PCP
PLANT & CELL PHYSIOLOGY

Calcium-Dependent Protein Kinases in Plants: Evolution, Expression and Function

Gardette R. Valmonte[1,2], Kieren Arthur[2], Colleen M. Higgins[1] and Robin M. MacDiarmid[2,3,*]

[1]Institute for Applied Ecology New Zealand, School of Applied Sciences, Auckland University of Technology, New Zealand
[2]The New Zealand Institute for Plant & Food Research Limited, New Zealand
[3]School of Biological Sciences, The University of Auckland, New Zealand
*Corresponding author: Fax: 6499257001, E-mail: robin.macdiarmid@plantandfood.co.nz
(Received June 18, 2013; Accepted December 17, 2013)

Calcium-dependent protein kinases (CPKs) are plant proteins that directly bind calcium ions before phosphorylating substrates involved in metabolism, osmosis, hormone response and stress signaling pathways. CPKs are a large multigene family of proteins that are present in all plants studied to date, as well as in protists, oomycetes and green algae, but are not found in animals and fungi. Despite the increasing evidence of the importance of CPKs in developmental and stress responses from various plants, a comprehensive genome-wide analysis of CPKs from algae to higher plants has not been undertaken. This paper describes the evolution of CPKs from green algae to plants using a broadly sampled phylogenetic analysis and demonstrates the functional diversification of CPKs based on expression and functional studies in different plant species. Our findings reveal that CPK sequence diversification into four major groups occurred in parallel with the terrestrial transition of plants. Despite significant expansion of the CPK gene family during evolution from green algae to higher plants, there is a high level of sequence conservation among CPKs in all plant species. This sequence conservation results in very little correlation between CPK evolutionary groupings and functional diversity, making the search for CPK functional orthologs a challenge.

Keywords: Calcium-dependent protein kinase • Calcium signaling • Gene family evolution • Plant development • Plant stress.

Abbreviations: AJ, autoinhibitory junction; BEAST, Bayesian Evolutionary Analysis Sampling Trees; CaM, calmodulin; CaMK, calmodulin-dependent protein kinase; CaML, calmodulin-like protein; CBD, calcium-binding domain; CBL, calcineurin B-like protein; CPK, calcium-dependent protein kinase; CT, C-terminal variable domain; EST, expressed sequence tag; GFP, green fluorescent protein; MAMP, microbe-associated molecular pattern; MAPK, mitogen-activated protein kinase; MCMC, Markov chain Monte Carlo; MRCA, most recent common ancestor; MSU release, Michigan State University release; MYA, million years ago; NJ, Neighbor–Joining; N-VD, N-terminus variable domain; PGSC, Potato Genome Sequencing Consortium; PHB, prohibitin; PK, protein kinase; RT–PCR, reverse transcription–PCR; TAIR, The Arabidopsis Information Resource; TIGR, The Institute for Genomic Research; WGD, whole-genome duplication.

Introduction

Calcium (Ca^{2+}) signaling is a highly integrated signaling network that plays a fundamental role in growth, development and stress responses in plants. Cytosolic Ca^{2+} concentrations change in complex spatio-temporal patterns in response to various stimuli. In plants, these altered Ca^{2+} signatures lead to specific cellular responses including stomatal movement, increased water retention, microbial detection and tip structure movement (DeFalco et al. 2010, Hashimoto and Kudla 2011). Plant Ca^{2+} signatures are decoded by a vast array of Ca^{2+} sensors, such as calmodulins (CaMs), calmodulin-like proteins (CaMLs), calcineurin B-like proteins (CBLs) and Ca^{2+}-dependent protein kinases (CPKs). These proteins undergo conformational changes upon binding Ca^{2+}, activating their respective Ca^{2+} responders, CaM-dependent protein kinases (CaMKs), Ca^{2+} and CaM-dependent protein kinases (CCa MKs) and CBL-interacting protein kinases (CIPKs) which phosphorylate specific downstream proteins (Harmon et al. 2001, DeFalco et al. 2010). Among Ca^{2+} sensors, CPKs are unique because they also function as Ca^{2+} responders.

CPKs have four major domains: an N-terminus variable domain (N-VD), a protein kinase (PK) domain, an autoinhibitory junction (AJ) and a calcium-binding domain (CBD) (**Fig. 1**) (Harmon et al. 2001). Binding of Ca^{2+} ions occurs in the CBD, which contains 1–5 (typically four) loops called EF-hands (Cheng et al. 2002). Each EF-hand loop is 12 amino acids long and is flanked by two α-helices, forming a helix–loop–helix conformation (Cheng et al. 2002, Grabarek 2006). The PK domain contains phosphorylation sites characteristic of serine/threonine kinases (Hardie 1999), which become active only upon conformational changes in the AJ and CBD as a result of Ca^{2+} binding (Wernimont et al. 2010). Some authors also

Plant Cell Physiol. 55(3): 551–569 (2014) doi:10.1093/pcp/pct200, available online at www.pcp.oxfordjournals.org

Fig. 1 Characteristic structure of CPKs. N-VD, N-terminal variable domain; PK-D, catalytic protein kinase domain; AJ, autoinhibitory junction domain; CBD, calcium binding domain usually containing four EF-hands (black boxes); and CT, C-terminal variable domain. The dashed line indicates the conserved region used in performing sequence alignments.

consider the C-terminal variable domain (CT) as a distinct domain; the CT is as variable as the N-VD but is typically shorter (Klimecka and Muszyńska 2007).

