

RESEARCH ARTICLE

Mix and match: Patchwork domain evolution of the land plant-specific Ca²⁺-permeable mechanosensitive channel MCA

Kanae Nishii^{1,2*}, Michael Möller¹, Hidetoshi Iida^{3*}

1 Royal Botanic Garden Edinburgh, Edinburgh, Scotland, United Kingdom, **2** Kanagawa University, Yokohama-shi, Kanagawa, Japan, **3** Department of Biology, Tokyo Gakugei University, Tokyo, Japan

* kanaenishii@gmail.com (KN); iida@u-gakugei.ac.jp (HI)



Abstract

Multidomain proteins can have a complex evolutionary history that may involve *de novo* domain evolution, recruitment and / or recombination of existing domains and domain losses. Here, the domain evolution of the plant-specific Ca²⁺-permeable mechanosensitive channel protein, MID1-COMPLEMENTING ACTIVITY (MCA), was investigated. MCA, a multidomain protein, possesses a Ca²⁺-influx-MCA^{func} domain and a PLAC8 domain. Profile Hidden Markov Models (HMMs) of domains were assessed in 25 viridiplantae proteomes. While PLAC8 was detected in plants, animals, and fungi, MCA^{func} was found in streptophytes but not in chlorophytes. Full MCA proteins were only found in embryophytes. We identified the MCA^{func} domain in all streptophytes including charophytes where it appeared in E3 ubiquitin ligase-like proteins. Our Maximum Likelihood (ML) analyses suggested that the MCA^{func} domain evolved early in the history of streptophytes. The PLAC8 domain showed similarity to *Plant Cadmium Resistance (PCR)* genes, and the coupling of MCA^{func} and PLAC8 seemed to represent a single evolutionary event. This combination is unique in MCA, and does not exist in other plant mechanosensitive channels. Within angiosperms, gene duplications increased the number of MCAs. Considering their role in mechanosensing in roots, MCA might be instrumental for the rise of land plants. This study provides a textbook example of *de novo* domain emergence, recombination, duplication, and losses, leading to the convergence of function of proteins in plants.

OPEN ACCESS

Citation: Nishii K, Möller M, Iida H (2021) Mix and match: Patchwork domain evolution of the land plant-specific Ca²⁺-permeable mechanosensitive channel MCA. PLoS ONE 16(4): e0249735. <https://doi.org/10.1371/journal.pone.0249735>

Editor: Berthold Heinze, Austrian Federal Research Centre for Forests BFW, AUSTRIA

Received: November 18, 2020

Accepted: March 23, 2021

Published: April 15, 2021

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0249735>

Copyright: © 2021 Nishii et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: R codes used in this study are published in protocol.io. [[dx.doi.org/10.17504/protocols.io.bkqwkvxe](https://doi.org/10.17504/protocols.io.bkqwkvxe)] Gene trees and their alignment are available from TreeBASE. [<http://purl.org/phylo/treebase/phylows/study/TB2:S26880>]

Introduction

Proteins are essential components in any biological organism, including plants. Each protein can be assembled from smaller units, termed domains, and a protein can consist of a single or multiple domains [1]. There exist several databases for the repository of protein domains found in biological organisms [2]. Pfam, for example, currently has 19,179 entries ([3]; Pfam v.34.0, released March 2021). During organismal evolution, protein domains can combine but also evolve *de novo*. These *de novo* domains can be further combined with other *de novo* or existing domains to create novel proteins [1]. During plant evolution, it has been suggested

The seed alignment of the MCA_{func} domain is available in the new March 2021 release, v. 34.0, of Pfam [<http://pfam.xfam.org/family/PF19584#tabview=tab0>], as well as in the S19 Appendix of this study.

Funding: This work was supported by the Japan Society for the Promotion of Science (JSPS) [KAKENHI Grant Number 25120708], Ministry of Education, Culture, Sports, Science & Technology of Japan, to HI. KN's stay at RBGE is financially supported by the Edinburgh Botanic Garden Sibbald Trust [2018#18], JSPS [JSPS KAKENHI Grant Number 18K06375], and the Sumitomo Foundation [170204].

Competing interests: We declare that we have no competing interests.

that at least 500 novel protein domains unique to this evolutionary lineage have emerged [4]. A search of *Arabidopsis thaliana* proteomes suggested that 75% of its proteins have domains registered in Pfam [5]. This indicates that there still exist a significant amount of unknown protein domains or domain combinations even in well studied plants, let alone plants in general. The combination of domains is perhaps a cost-effective way for organisms to create novel proteins [1], and in *A. thaliana*, at least 25% of proteins have multiple domains [5].

Integral membrane proteins that mediate ion fluxes in response to mechanical stresses, including touch, wind, water flow, osmotic pressure, gravity, and cell division- and cell expansion-generated forces, are called mechanosensitive channels. To date, five groups of mechanosensitive channels are found in plants [6]. One of them is a group of MID1-COMPLEMENTARY ACTIVITY (MCA) proteins, which are shown to function as Ca²⁺-permeable mechanosensitive channels [7, 8]. The genes encoding MCAs are found exclusively in the plant kingdom [7, 9], whereas genes encoding other groups of mechanosensitive channels are found in prokaryotes and/or eukaryotes. Therefore, MCAs are unique in terms of molecular evolution and it is interesting to investigate when and where the MCA genes appeared during plant evolution.

In *A. thaliana*, two paralogous MCA genes, *AtMCA1* and *AtMCA2* have been isolated, and their functions examined in great detail. The *AtMCA1* protein is involved in touch sensing at the root tip and a hypoosmotic shock-induced increase in the cytosolic free Ca²⁺ concentration [7]. *AtMCA2* was reported to participate in Ca²⁺ uptake at the roots [10]. In addition, *AtMCA1* and *AtMCA2* respond to membrane stretch to generate cation currents when expressed in *Xenopus laevis* oocytes [8]. Furthermore, MCA channels appear to have common functions in plants, based on studies on *Oryza sativa* OsMCA1 [11–13], *Nicotiana tabacum* NtMCA1, NtMCA2 [14], *Zea mays* CNR13 [15], and *Streptocarpus* MCA-like gene (as *Saint-paulia* in [16]; see [17]).

MCAs are approximately 420 amino acid (aa) residues long multidomain proteins. They retain the provisionally advocated ARPK domain (Amino-terminal domain of Rice putative Protein Kinases; 1–143 aa) [7], overlapping with the EF hand-like region at the N-terminal region (136–180 aa) (InterPro: IPR002048), and well-curated PLAC8 domain (Pfam ID: PF04749) at the C-terminal region (S1 Appendix). A coiled-coil motif is located in the middle of the proteins. An approximately 170 aa region at the N-terminus, covering the ARPK and the EF hand-like domains, has Ca²⁺ influx activity and is proposed to be a functional domain of MCAs [18]. In this study, we defined the N-terminal region as the MCA functional (MCA^{func}) domain.

In previous work, an MCA Neighbor-Joining tree was published that included only a limited number of plants, *i.e.* one moss, one lycophyte, one gymnosperm, and eight angiosperms. The unrooted tree showed that MCA proteins were mostly grouped following the tree of life (*e.g.* tolweb.org/tree/), except for *Picea sitchensis* (gymnosperm) and *Linum usitatissimum* (angiosperm) [9]. However, information from this tree is insufficient to elucidate the evolutionary history of the protein family or their domains. To better understand the origin and evolution of MCA proteins in plants, a more comprehensive study is required. Thus, in the present study, wide-ranging phylogenetic analyses of MCA proteins were carried out on 25 viridiplantae proteomes and full MCA proteins of 55 streptophyte species. Here, for ranks, we followed the definition by Leliaert et al. [19] and NCBI Taxonomy Browser (<https://www.ncbi.nlm.nih.gov/guide/taxonomy/>), where viridiplantae include green algae (chlorophytes) and streptophytes, streptophytes include charophytes and embryophytes, and embryophytes (also termed as “land plants”) include bryophytes (Hornworts, Liverworts, Mosses), lycophytes, ferns, gymnosperms and angiosperms. Since MCA is a multidomain protein, we focused on the evolution, origin and fate of each domain (MCA^{func} and PLAC8) as well as the full MCA protein. Comprehensive domain searches were carried out against the viridiplantae proteomes

that included two chlorophytes and two charophytes. The study represents an example for the evolutionary dynamics of a multidomain protein in plants.

Materials and methods

Proteomes, genome, and transcriptomes used in this study

Twenty-five proteomes including species ranging from chlorophytes to angiosperms, were downloaded from Uniprot (<https://www.uniprot.org/>) and plaza (<https://bioinformatics.psb.ugent.be/plaza/versions/gymno-plaza/>) (S2 Appendix). Genomes / transcriptomes of 55 plant species were explored to find the full MCA genes (S3 Appendix; KEGG: [20]; Phytozome: <https://phytozome.jgi.doe.gov/pz/portal.html#>; OneKP: [21]; NCBI Genome: <https://www.ncbi.nlm.nih.gov>; Fernbase: <https://www.fernbase.org>; EnsemblPlants: <https://plants.ensembl.org>). Recently, systematic studies returned the genus *Physcomitrella* to the genus *Physcomitrium* [22], but we used the name *Physcomitrella* in this study for consistency with the registered names in the databases. The proteome completeness information, *i.e.* BUSCO completeness values (BUSCO-C) were available for most taxa on the Uniprot database. The BUSCO-C values of proteomes from plaza database (*Cycas micholitzii*, *Taxus baccata*) were newly obtained in this study using BUSCO v.4.0.6 [23], by comparisons against viridiplantae_odb10 lineage datasets.

