1 Phylogenetic trees and evolutionary profile of *CNX* and *CRT* genes in green plants. **a** Phylogenetic tree showing the evolutionary relationship between plants *CNXs*. Tree topology is a consensus from NJ, MP, and Bayesian analyses. Bootstrap values and posterior probabilities from the original trees higher than 50% are shown (NJ/MP/Bayesian). *Triangles* represent compacted groups of orthologs that appear in detail in Supplemental Fig. 1. **b** Phylogenetic tree showing the evolutionary relationship between plants *CRTs*. Tree topology is a consensus from NJ, MP, and Bayesian analyses. Bootstrap values and posterior probabilities from the original trees higher than 50% are shown (NJ/MP/Bayesian). *Triangles* represent compacted groups of orthologs that appear in detail in Supplemental Fig. 2. **c** Evolutionary profile of *CNX* and *CRT* genes in green plants. The *arrows mark* duplication events shared along the descendent lineages, and the *squares mark* lineage-specific duplication events

ebi.ac.uk/interpro/IEntry?ac=IPR018124), which are indicative of a possible common origin. The fact that CRTs and CNXs genes are present along animals and plants along with my results showing the presence of those genes in genomes of green algae such as *Micromonas*, *Volvox*, *Chlorella* and *Ostreococcus* (Fig. 1a, b; Supplemental Figs. 1 and 2; Supplemental Table 1) strongly indicate that they originated by an ancestral gene duplication prior to the divergence between Chlorophyta and Embryophyta. This duplication event could even take place in early eukaryotes.

CNX genes in green plants were further classified in a single possible group of orthologs (PoGO) that integrate genes from Chlorophyta algae to angiosperms (Fig. 1a; Supplemental Table 1; Supplemental Fig. 1). This PoGO generally remained as a single-copy gene in a very diverse taxonomic ranking of green plants such as in the Chlorophyta *Volvox carteri* or even in the monocots sorghum and rice. This observation suggests that single-copy green plant's *CNX* genes probably retained the ancestral eukaryotic function that is thought to be related to glycoprotein folding quality control (Schrag et al. 2001). In contrast, soybean and the moss *Physcomitrella* genomes, probably due to recent large-scale genome duplications, contain four *CNX* paralogs (Supplemental Table 1). I also analysed *Arabidopsis thaliana* (eudicot), sorghum (monocot), *Physcomitrella patens patens* (moss), and *Volvox carteri* (green algae) *CNX* genes for shared intron positions within their coding sequences (Supplemental Fig. 3). This analysis helped support the suggested phylogenetic relationship between green plants *CNX* genes.

Differently from *CNX*, *CRT* genes in Viridiplantae kingdom have diversified specifically in land plants by an ancient event of gene duplication in the last common ancestor of Tracheophyta (Fig. 1b; Supplemental Table 1; Supplemental Fig. 2). While chlorophytes' *CRTs* formed a single PoGO, land plant's *CRTs* were further divided into *CRT1/2* and *CRT3* PoGOs in agreement with the previous literature (Persson et al. 2003). *CRT3* PoGO is embryophyte-specific, which means that genes from this group emerged as a single gene in early land plants' genomes evolving directly from chlorophyte's ancestral single-copy *CRT* gene (PoGO Proto-CRT in Fig. 1b). I interpreted *CRT1/2* PoGO as been derived from a *CRT3* PoGO gene by an ancestral duplication taking place in the last common ancestor of tracheophytes, as evidenced by the presence of a *CRT* gene from *Selaginella moellendorffii* in both PoGOs and the absence of Marchantiophyta and Bryophyta genes in *CRT1/2* PoGO. The proposed phylogenetic classification

is further supported by shared intron positions analysis (Supplemental Fig. 4). However, I cannot fully discard an alternative scenario where *CRT1/2* genes were specifically lost in non-vascular land plants.

Genes belonging to *CRT3* PoGO appear to have evolved specialized functions in land plants when compared to other eukaryotic CRTs, as illustrated by the specific function in retention of defective EFR brassinosteroid receptor in *Arabidopsis thaliana* (Jin et al. 2009; Saijo et al. 2009). In fact, the three *Arabidopsis CRT* genes were shown to be up-regulated by short-term treatments with exogenous glucose and sucrose along with a sugarcane gene belonging to *CRT1/2* group (Papini-Terzi et al. 2009). This finding supports the notion that plant's *CRTs* expression response to sugars is conserved at least in angiosperms. Grasses also appear to have duplicated *CRT1/2* and *CRT3* angiosperms' ancestral genes earlier in their evolution (Fig. 1c; Supplemental Table 1).

My analysis suggests that *CRT1/2* founder gene in Spermatophyta has undergone at least three independent duplication events: one occurring in Pinophyta (as seen for *Pinus taeda* and *Picea sitchensis*), other in grasses (rice and sorghum), and finally in Brassicaceae *Arabidopsis thaliana* and *A. lyrata* (Supplemental Fig. 2; Supplemental Table 1; Fig. 1c). The expression of *CRT1* and *CRT3* from *Arabidopsis thaliana* into calreticulin-deficient (*crt* $^{-}/^{-}$) mouse embryonic fibroblasts restored the wild-type phenotype of ER Ca^{2+}-holding potential and putative chaperone capabilities, suggesting a strong conserved biochemical function of plants and mammals CRTs (Christensen et al. 2008, 2010.