Building profile Hidden Markov Models (HMMs)

Three profile HMMs were used in this study: for the full MCA protein the model in PANTHER (<http://www.pantherdb.org/>), 'PROTEIN MID1-COMPLEMENTARY ACTIVITY 1 (MCA1): PTHR46604.SF3.pir.hmm' (422 aa) was used. A new profile HMM was created with *hmmbuild* in HMMER v.3.1b2 package (<http://hmmerr.org/>), for the 1–167 aa region of the putative MCA^{func} domain (MCA^{func}.hmm; 167 aa). MCA^{func}.hmm was registered in Pfam v.34.0 (PF19584). The profile HMM of PLAC8, was obtained from Pfam v.33.1 (PLAC8.hmm: PF04749; 91 aa). Logos of the profile HMMs were generated with Skyline (<http://www.skyline.org>) (S4 and S5 Appendices).

Building domain matrices

Proteomes were interrogated for the presence of the MCA^{func} and PLAC8 domains using *hmmsearch* (HMMER package), and the default setting (*E* value < 10.00). In these, proteins with the 'full E-value' < 0.001 and > 30 aa homologous regions were kept for further analyses. The domain sequences were aligned with MCA^{func}.hmm and PLAC8.hmm, using *hmmalign*, respectively. The alignments were manually checked and corrected in BioEdit v.7.2.5 [24]. They were further trimmed to remove hypervariable regions with BMGE v.1.12 [25] on the Galaxy server (<https://galaxy.pasteur.fr/>).

The proteome of *M. polymorpha* subsp. *ruderalis* (UP000077202) did not include proteins with both domains, but the closely related *M. polymorpha* did. The full MCA sequence was found in the genome database of *M. polymorpha* subsp. *ruderalis* (NCBI Genome GCA_001641455.1; Mp_v4; LVLJ01003617.1:83933–90898), and was highly homologous to that in *M. polymorpha* (Phytozome v.12.0: Mapoly0134s0009) (S6 Appendix). Thus, the translated amino acid sequences of the genome region (LVLJ01003617.1:83933–90898) was used as "A0A176VHI1_MARPO*" (S7 and S8 Appendices).

Domain-based phylogenetic analyses

Maximum likelihood (ML) analyses were carried out with PhyML v.3.0 [26] on the ATGC server (www.atgc-montpellier.fr), with Smart Model Selection (SMS) [27]. Tree topology searches using SPR were carried out, and SH-like α LRT values obtained for branch support. ML rapid bootstrapping analyses of 2000 replicates were performed for additional clade support with RAxML v.8 [28], using models selected with ToPALi v.2 [29].

The Phyml trees were examined with Notung v.2.9 [30] for determining the root of the trees. The required species tree for this analysis (S9 Appendix) followed the Tree of Life Web Project (<http://tolweb.org>) and Angiosperm phylogeny website v.14 [31]. The bryophyte relationships followed [32]. For the MCA^{func}, the proteins in the charophyte *K. nitens* were suggested as root (S10 and S11 Appendices). For the PLAC8 domain tree, no strong root position

Table 1. Number of proteins found in proteomes. Result of profile HMM searches of MCA^{func} and PLAC8 domains in proteomes of 25 taxa across viridiplantae. Number of proteins retaining the MCA^{func} or PLAC8 domains (E value $< 10^{-3}$) are listed and arranged following the Tree of Life (see S2, S7, and S8 Appendices).

Vernacular name	ID	Taxon	No of proteins with	
			MCA ^{func}	PLAC8
Chlorophytes	CHLRE	<i>Chlamydomonas reinhardtii</i>	0	12
Chlorophytes	VOLCA	<i>Volvox carterii</i> f. <i>nagariensis</i>	0	7
Charophytes	KLENI	<i>Klebsormidium nitens</i>	2	13
Charophytes	CHABU	<i>Chara braunii</i>	8	12
Bryophytes	MARPO	<i>Marchantia polymorpha</i>	9	13
Bryophytes	MapoRu	<i>Marchantia polymorpha</i> subsp. <i>ruderalis</i>	9	7
Bryophytes	PHYPA	<i>Physcomitrella patens</i>	9	23
Lycophytes	SELML	<i>Selaginella moellendorffii</i>	17	15
Gymnosperm	CMI	<i>Cycas micholitzii</i>	1	12
Gymnosperm	TBA	<i>Taxus baccata</i>	7	10
Angiosperm	AMBTC	<i>Amborella trichopoda</i>	7	9
Angiosperm	MUSAM	<i>Musa acuminata</i> subsp. <i>malaccensis</i>	9	26
Angiosperm	ORYSJ	<i>Oryza sativa</i> subsp. <i>japonica</i>	29	23
Angiosperm	MAIZE	<i>Zea mays</i>	11	25
Angiosperm	SORBI	<i>Sorghum bicolor</i>	13	19
Angiosperm	AQUCA	<i>Aquilegia coerulea</i>	10	10
Angiosperm	VITVI	<i>Vitis vinifera</i>	7	20
Angiosperm	POPTR	<i>Populus trichocarpa</i>	12	26
Angiosperm	MEDTR	<i>Medicago truncatula</i>	10	22
Angiosperm	CUCSA	<i>Cucumis sativus</i>	5	10
Angiosperm	GOSRA	<i>Gossypium raimondii</i>	14	29
Angiosperm	BRAOL	<i>Brassica oleracea</i> var. <i>oleracea</i>	7	29
Angiosperm	ARATH	<i>Arabidopsis thaliana</i>	5	20
Angiosperm	ERYGU	<i>Erythranthe guttata</i>	9	27
Angiosperm	SOLLC	<i>Solanum lycopersicum</i>	7	19
	<i>sum</i>		217	438

The number of MCA^{func} domain proteins within a proteome varied between species. In charophytes, *Klebsormidium nitens* had two proteins, but *Chara braunii* eight. In angiosperms, monocots possessed generally higher numbers between nine and 29, whereas dicots five to 14 per species. The number of PLAC8 domain proteins was between seven and 30 per species, the lowest in the liverwort *Marchantia polymorpha* subsp. *ruderalis* and the highest in *Brassica oleracea* var. *oleracea*. The more complete liverwort proteome of *Marchantia polymorpha* (BUSCO 96.7% in UniProt) had 14 PLAC8 genes. The low number in *M. polymorpha* subsp. *ruderalis* (BUSCO 91.3%) might be explained by the incompleteness of its proteome. The full MCA protein with both of MCA^{func} and PLAC8 domains was not found in charophytes. In streptophytes, at least one and up to three full MCA proteins were found per species.

<https://doi.org/10.1371/journal.pone.0249735.t001>

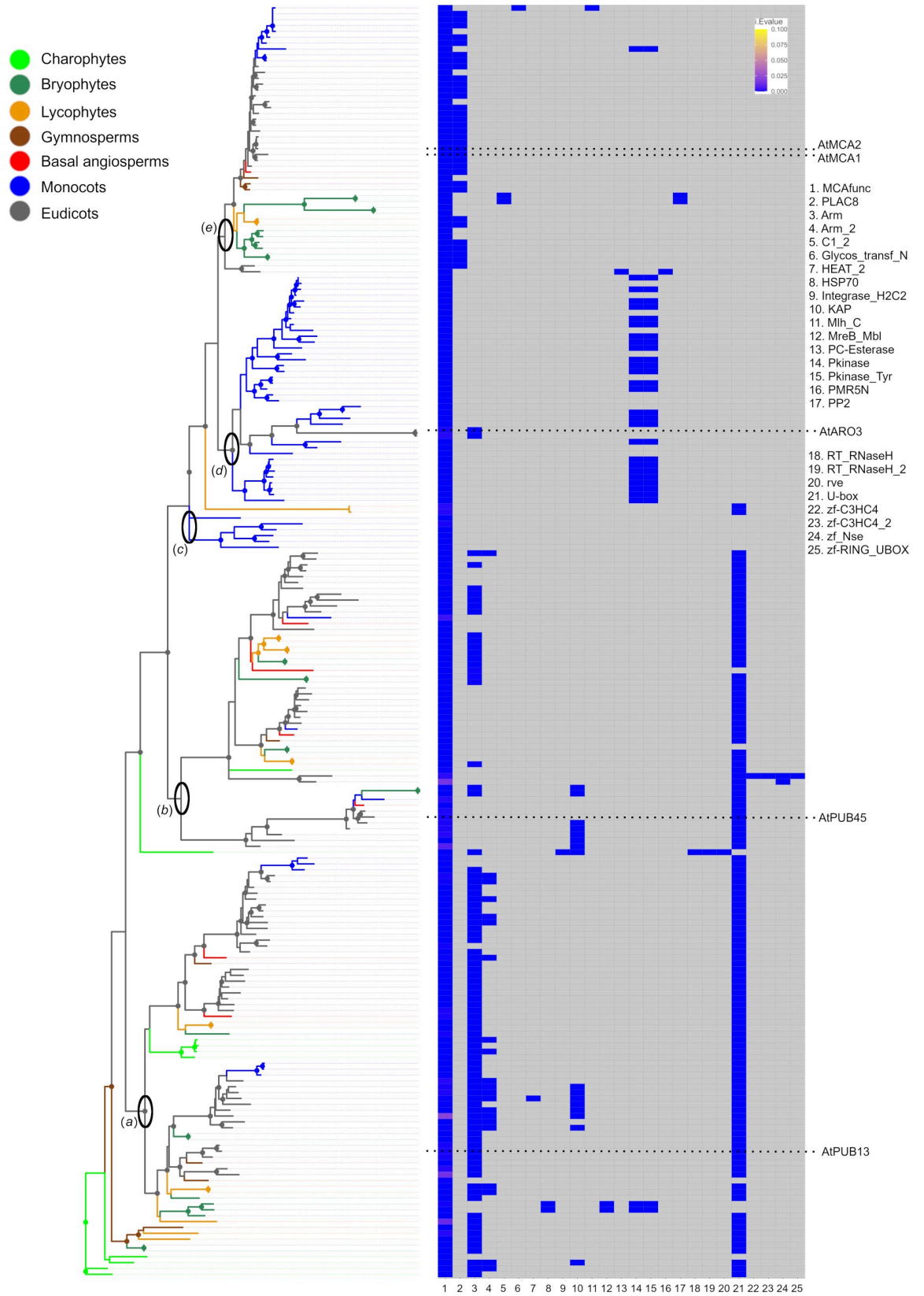


Fig 1. ML tree of MCA^{func} domains and their partner domains predicted by HMMER website searches. Left: ML tree. (a) Clade associated with E3 ubiquitin ligase AtPUB13. (b) Clade associated with AtPUB45. (c) MCA^{func} only proteins. (d) Clade associated with AtARO3 and monocot U-box containing protein kinase like proteins. (e) MCA clade. Clades supported with α -LRT SH-like values > 0.8 indicated with circles at the nodes. Right: Domain individual *E* values (i.Evalue) resulting from HMMER website searches are shown as a heatmap. Absence of domains indicated in grey. The original tree and heatmap are shown in [S11](#) and [S16](#) Appendices. Numbers in legends to the right indicate the domain names (x-axis of the heatmap).

<https://doi.org/10.1371/journal.pone.0249735.g001>

was indicated ([S12 Appendix](#)), and thus a midpoint rooted tree was shown for ease of visualization ([S13 Appendix](#)).

Partner domain HMMER searches of MCA^{func} and PLAC8 domains

The retained proteins possessed either MCA^{func} or PLAC8, or both domains. To determine the exact domain composition of these proteins, they were searched against the Pfam with *hmmsearch* in HMMER (<https://www.ebi.ac.uk/Tools/hmmer/>). Based on their *E*-values they were visualized through R [33], as a colour-coded rooted phylogeny and heatmap utilizing *ggplot2* [34], *ggtree* [35], *ape* v.5.0 [36], and *phytools* [37]. In some cases where two closely related domains were predicted for the same genome position, or domain duplications were involved, the domains with the lowest *E* values were selected. The data was also used for the schematic illustrations of representative domain structures visualized by R with a modified script based on Brennan (<https://rforbiochemists.blogspot.com/2015/11/drawing-protein-domain-structure-using-r.html>).

Phylogeny analysis of full MCA protein sequences

BLAST searches were carried out on plant genome and transcriptome databases using the AtMCA1 protein sequence. The found sequences were further evaluated using *hmmsearch* with MCA^{func}.hmm and PLAC8.hmm. Only genes possessing both domains were included in the phylogenetic analysis. The positions of MCA sequences in genomes were examined where it was possible, and only one transcript sequence involved, e.g. in *Selaginella moellendorffii*, one MCA genome sequence found, whereas two identical proteins are present in the proteome (UP000001514). Thus, only one MCA from *S. moellendorffii* was included in the analyses. A phylogenetic tree was built with *Phyml* v.3.0, and subjected to Notung analyses for rooting. The bryophytes were suggested as likely root ([S14 Appendix](#)).

Because the study focussed on MCA, we specifically analysed gene duplication events for the full MCA protein tree in a Notung reconcile analysis (tree rearranged with Edge Weight Threshold = 0.6). The species tree used here ([S15 Appendix](#)) followed the Angiosperm phylogeny website v.14 [31]. The bryophyte relationship followed [32]. The relationships within angiosperm followed [38], for Brassicaceae [39], for Fabaceae [40], and for Poaceae [41].

Results

MCA^{func} domain found in streptophytes, MCA^{func}+PLAC8 in land plants

To determine the distribution of MCA proteins in viridiplantae, 25 proteomes (see [Table 1](#)) were interrogated for domains of the MCA protein, MCA^{func} and PLAC8, with profile HMMs using HMMER. In total, 217 proteins were found possessing only the MCA^{func} domain, 438 with only the PLAC8 domain, and 32 possessing both domains ([Table 1](#); [S7](#) and [S8](#) Appendices). The MCA^{func} domain was only present in streptophytes, whereas the PLAC8 domain was found in all proteomes examined in this study ([Table 1](#)).

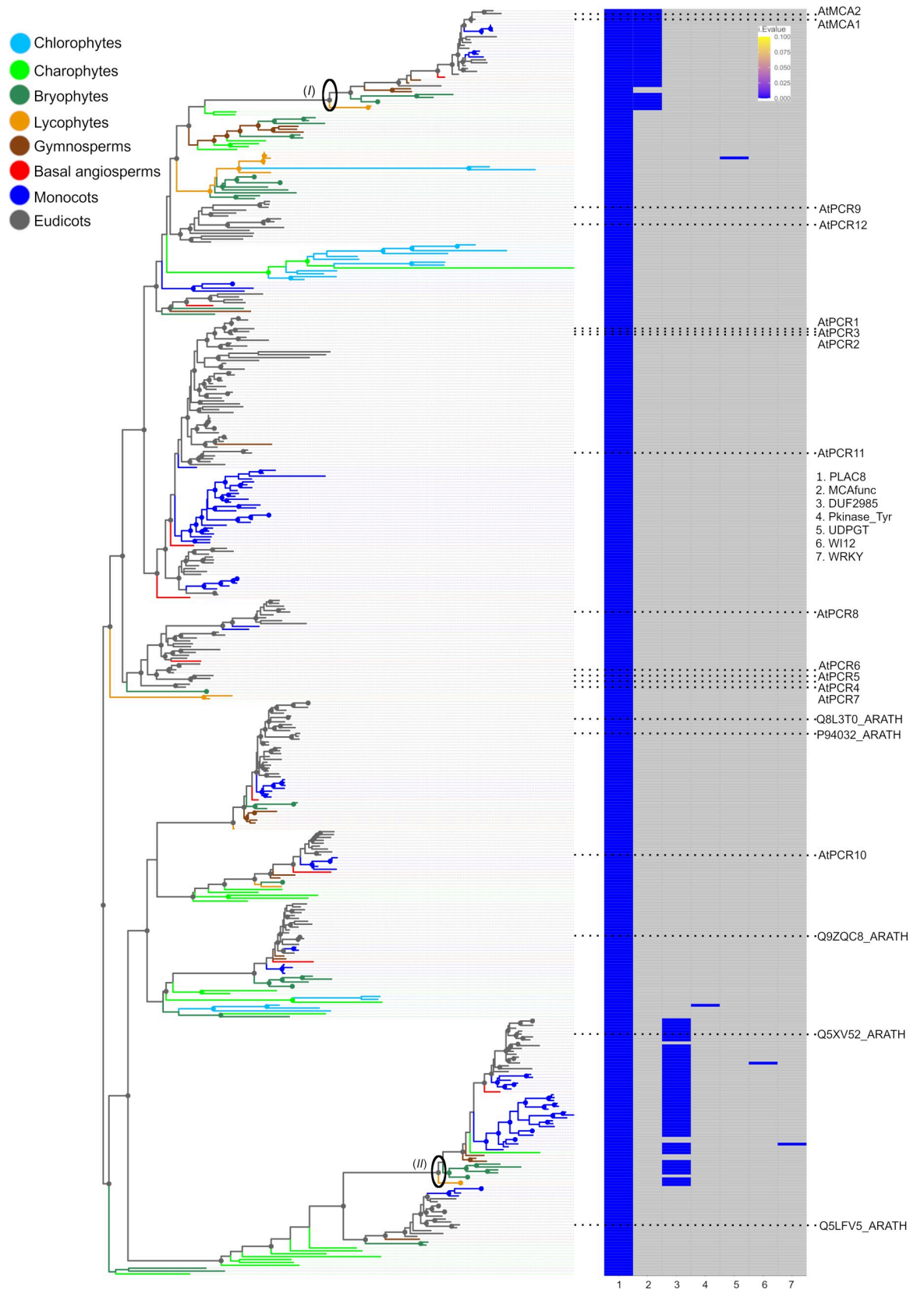


Fig 2. PLAC8 domain tree and their partner domains. Left: PLAC8 domain ML tree. (I) MCA ‘Clade I’. (II) ‘Clade II’ with proteins retaining DUF2985 + PLAC8. Clades supported with α -LRT SH-like support values > 0.8 indicated with circles at the node. Right: Domain combinations observed in the PLAC8 domain containing proteins associated with the ML tree (left). Domain individual *E* values (i.Evalue) resulting from HMMER website searches are shown as a heatmap. Absence of domains indicated in grey. The original tree and heatmap are shown in S13 and S17 Appendices. Numbers in legend to the right indicate the domain names (x-axis of the heatmap).

<https://doi.org/10.1371/journal.pone.0249735.g002>

MCA^{func} and PLAC8 domain phylogenies

Since MCA is a multidomain protein, we studied the phylogenetic relationships of the domains MCA^{func} and PLAC8 separately. In the MCA^{func} domain Maximum Likelihood (ML) tree of 217 domain sequences, the samples included clustered according to the presence of partner domains (Fig 1; S11 and S16 Appendices). For example, samples of charophytes and ‘Clade a’, that included AtPUB13 (RING-type E3 ubiquitin ligase), had U-box (PF04564.15), Arm (PF00514.23) or Arm_2 (PF04826.13) as partner domains to MCA^{func} (Fig 1). Arm and Arm_2 are overlapping domains. ‘Clade a’ contained two major clades each including all streptophyte lineages, suggesting a gene duplication. ‘Clade b’ (Fig 1; S11 Appendix) also contained two main clades including most streptophyte lineages, suggesting a further duplication, where most proteins in one clade had lost the Arm domain. The following clades ‘c’ and ‘d’ contained mostly monocot-specific undescribed or potential protein kinase proteins (Fig 1) (e.g. rice Q2QZY3). ‘Clade e’, is the MCA protein clade including AtMCA1 and AtMCA2, where the majority of MCA^{func} domain proteins were partnered with the PLAC8 domain, which suggested that MCA as the derived proteins. A few proteins scattered across ‘Clade e’ had lost PLAC8 (Fig 1E), but there always was at least one protein with MCA^{func} plus PLAC8 present in each species (Table 1). Some MCA proteins had obtained an alternative partner domain such as C1_2 (PF03107.16) and PP2 (PF14299.6) (*M. polymorpha*), Pkinase (PF00069.25), or Pkinase-Tyr (PF07714.17) (*M. acuminata*) C1_2 and Mlh1_C (PF16413.5) (*Z. mays*).