In conclusion, I presented a catalog of *CRT* and *CNX* genes from 18 completed genomes and ESTs from several green plants species (Supplemental Table 1), as well the evolutionary profile of *CRT* and *CNX* genes in the Viridiplantae kingdom (Fig. 1).

Acknowledgment I would like to thank the JGI (www.jgi.doe.gov/) for providing most of the data used in this work.

References

Baluska F, Samaj J, Napier R, Volkmann D (1999) Maize calreticulin localizes preferentially to plasmodesmata in root apex. Plant J 19:481–488

Chen MH, Tian GW, Gafni Y, Citovsky V (2005) Effects of calreticulin on viral cell-to-cell movement. Plant Physiol 138:1866–1876

Christensen A, Svensson K, Persson S, Jung J, Michalak M, Widell S, Sommarin M (2008) Functional characterization of Arabidopsis calreticulin1a: a key alleviator of endoplasmic reticulum stress. Plant Cell Physiol 49:912–924

Christensen A, Svensson K, Thelin L, Zhang W, Tintor N, Prins D, Funke N, Michalak M, Schulze-Lefert P, Saijo Y, Sommarin M, Widell S, Persson S (2010) Higher plant calreticulins have acquired specialized functions in Arabidopsis. PLoS One 5(6):e11342

Crofts AJ, Denecke J (1998) Calreticulin and calnexin in plants. Trends Plants Sci 3:396–399

Dayhoff MO, Schwartz RM, Orcutt BC (1978) A model of evolutionary change in proteins. In: Dayhoff MO (ed) Atlas of protein sequence and structure, vol 5, suppl 3. Natl Biomed Res Found, Washington, pp 345–352

Del Bem LEV, Vincentz MGA (2010) Evolution of xyloglucan-related genes in green plants. BMC Evol Biol 10:341

Eck RV, Dayhoff MO (1966) Atlas of protein sequence and structure. National biomedical research foundation, Silver Springs

Felsenstein J (1989) PHYLIP–Phylogeny inference package (Version 3.2). Cladistics 5:164–166

Hammond C, Braakman I, Helenius A (1994) Role of N-linked oligosaccharide recognition, glucose trimming, and calnexin in glycoprotein folding and quality control. Proc Natl Acad Sci USA 91(3):913–917

Huang L, Franklin AE, Hoffman NE (1993) Primary structure and characterization of an *Arabidopsis thaliana* calnexin-like protein. J Biol Chem 268:6560–6566

Jia XY, Xu CY, Jing RL, Li RZ, Mao XG, Wang JP, Chang XP (2008) Molecular cloning and characterization of wheat calreticulin (CRT) gene involved in drought-stressed responses. J Exp Bot 59:739–751

Jin H, Hong Z, Su W, Li J (2009) A plant-specific calreticulin is a key retention factor for a defective brassinosteroid receptor in the endoplasmic reticulum. PNAS 106(32):13612–13617

Katoh K, Toh H (2008) Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform 9:286–298

Laporte C, Vetter G, Loudes AM, Robinson DG, Hillmer S, Stussi-Garaud C, Ritzenthaler C (2003) Involvement of the secretory pathway and the cytoskeleton in intracellular targeting and tubule assembly of Grapevine fanleaf virus movement protein in tobacco BY-2 cells. Plant Cell 15:2058–2075

Li J, Zhao-Huia C, Batoux M, Nekrasov V, Roux M, Chinchilla D, Zipfel C, Jones JDG (2009) Specific ER quality control components required for biogenesis of the plant innate immune receptor EFR. PNAS 106(37):15973–15978

Michalak M, Groenendyk J, Szabo E, Gold LI, Opas M (2009) Calreticulin, a multi-process calcium-buffering chaperone of the endoplasmic reticulum. Biochem J 417:651–666

Papini-Terzi FS, Rocha FR, Vêncio RZ, Felix JM, Branco DS, Waclawovsky AJ, Del Bem LEV, Lembke CG, Costa MD, Nishiyama MY Jr, Vicentini R, Vincentz MG, Ulian EC, Menossi M, Souza GM (2009) Sugarcane genes associated with sucrose content. BMC Genomics 10:120

Persson S, Wyatt SE, Love J, Thompson WF, Robertson D, Boss WF (2001) The Ca(2+) status of the endoplasmic reticulum is altered by induction of calreticulin expression in transgenic plants. Plant Physiol 126:1092–1104

Persson S, Rosenquist M, Svensson K, Galvão R, Boss WF, Sommarin M (2003) Phylogenetic analyses and expression studies reveal two distinct groups of calreticulin isoforms in higher plants. Plant Physiol 133:1385–1396

Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

Saijo Y, Tintor N, Lu X, Rauf P, Pajerowska-Mukhtar K, Haweker H, Dong X, Robatzek S, Schulze-Lefert P (2009) Receptor quality control in the endoplasmic reticulum for plant innate immunity. EMBO J 28:3439–3449

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4(4):406–425

Schrag JD, Bergeron JJM, Li Y, Borisova S, Hahn M, Thomas DY, Cygler M (2001) The structure of calnexin, an ER chaperone involved in quality control of protein folding. Mol Cell 8(3):633–644

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Ware FE, Vassilakos A, Peterson PA, Jackson MR, Lehrman MA, Williams DB (1995) The molecular chaperone calnexin binds $Glc_1Man_9GlcNAc_2$ oligosaccharide as an initial step in recognizing unfolded glycoproteins. J Biol Chem 270:4697–4704

Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. Mol Biol Evol 18(5):691–699