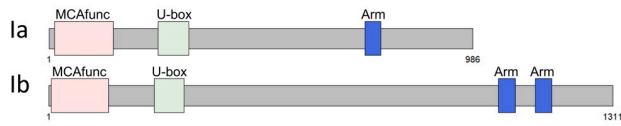
In *Zea mays*, A0A1D6PNG8 and A0A1D6F850 hold the protein name “MCA1” in UniProt, but they in ‘Clade d’ and also lacked PLAC8 but retained Pkinase or Pkinase_Tyr. On the other hand, CNR13 and A0AD6JP06 were found to be proper MCAs since they were in the MCA clade (‘Clade e’) and possessed PLAC8 (S11 and S16 Appendices), as previously reported [15].

The phylogeny of the other MCA domain, PLAC8, was also examined phylogenetically. In the PLAC8 domain ML tree of 438 domain sequences, the samples also clustered according to their partner domains (Fig 2; S13 and S17 Appendices). Most of PLAC8 domain proteins appeared as single domain proteins, but the MCA clade (‘Clade I’) retained MCA^{func}, while another clade (‘Clade II’) retained a DUF2985 (PF11204) domain (Fig 2; S13 and S17 Appendices) with unknown functions. In *A. thaliana*, PLAC8 single domain proteins are registered as “Plant Cadmium Resistance proteins (PCR)”, with the function to reduce cadmium uptake [42]. The MCA clade appeared to be closely related to a clade including AtPCR9 and AtPCR12. All proteins in Clade I, except for two, retained MCA^{func} and PLAC8. In gymnosperms, two proteins per species were found, with one having MCA^{func} while the other lacked it. The ML tree topology and distribution of partner domains suggested that the coupling between MCA^{func} and PLAC8 domains occurred once in the plant lineage, possibly in the common ancestor of embryophytes, and was sometimes lost after gene duplication events but was always retained in at least one copy.

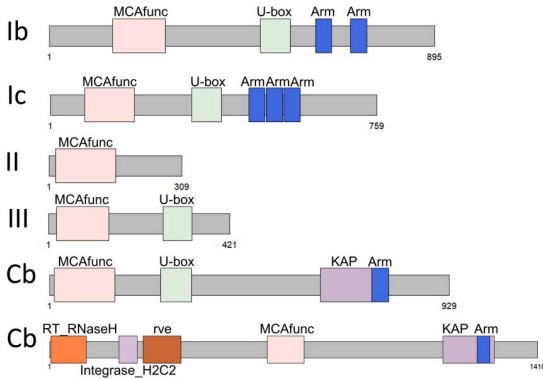
U-box and Arm are original partners of the MCA^{func} domain

While the PLAC8 domain commonly existed within the plant, animal, and fungi kingdoms, the MCA^{func} domain was only observed in streptophytes in the plant kingdom. Thus, the

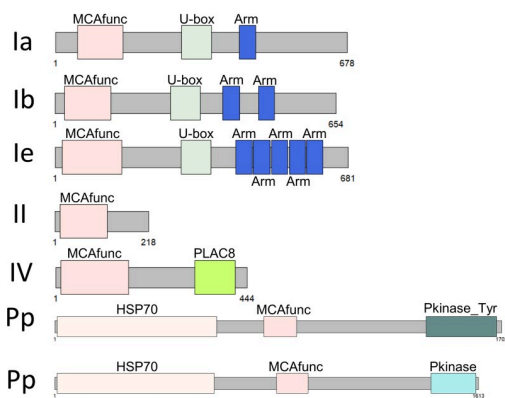
Charophytes: *Klebsormidium nitens*



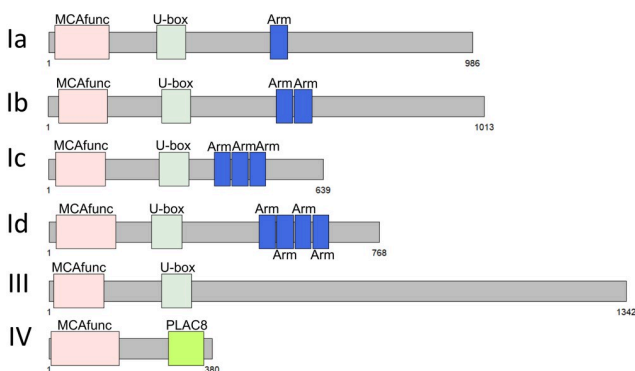
Charophytes: *Chara braunii*



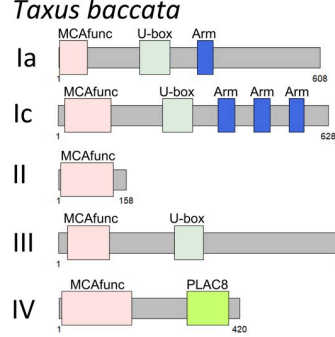
Mosses: *Physcomitrella patens*



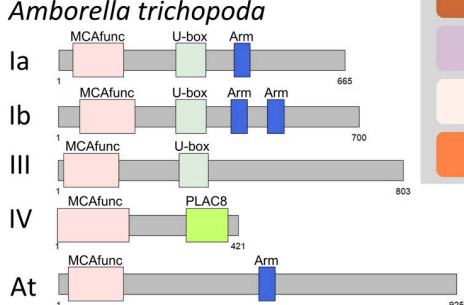
Lycophytes: *Selaginella moellendorffii*



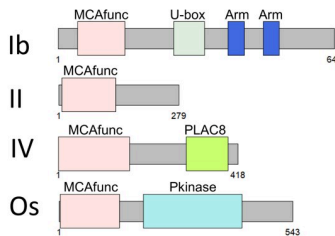
Gymnosperms: *Taxus baccata*



Basal angiosperms: *Amborella trichopoda*



Monocots: *Oryza sativa*



Eudicots: *Arabidopsis thaliana*

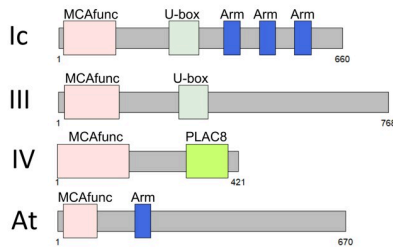


Fig 3. Schematic illustration of protein domain combinations of the MCA^{func} domain proteins in representative proteomes. Type I, MCA^{func} + U-box + Arm, number of Arm copies varies from 1 to 5 (Ia to Ie). Type II, MCA^{func} only. Type III, MCA^{func} + U-box. Type IV, MCA^{func} + PLAC8. The well-studied MCA proteins in *A. thaliana*, AtMCA1 and AtMCA2, are type IV proteins. Lineage specific combinations are indicated by initials of the species, Cb, Pp, Os, At respectively.

<https://doi.org/10.1371/journal.pone.0249735.g003>

Table 2. Domain partners of MCA^{func} domain and their combinations found in proteomes across viridiplantae.

ID	Taxon	Domain combinations (Types)						sum
		I	II	III	IV	Os	At	
CHLRE	<i>Chlamydomonas reinhardtii</i>	0	0	0	0	0	0	0
VOLCA	<i>Volvox carteri f. nagariensis</i>	0	0	0	0	0	0	0
KLENI	<i>Klebsormidium nitens</i>	2	0	0	0	0	0	2
CHABU	<i>Chara braunii</i>	4	1	1	0	0	0	6
MARPO	<i>Marchantia polymorpha</i>	4	1	1	1	0	0	7
MapoRu	<i>Marchantia polymorpha</i> subsp. <i>ruderalis</i>	4	2	0	*1	0	0	7
PHYPA	<i>Physcomitrella patens</i>	3	2	0	2	†2	0	9
SELML	<i>Selaginella moellendorffii</i>	12	0	3	2	0	0	17
CMI	<i>Cycas micholitzii</i>	0	0	0	1	0	0	1
TBA	<i>Taxus baccata</i>	3	2	1	1	0	0	7
AMBTC	<i>Amborella trichopoda</i>	3	0	3	1	0	1	8
MUSAM	<i>Musa acuminata</i> subsp. <i>malaccensis</i>	1	0	4	3	1	0	9
ORYSJ	<i>Oryza sativa</i> subsp. <i>japonica</i>	1	14	0	1	7	0	23
MAIZE	<i>Zea mays</i>	1	2	1	2	4	0	10
SORBI	<i>Sorghum bicolor</i>	1	2	0	1	9	0	13
AQUCA	<i>Aquilegia coerulea</i>	4	0	2	1	0	0	7
VITVI	<i>Vitis vinifera</i>	3	0	3	1	0	0	7
POPTR	<i>Populus trichocarpa</i>	6	0	2	2	0	0	10
MEDTR	<i>Medicago truncatula</i>	4	0	2	2	0	0	8
CUCSA	<i>Cucumis sativus</i>	2	1	0	1	0	0	4
GOSRA	<i>Gossypium raimondii</i>	10	0	2	1	0	0	13
BRAOL	<i>Brassica oleracea</i> var. <i>oleracea</i>	3	0	0	3	0	1	7
ARATH	<i>Arabidopsis thaliana</i>	1	0	1	2	0	1	5
ERYGU	<i>Erythranthe guttata</i>	3	1	3	2	0	0	9
SOLLC	<i>Solanum lycopersicum</i>	4	0	1	1	0	0	6
sum		79	28	30	32	21	3	

The result of HMMER searches of MCA^{func} domain partners and their combinations are listed and arranged following the Tree of Life. The types of domain combinations are described as follows: Type I: MCA^{func} + U-box + Arm/Arm_2, Type II: MCA^{func} only, Type III: MCA^{func} + U-box, Type IV: MCA^{func} + PLAC8 (MCA protein type), Os (monocot type): MCA^{func} + Pkinase/Pkinase_Tyr. † -with HSP70, At (ARO3 type): MCA^{func} + Arm
 *—not found in the proteome but in the genome (see S3, S6–S8 Appendices).

<https://doi.org/10.1371/journal.pone.0249735.t002>

MCA^{func} domain might be the key domain for the MCA protein, and we further assessed the coupling of the MCA^{func} domain with its partner domains. The predicted domain combination for *K. nitens* (charophyte) was MCA^{func} + U-box + Arm (type I) (Fig 3). Type I was found in all species, but the number of Arm domains varied from one to five. In *C. braunii* (charophyte), in addition to type I, MCA^{func} only (type II) and MCA^{func} + Ubox (type III) and lineage specific types (Fig 3Cb) were found. In *P. patens* (bryophyte-moss), type I, type II, the MCA type (MCA^{func} + PLAC8: type IV), and also lineage specific types were found (Fig 3Pp). The proteome of ferns was not available, but type I to type IV, and lineage specific combinations were widely observed from lycophytes to angiosperms. The *O. sativa* MCA^{func} + Pkinase (Os) type was widely present in angiosperm monocots, and in the moss *P. patens*, possessing an additional HSP70 domain (Pp). The *A. thaliana* MCA^{func} + Arm: ARO3 (At) type was only observed in the angiosperms *A. thaliana*, *Brassica oleracea*, and *Amborella trichopoda* (Fig 3; Table 2).

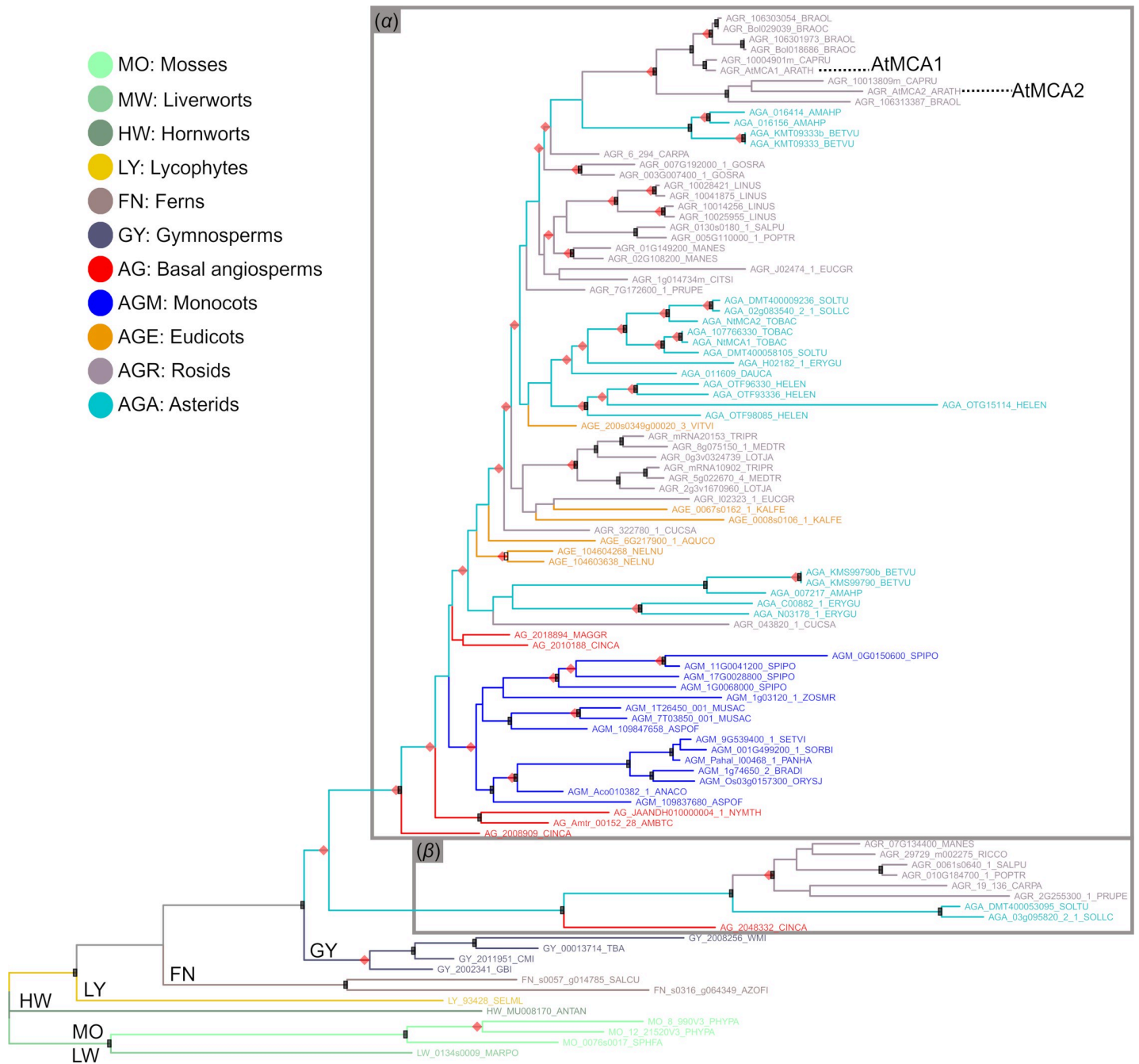


Fig 4. MCA protein ML tree, rooted on the proteins of bryophytes. Proteins have diverged into two clades in angiosperms (α , β). (α) Main MCA clade with all angiosperm species. (β) Diversified MCA clade showing variation at the N-terminus (see Fig 5). Black squares at the nodes indicate high values in both Phylml α LRT (> 0.7) and RaxML ($> 70\%$) clade support analyses. Red diamonds at the nodes indicate Notung-inferred duplication events.

<https://doi.org/10.1371/journal.pone.0249735.g004>

Full MCA protein phylogeny, duplication and diversification in land plants

In order to unravel the history of MCA proteins in plants, a phylogeny of 106 full MCA proteins from 55 embryophyte species was reconstructed. The full MCA proteins include both MCA^{func} and PLAC8 domain sequences. In this analysis, the basal grades in the ML tree, from bryophytes to gymnosperms followed the tree of life relationships (Fig 4). The angiosperm

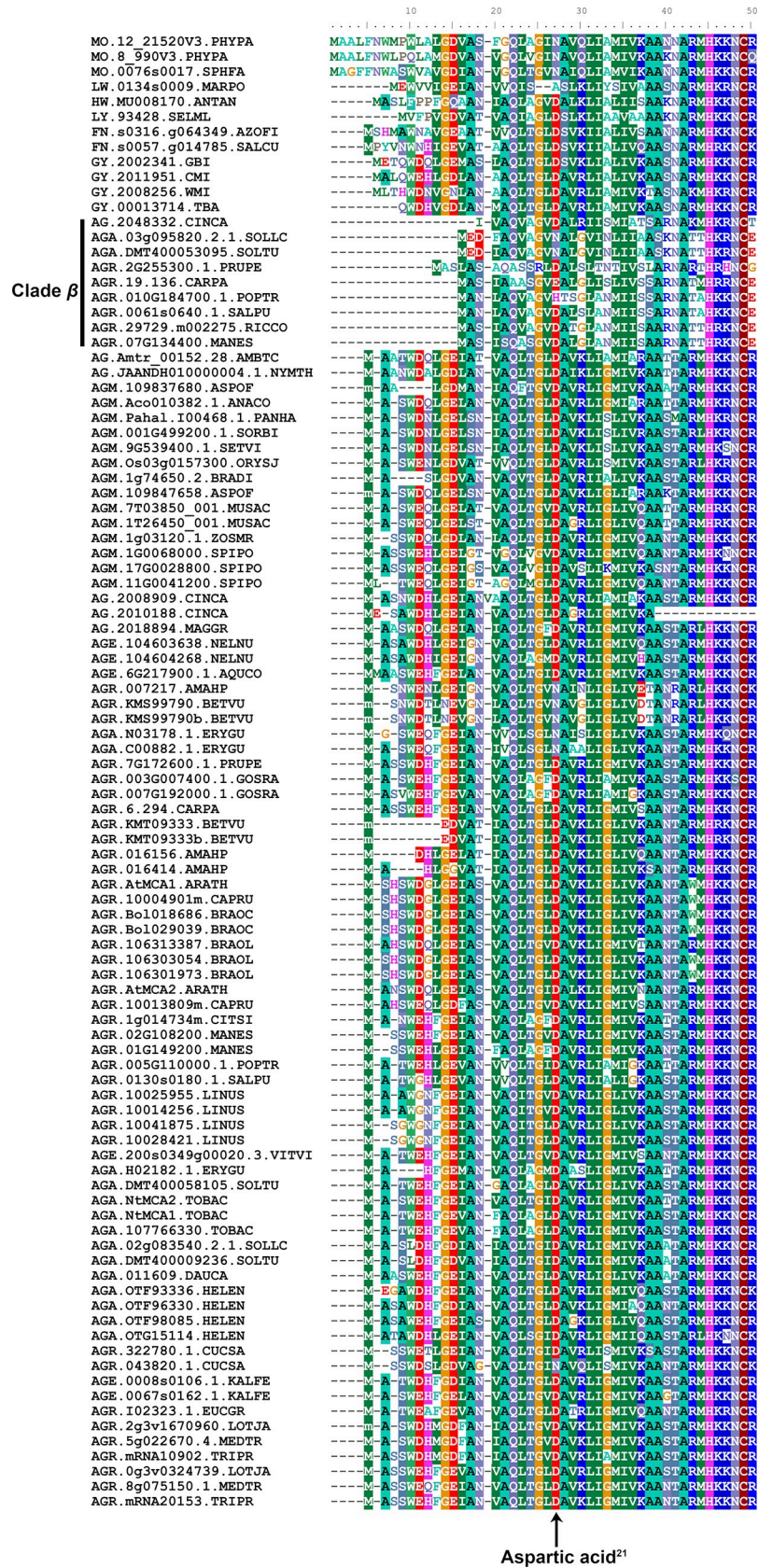


Fig 5. Conservation and variations in the N-terminus region of MCA proteins. The gene IDs are listed in [S3 Appendix](#). The functional site of the 21st position of Aspartic acid (Aspartic acid²¹, arrow) in AtMCA1 (“AGR. AtMCA1.ARATH” in the Fig 5) was well conserved within the MCA proteins, except in mosses and liverworts. Variation in the N-terminus region was observed within angiosperm species, while at least one MCA protein kept the Aspartic acid²¹. The angiosperm clade β MCA proteins lack approximately 10 amino acids in the N-terminal region.

<https://doi.org/10.1371/journal.pone.0249735.g005>

MCAs formed two clades. The majority of proteins fell in ‘Clade α ’ including proteins of all angiosperm species analysed in this study. Only nine proteins formed ‘Clade β ’, representing the orders Laureales (*Cinnamomum camphora*), Malpighiales, Rosales, Solanales and Brassicales (*Carica papaya*). These showed an MCA diversification and lacked approximately 10 aa in the N-terminal region (Fig 5). The predicted functional site of AtMCAs, the 21st position of aspartic acid (Asp²¹; Fig 5, arrow) [18], was different in mosses (asparagine) and liverworts (alanine). Hornworts, on the other hand, retained Asp²¹. At least one MCA per species retained Asp²¹ from lycophytes to angiosperms (Fig 5).

A maximum of 39 duplication events were estimated across the ML tree, with two outside angiosperms (Fig 4, [S18 Appendix](#)). Three duplication events were inferred prior or at the point of diversification of angiosperms. Within the angiosperms, duplications were scattered among the lineages, but the superrosid clade stood out with an accumulation of six inferred duplications events. For several species repeated duplication events were inferred, e.g. three in *Linum usitatissimum* and two in *Beta vulgaris* and *Helianthus annuus* ([S18 Appendix](#)).

Discussion

The evolution of multidomain proteins can be complex, and may involve *de novo* domain evolution, recruitment of existing domains as partners, and recombination and domain losses [1]. In the present study where the evolution of the multidomain protein MCA was examined in detail, the results showed that it represents an example with a complex evolutionary history.

Our comprehensive proteome interrogation with profile HMMs suggested that the MCA^{func} domain [18], formerly subscribed as ARPK domain plus EF hand-like [7], is a well conserved domain among plants. Accordingly, MCA can be described as a multidomain protein composed of the MCA^{func} and the PLAC8 domains. PLAC8 is widely observed in eukaryotes as seen in our profile HMM searches, in which we found it in all proteomes we examined (Table 1). The MCA^{func} domain, on the other hand, was streptophyte-specific and not found in chlorophytes, suggesting that the domain originated in the common ancestor of streptophytes, i.e. charophytes plus embryophytes (Fig 6) [4].

The MCA^{func} evolution further included domain recruitment, recombination and losses. The E3 ubiquitin ligase-type proteins (type I in Fig 3) found in charophytes, represent an ancestral combination (Table 1). Type I proteins were found in most streptophytes, except *Cycas micholitzii* possibly due to the incompleteness of its proteome. The ancestral charophyte *K. nitens* retained only the type I, while in the more derived *C. braunii*, MCA^{func} obtained different partner domains or lost them all (Fig 3). Although domain-losses need to be seen with caution in some species included here due to their proteome incompleteness, single MCA^{func} domain proteins were also observed in well-assembled genomes such as *P. patens* and *O. sativa*, supporting the existence of single-domain MCA^{func} proteins (Fig 1 and Fig 3). Lineage-specific domain combinations were also observed in angiosperms, such as the Os- and At-types. Intriguingly, the At-type was only found in Brassicaceae and *A. trichopoda*, but perhaps due to unrelated parallel evolutionary events (Table 1).

A key event for the MCA evolution seemed to be the partnering of MCA^{func} and PLAC8 first recruited in the common ancestor of embryophytes. MCA is seemingly streptophyte-specific and might play some basic roles, perhaps as a mechanosensor, for habit expansion to terra

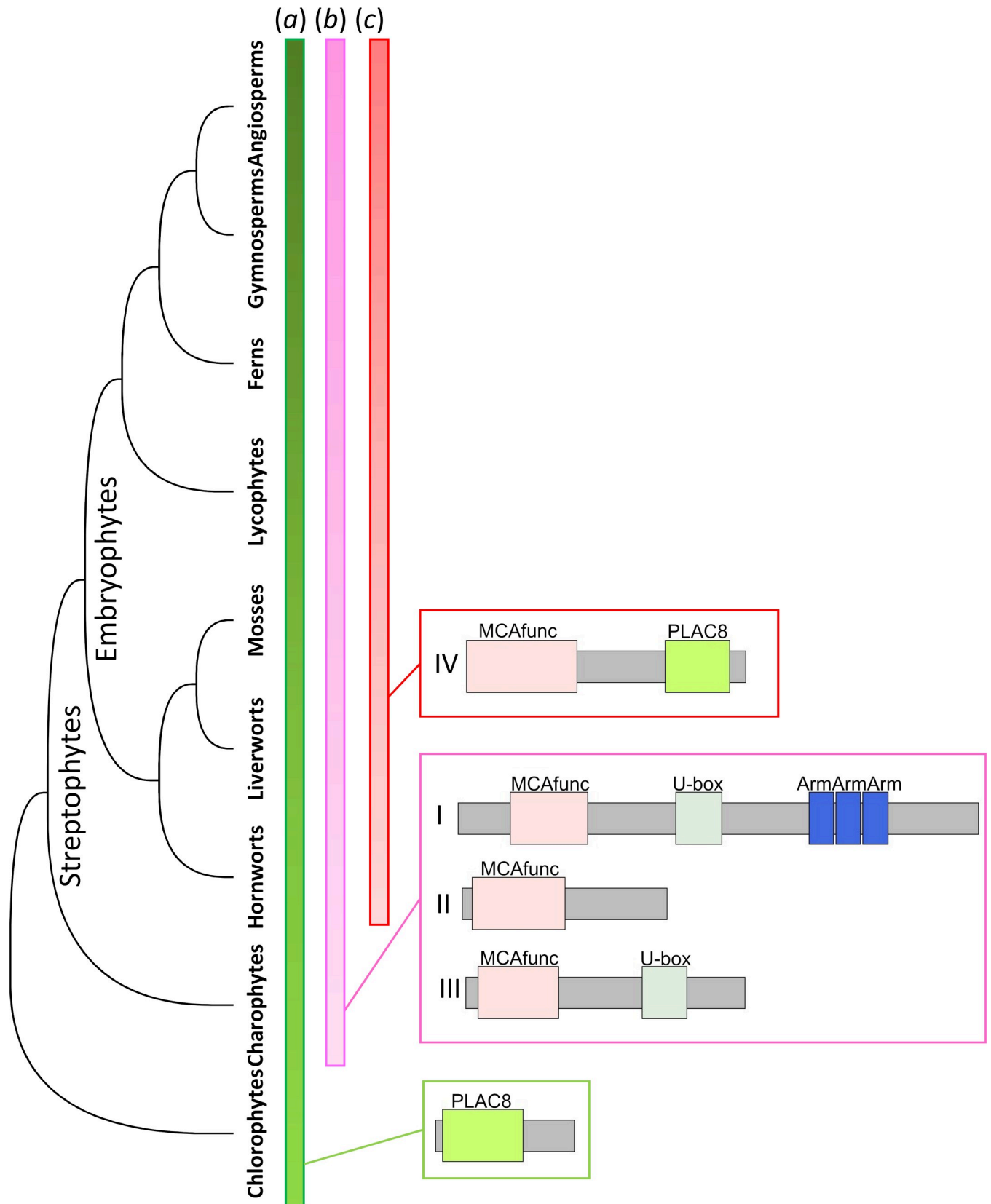


Fig 6. Schematic illustration of the MCA protein evolution. PLAC8 domain genes were observed from chlorophytes to angiosperms, and many were single domain proteins (a). In the common ancestor of streptophytes, type I, type II, and type III domain combinations evolved (b). In the common ancestor of land plants, *i.e.* embryophytes, type IV domain combinations (MCA type, *i.e.* MCA^{func} + PLAC8) evolved (c). In addition to the commonly observed types I-IV, lineage specific combinations were also observed.

<https://doi.org/10.1371/journal.pone.0249735.g006>

firma [9]. A previous study indicated that the Asp²¹ of MCA^{func} domain is crucial for Ca²⁺ uptake [18]. Since Asp²¹ is diversified in mosses and liverworts, it could be hypothesized that their MCAs do not have Ca²⁺ uptake function. In angiosperms, MCA diverged into two clades and one might have changed functions from proper MCA. However, further studies would be needed to support these hypotheses.

Intriguingly, the E3 ubiquitin ligase type, with MCA^{func} + U-box + Arm, seems to represent the most ancient MCA^{func} protein (Fig 1). E3 ubiquitin ligase mediate substrate specificity for ubiquitylation [43] and is a large protein family. In the E3 ubiquitin ligases of *A. thaliana*, only plant U-Box13 (PUB13) and PUB45 retained MCA^{func} domains (Fig 1; S11 and S16 Appendices). PUB13 was suggested to be involved in the abscisic acid signalling pathway, flowering time, and abiotic stress resistance [44]. The expression level of PUB45 seemed to be affected by nutrients [45]. It could be postulated that the MCA^{func} domain was first utilized for these E3 ubiquitin ligases for roles for environmental adaptation, though further studies are required here. In addition, there are proteins only retaining the MCA^{func} domain, but their function is not yet reported and remains unknown (S7 Appendix).

The PLAC8 domain exists in the Plant Cadmium Resistant (PCR) protein family as single domain proteins. PCRs are possibly transmembrane proteins and have roles in cadmium resistance [42] and zinc transport (PCR2; [46]). It is possible that the MCA^{func} domain, initially part of E3 ubiquitin ligase, and PLAC8, an ion transporter, combined at some point in time and resulted in a novel protein, MCA, as a mechanosensor reacting to environmental calcium ions [11, 14]. The sequences between MCA and other plant mechanosensitive channels, such as MSL, are different [9], and the evolutionary history of MCA is different from that of MSL, which has originated in prokaryotes [47], and may represent an example of convergence in function.

Conclusions

In conclusion, MCA is an example of a multidomain protein, whose MCA^{func} domain emerged *de novo* in the ancestor of streptophytes, and recruited an existing domain PLAC8 in the ancestor of embryophytes. The full MCA protein further duplicated and diversified during the evolution of land plants, involving recombination and losses of domains. However, each streptophyte species analysed had at least one complete full MCA copy, pointing to the importance of the protein. The functions of many MCA proteins are not investigated yet but they appear somewhat related to environment sensing, protein-protein interactions, and ion transport. In the basal lineage of streptophytes, *i.e.* charophytes, the MCA^{func} domain is associated with U-box and Arm domains, supposed to play roles in the E3 ubiquitin ligase pathway. On the other hand, MCA proteins with MCA^{func} and PLAC8 domains show quite different roles in ion transport. This further supports a hypothesis where domain swapping is an efficient mechanism to increase protein numbers with diversified functions during organismal evolution. Future studies will shed more light on the roles of these proteins and their interactions in relation to land plant evolution.

Supporting information

S1 Appendix. Proposed domain structure of MCA. (a) For the MCA protein, the ARPK domain, at the N-terminus, and PLAC8, at the C-terminus, was proposed previously, and it

was shown that MCA has an EF hand-like and a coiled-coil region [7]. (b) The biological function of MCA proteins was tested. As a result, the MCA functional domain at the N-terminus, which was previously described as ARPK domain and part of the EF hand-like region, was proposed [18]. This domain was analyzed in the present study and has been registered as MCA^{func} domain in the Pfam database.

(PDF)

S2 Appendix. List of proteomes used in this study.

(PDF)

S3 Appendix. List of genome and transcriptome databases interrogated in this study, and the list of MCA genes included in the MCA phylogenetic analyses.

(PDF)

S4 Appendix. Profile HMM logo of the full MCA protein, PTHR46604:SF3.

(PDF)

S5 Appendix. Profile HMM logos of domains in the MCA protein. (a) MCA^{func}.hmm generated in this study, (b) PLAC8.hmm (PF04749).

(PDF)

S6 Appendix. MCA genes found in the genome of *M. polymorpha* subsp. *ruderalis*. Top: *M. polymorpha* cDNA 0134s0009.1, middle: *M. polymorpha* genome, bottom: *M. polymorpha* subsp. *ruderalis* genome.

(PDF)

S7 Appendix. MCA^{func} domain genes retrieved from proteomes.

(PDF)

S8 Appendix. PLAC8 domain genes retrieved from proteomes.

(PDF)

S9 Appendix. Species tree used for Notung rooting analyses of domain trees.

(PDF)

S10 Appendix. Result of the Notung rooting analysis of the MCA^{func} domain tree. The possible root position is marked in red (arrow).

(PDF)

S11 Appendix. MCA^{func} domain ML tree. (a) Clade associated with E3 ubiquitin ligase AtPUB13. (b) Clade associated with AtPUB45. (c) MCA^{func} only proteins. (d) Clade associated with AtARO3 and monocot U-box containing protein kinase like proteins. (e) MCA clade associated with AtMCA1 and AtMCA2.

(PDF)

S12 Appendix. Result of the Notung rooting analysis of the PLAC8 domain tree. Multiple branches show equally strong estimates as possible root positions (marked in red).

(PDF)

S13 Appendix. PLAC8 domain ML tree. (I) MCA clade (II) Clade with proteins retaining DUF2985 + PLAC8.

(PDF)

S14 Appendix. Notung analyses of the MCA tree. (a) Results of the Notung rooting analyses of the MCA tree. Possible root position is marked in red (arrow). (b) Results of Notung rearrangement of the MCA tree. Rearranged branches are marked in yellow. D: inferred

duplication.
(PDF)

S15 Appendix. Species tree used for Notung analyses for full MCA protein sequence tree.
(PDF)

S16 Appendix. Domain partners of MCA^{func} domain. Domain partners observed in the MCA^{func} domain containing proteins associated with the tree shown in Fig 1 (left). Domain individual *E* values (i.Evalue) resulting from HMMER website searches are shown as heatmap (right). Absence of domains indicated in grey.
(PDF)

S17 Appendix. Domain partners of PLAC8 domain. Domain partners observed in the PLAC8 domain containing proteins associated with the ML tree (left). Domain individual *E* values (i.Evalue) resulting from HMMER website searches are shown as heatmap (right). Absence of domains indicated in grey.
(PDF)

S18 Appendix. MCA gene duplication events estimated by the Notung analysis. “D” indicates inferred gene duplication events.
(PDF)

S19 Appendix. The seed alignment and profile HMM of MCA^{func} domain.
(ZIP)

Acknowledgments

The authors are indebted to Dannie Durand for helpful comments and discussions pertaining to this study, particularly relating to the Notung rooting analyses. The authors also thank Daniel Barker and Frank Wright for helpful discussions. KN is grateful to the following persons for facilitating research associateships, to Pete Hollingsworth and Mark Newman at the Royal Botanic Garden Edinburgh (RBGE), UK, and to Akitoshi Iwamoto at Kanagawa University, Japan. This work was logistically supported by RBGE’s Science and ICT divisions. We also thank Duncan Reddish and Catherine Kidner for facilitating RBGE Linux server access and support, and Iain Milne for organizing access to the CropDiversity server, James Hutton Institute, Dundee, UK. We acknowledge the National Institute of Genetics, Japan, for allowing the use of their NIG-supercomputer system. We thank three anonymous reviewers and the editor for their constructive comments.

Author Contributions

Conceptualization: Kanae Nishii, Hidetoshi Iida.

Data curation: Kanae Nishii, Michael Möller.

Formal analysis: Kanae Nishii, Michael Möller.

Funding acquisition: Kanae Nishii, Hidetoshi Iida.

Investigation: Kanae Nishii, Michael Möller.

Methodology: Kanae Nishii.

Supervision: Hidetoshi Iida.

Writing – original draft: Kanae Nishii, Michael Möller.

Writing – review & editing: Kanae Nishii, Michael Möller, Hidetoshi Iida.

References

1. Buljan M, Bateman A. The evolution of protein domain families. *Biochem Soc Trans*. 2009; 37: 751–755. <https://doi.org/10.1042/BST0370751> PMID: 19614588
2. Bagowski CP, Bruins W, Te Velthuis AJ. The nature of protein domain evolution: shaping the interaction network. *Curr Genomics*. 2010; 11: 368–376. <https://doi.org/10.2174/138920210791616725> PMID: 21286315
3. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al. The Pfam protein families database in 2019. *Nucl Acids Res*. 2019; 47: D427–D432. <https://doi.org/10.1093/nar/gky995> PMID: 30357350
4. Kersting AR, Bornberg-Bauer E, Moore AD, Grath S. Dynamics and adaptive benefits of protein domain emergence and arrangements during plant genome evolution. *Genome Biol Evol*. 2012; 4: 316–329. <https://doi.org/10.1093/gbe/evs004> PMID: 22250127
5. Zhang XC, Wang Z, Zhang X, Le MH, Sun J, Xu D, et al. Evolutionary dynamics of protein domain architecture in plants. *BMC Evol Biol*. 2012; 12: 6. <https://doi.org/10.1186/1471-2148-12-6> PMID: 22252370
6. Basu D, Haswell ES. Plant mechanosensitive ion channels: an ocean of possibilities. *Curr Opin Plant Biol*. 2017; 40: 43–48. <https://doi.org/10.1016/j.pbi.2017.07.002> PMID: 28750206
7. Nakagawa Y, Katagiri T, Shinozaki K, Qi Z, Tatsumi H, Furuichi T, et al. *Arabidopsis* plasma membrane protein crucial for Ca^{2+} influx and touch sensing in roots. *Proc Natl Acad Sci USA*. 2007; 104: 3639–3644. <https://doi.org/10.1073/pnas.0607703104> PMID: 17360695
8. Furuichi T, Iida H, Sokabe M, Tatsumi H. Expression of *Arabidopsis* MCA1 enhanced mechanosensitive channel activity in the *Xenopus laevis* oocyte plasma membrane. *Plant Signal Behav*. 2012; 7: 1022–1026. <https://doi.org/10.4161/psb.20783> PMID: 22751361
9. Kurusu T, Kuchitsu K, Nakano M, Nakayama Y, Iida H. Plant mechanosensing and Ca^{2+} transport. *Trends Plant Sci*. 2013; 18: 227–233. <https://doi.org/10.1016/j.tplants.2012.12.002> PMID: 23291244
10. Yamanaka T, Nakagawa Y, Mori K, Nakano M, Imamura T, Kataoka H, et al. MCA1 and MCA2 that mediate Ca^{2+} uptake have distinct and overlapping roles in *Arabidopsis*. *Plant Physiol*. 2010; 152: 1284–1296. <https://doi.org/10.1104/pp.109.147371> PMID: 20097794
11. Kurusu T, Nishikawa D, Yamazaki Y, Gotoh M, Nakano M, Hamada H, et al. Plasma membrane protein OsMCA1 is involved in regulation of hypo-osmotic shock-induced Ca^{2+} influx and modulates generation of reactive oxygen species in cultured rice cells. *BMC Plant Biol*. 2012a; 12: 11. <https://doi.org/10.1186/1471-2229-12-11> PMID: 22264357
12. Liu Z, Cheng Q, Sun Y, Dai H, Song G, Guo Z, et al. A SNP in OsMCA1 responding for a plant architecture defect by deactivation of bioactive GA in rice. *Plant Mol Biol*. 2015; 87: 17–30. <https://doi.org/10.1007/s11103-014-0257-y> PMID: 25307286
13. Liang J, He Y, Zhang Q, Wang W, Zhang Z. Plasma membrane Ca^{2+} permeable mechanosensitive channel *OsDMT1* is involved in regulation of plant architecture and ion homeostasis in rice. *Int J Mol Sci*. 2020; 21: 1097. <https://doi.org/10.3390/ijms21031097> PMID: 32046032
14. Kurusu T, Yamanaka T, Nakano M, Takiguchi A, Ogasawara Y, Hayashi T, et al. Involvement of the putative Ca^{2+} -permeable mechanosensitive channels, NtMCA1 and NtMCA2, in Ca^{2+} uptake, Ca^{2+} -dependent cell proliferation and mechanical stress-induced gene expression in tobacco (*Nicotiana tabacum*) BY-2 cells. *J Plant Res*. 2012b; 125: 555–568. <https://doi.org/10.1007/s10265-011-0462-6> PMID: 22080252
15. Rosa M, Abraham-Juárez MJ, Lewis MW, Fonseca JP, Tian W, Ramirez V, et al. 2017 The maize MID-COMPLEMENTING ACTIVITY homolog CELL NUMBER REGULATOR13/NARROW ODD DWARF coordinates organ growth and tissue patterning. *Plant Cell* 2017; 29: 474–490. <https://doi.org/10.1105/tpc.16.00878> PMID: 28254777
16. Ohnishi M, Kadohama N, Suzuki Y, Kajiyama T, Shichijo C, Ishizaki K, et al. Involvement of Ca^{2+} in vacuole degradation caused by a rapid temperature decrease in *Saintpaulia* palisade cells: a case of gene expression analysis in a specialized small tissue. *Plant Cell Physiol*. 2015; 56: 1297–1305. <https://doi.org/10.1093/pcp/pcv048> PMID: 25941231
17. Nishii K, Hughes M, Briggs M, Haston E, Christie F, DeVilliers MJ, et al. *Streptocarpus* redefined to include all Afro-Malagasy Gesneriaceae: Molecular phylogenies prove congruent with geography and cytology and uncovers remarkable morphological homoplasies. *Taxon* 2015; 64: 1243–1274. <https://doi.org/10.12705/646.8>

18. Nakano M, Iida K, Nyunoya H, Iida H. Determination of structural regions important for Ca²⁺ uptake activity in Arabidopsis MCA1 and MCA2 expressed in yeast. *Plant Cell Physiol.* 2011; 52: 1915–1930. <https://doi.org/10.1093/pcp/pcr131> PMID: 21949028
19. Leliaert F, Smith DR, Moreau H, Herron MD, Verbruggen H, Delwiche CF, et al. Phylogeny and molecular evolution of the green algae. *Critical Rev Plant Sci* 2012; 31: 1–46. <https://doi.org/10.1080/07352689.2011.615705>
20. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. *Nucl Acids Res.* 2019; 47: D590–D595. <https://doi.org/10.1093/nar/gky962> PMID: 30321428
21. Leebens-Mack JH, Barker MS, Carpenter EJ, Deyholos MK, et al. One thousand plant transcriptomes and the phylogenomics of green plants. *Nature* 2019; 574: 679–685. <https://doi.org/10.1038/s41586-019-1693-2> PMID: 31645766
22. Medina R, Johnson MG, Liu Y, Wickett NJ, Shaw AJ, Goffinet B. Phylogenomic delineation of *Physcomitrium* (Bryophyta: Funariaceae) based on targeted sequencing of nuclear exons and their flanking regions rejects the retention of *Physcomitrella*, *Physcomitridium* and *Aphanorhagma*. *J Syst Evol.* 2019; 57: 404–417. <https://doi.org/10.1111/jse.12516>
23. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 2015; 31: 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351> PMID: 26059717
24. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acid S.* 1999; 41: 95–98. <https://doi.org/10.1021/bk-1999-0734.ch008>
25. Criscuolo A, Gribaldo S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol.* 2010; 10: 210. <https://doi.org/10.1186/1471-2148-10-210> PMID: 20626897
26. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010; 59: 307–321. <https://doi.org/10.1093/sysbio/syq010> PMID: 20525638
27. Lefort V, Longueville J-E, Gascuel O. SMS: Smart Model Selection in PhyML. *Mol Biol Evol.* 2017; 34: 2422–2424. <https://doi.org/10.1093/molbev/msx149> PMID: 28472384
28. Stamatakis A RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014; 30: 1312–1313. <https://doi.org/10.1093/bioinformatics/btu033> PMID: 24451623
29. Milne I, Lindner D, Bayer M, Husmeier D, McGuire G, Marshall DF, et al. TOPALI v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. *Bioinformatics.* 2009; 25: 126–127. <https://doi.org/10.1093/bioinformatics/btn575> PMID: 18984599
30. Chen K, Durand D, Farach-Colton M. NOTUNG: a program for dating gene duplications and optimizing gene family trees. *J Comput Biol.* 2000; 7: 429–447. <https://doi.org/10.1089/106652700750050871> PMID: 11108472
31. Stevens, P. F. 2001 onwards. Angiosperm Phylogeny Website. Version 14, July 2017 [and more or less continuously updated since] Available from: <http://www.mobot.org/MOBOT/research/APweb/>
32. Puttick MN, Morris JL, Williams TA, Cox CJ, Edwards D, Kenrick P, et al. The interrelationships of land plants and the nature of the ancestral embryophyte. *Curr Biol.* 2018; 28: 733–745 e732. <https://doi.org/10.1016/j.cub.2018.01.063> PMID: 29456145
33. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020. Available from: <https://www.R-project.org/>
34. Wickham H. ggplot2: elegant graphics for data analyses. Springer-Verlag New York. 2016. Available from: <https://ggplot2.tidyverse.org>
35. Yu G-C, Smith D, Zhu H, Guan Y, Tommy Lam T T-Y. ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol.* 2017; 8: 28–36. <https://doi.org/10.1111/2041-210X.12628>
36. Paradis E, Schliep K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics.* 2019; 35: 526–528. <https://doi.org/10.1093/bioinformatics/bty633> PMID: 30016406
37. Revell LJ. phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol.* 2012; 3: 217–223. <https://doi.org/10.1111/j.2041-210X.2011.00169.x>
38. Edger PP, Jocelyn C, Hall JC, Harkess A, Tang M, Coombs J, Mohammadin S, et al. Brassicales phylogeny inferred from 72 plastid genes: A reanalysis of the phylogenetic localization of two paleopolyploid events and origin of novel chemical defenses. *Amer J Bot.* 2017; 105: 463–469. <https://doi.org/10.1002/ajb2.1040> PMID: 29574686

39. Nikolov LA, Shushkov P, Nevado B, Gan X, Al-Shehbaz IA, Filatov D, et al. Resolving the backbone of the Brassicaceae phylogeny for investigating trait diversity. *New Phytol.* 2019; 222: 1638–1651. <https://doi.org/10.1111/nph.15732> PMID: 30735246
40. Choi IS, Ruhlman TA, Jansen RK. Comparative mitogenome analysis of the genus *Trifolium* reveals independent gene fission of *ccmFn* and intracellular gene transfers in Fabaceae. *Int J Mol Sci.* 2020; 21: 6. <https://doi.org/10.3390/ijms21061959> PMID: 32183014
41. Saarela JM, Burke SV, Wysocki WP, Barrett MD, Clark LG, Craine JM, et al. A 250 plastome phylogeny of the grass family (Poaceae): topological support under different data partitions. *PeerJ.* 2018; 6: e4299. <https://doi.org/10.7717/peerj.4299> PMID: 29416954
42. Song WY, Martinoia E, Lee J, Kim D, Kim DY, Vogt E, et al. A novel family of cys-rich membrane proteins mediates cadmium resistance in *Arabidopsis*. *Plant Physiol.* 2004; 135: 1027–1039. <https://doi.org/10.1104/pp.103.037739> PMID: 15181212
43. Metzger MB, Pruneda JN, Klevit RE, Weissman AM. RING-type E3 ligases: master manipulators of E2 ubiquitin-conjugating enzymes and ubiquitination. *Biochim Biophys Acta.* 2014; 1843: 47–60. <https://doi.org/10.1016/j.bbamcr.2013.05.026> PMID: 23747565
44. Liao D, Cao Y, Sun X, Espinoza C, Nguyen CT, Liang Y, et al. Arabidopsis E3 ubiquitin ligase PLANT U-BOX13 (PUB13) regulates chitin receptor LYSIN MOTIF RECEPTOR KINASE5 (LYK5) protein abundance. *New Phytol.* 2017; 214: 1646–1656. <https://doi.org/10.1111/nph.14472> PMID: 28195333
45. Pradhan M. Understanding the regulatory basis of defense signalling in plants: The role of *Argonautes* in modulating defense responses in *Nicotiana attenuate*. PhD Dissertation Friedrich-Schiller- Universität Jena. 2019.
46. Song WY, Choi KS, Kim DY, Geisler M, Park J, Vincenzetti V, et al. Arabidopsis PCR2 is a zinc exporter involved in both zinc extrusion and long-distance zinc transport. *Plant Cell* 2010; 22: 2237–2252. <https://doi.org/10.1105/tpc.109.070185> PMID: 20647347
47. Cox CD, Nakayama Y, Nomura T, Martinac B. The evolutionary ‘tinkering’ of MscS-like channels: generation of structural and functional diversity. *Pflugers Arch.–Eur. J. Physiol.* 2014; 467: 3–13. <https://doi.org/10.1007/s00424-014-1522-2> PMID: 24819593