CPKs are encoded by large gene families present in protists (Billker et al. 2009), oomycetes (Broad Institute of Harvard and MIT 2010), green algae (McCurdy and Harmon 1992, Baillie et al. 2000) and plants (Harmon et al. 2001, Li et al. 2008b, Asano et al. 2010, Kiselev et al. 2010), but have not been identified in any animal or fungal genome (Zhang and Choi 2001, Hrabak et al. 2003). Most flowering plants have between 20 and 40 CPK genes; there are 34 CPK genes in the model eudicot, *Arabidopsis thaliana* (thale cress or Arabidopsis), 31 genes in the model monocot, *Oryza sativa* (rice) (Cheng et al. 2002, Hrabak et al. 2003, Asano et al. 2005), and 30 genes in the model tree species, *Populus trichocarpa* (poplar) (Zuo et al. 2013).

Despite increasing evidence supporting the involvement of different CPKs in plant stress and development responses, a recent comprehensive genome-wide analysis of CPKs to demonstrate their evolution in plants has not been undertaken. A phylogenetic analysis of CPKs from protists through to plants was reported over a decade ago, but this was limited to the CPK sequences available in 2001 (Zhang and Choi 2001). Comparative genome-wide phylogenetic analyses of CPKs and their closely related gene families have so far been described only in apicomplexan protists (Nagamune and Sibley 2006, Billker et al. 2009) and a small number of plants, namely Arabidopsis (Cheng et al. 2002, Hrabak et al. 2003), rice (Asano et al. 2011) and *Triticum aestivum* (wheat) (Li et al. 2008a). The 34 CPKs of Arabidopsis separated into four major evolutionary groups (I–IV) (Cheng et al. 2002). Upon the inclusion of rice and wheat CPKs, Group II and III were separated into subgroups (IIa, IIb, IIIa and IIIb) (Asano et al. 2005, Li et al. 2008b). Additionally, phylogenetic analyses that consisted of CPK sequences from various plants (some analyses included a few protist and algae CPKs) were also undertaken to describe evolution and function among CPKs; but these were also limited by the number of CPK genes included to represent the genome of each species (Harmon et al. 2000, Hrabak et al. 2003, Boudsocq and Sheen 2013).

CPKs studied to date have different tissue and cellular localizations, substrate specificities, Ca^{2+} sensitivity and expression patterns in response to development and stress, but it is unclear whether functional distinctions and overlaps between related CPKs mirror the evolution of CPK genes. With the recent completion of several plant genomes, this paper explores the evolution of CPKs from green algae to higher plants using a broadly sampled phylogenetic analysis and examines its correlation with the functional diversification of CPKs based on expression and functional studies reported in different plant species.

Results

Genome-wide identification of CPKs in algae and plants

CPK sequences were mined from the genomes of 15 selected species representing major taxonomic groups from green algae to higher plants. This included two green algae, *Volvox carteri* and *Chlamydomonas reinhardtii* (Fukuzawa et al. 2008); a bryophyte, *Physcomitrella patens* (Rensing et al. 2008); a pteridophyte, *Selaginella moellendorfii* (Banks et al. 2011); a gymnosperm, *Picea sitchensis* (Ralph et al. 2008); four monocots, *O. sativa* (Asano et al. 2005), *T. aestivum* (Li et al. 2008b), *Sorghum bicolour* (sorghum) (Paterson et al. 2009) and *Zea mays* (maize) (Schnable et al. 2009); and six eudicots, *A. thaliana* (Cheng et al. 2002), *Vitis vinifera* (grape) (Jaillon et al. 2007), *Glycine max* (soybean) (Schmutz et al. 2010), *P. trichocarpa* (poplar) (Tuskan et al. 2006, Zuo et al. 2013), *Carica papaya* (papaya) (Ming et al. 2012) and *Solanum tuberosum* (potato) (Potato Genome Sequencing Consortium 2011) (**Fig. 2**; **Supplementary Table S1**). At the time of analysis, nearly complete genomes or high quality draft assemblies were available for the genomes of *C. reinhardtii*, *O. sativa*, *S. bicolour*, *Z. mays*, *V. vinifera*, *G. max*, *P. trichocarpa*, *A. thaliana* and *S. tuberosum* (Engstrom 2011, Rouard et al. 2011, Goodstein et al. 2012, Zhang et al. 2012, Du et al. 2013). *Volvox carteri*, *P. patens*, *S. moellendorfii*, *P. sitchensis* and *C. papaya* only had scaffold assemblies or tentative consensus data available. Sequences from Arabidopsis were used to identify CPKs from these selected species. The query sequences used in the BLAST searches against each of the genomes included one representative sequence for each of the four evolutionary groups (AtCPK1, 8, 21 and 16) and one consensus sequence derived from all 34 Arabidopsis CPKs (AtCPKs). True CPK sequences were distinguished from CPK-related sequences and other Ca^{2+} sensors and/or responders using InterProScan (see the Materials and Methods for the criteria used). This search strategy was tested against Arabidopsis, rice, poplar and maize, and detected all previously reported CPKs from these genomes. At the time of writing, there was no full genome available for *T. aestivum*, but an extensive evolutionary and functional study of the CPK gene family performed by Li et al (2008b) was used as reference.

A total of 352 CPK sequences were identified, which varied in length. Full-length CPK proteins ranged in size from 393 to 764 amino acids; except for two putative CPKs from *C. reinhardtii* which were 1,042 and 1,801 amino acids long (both had long CT domains). Variation in length of the entire CPK gene is usually due to differences in the length of the N-VD and CT

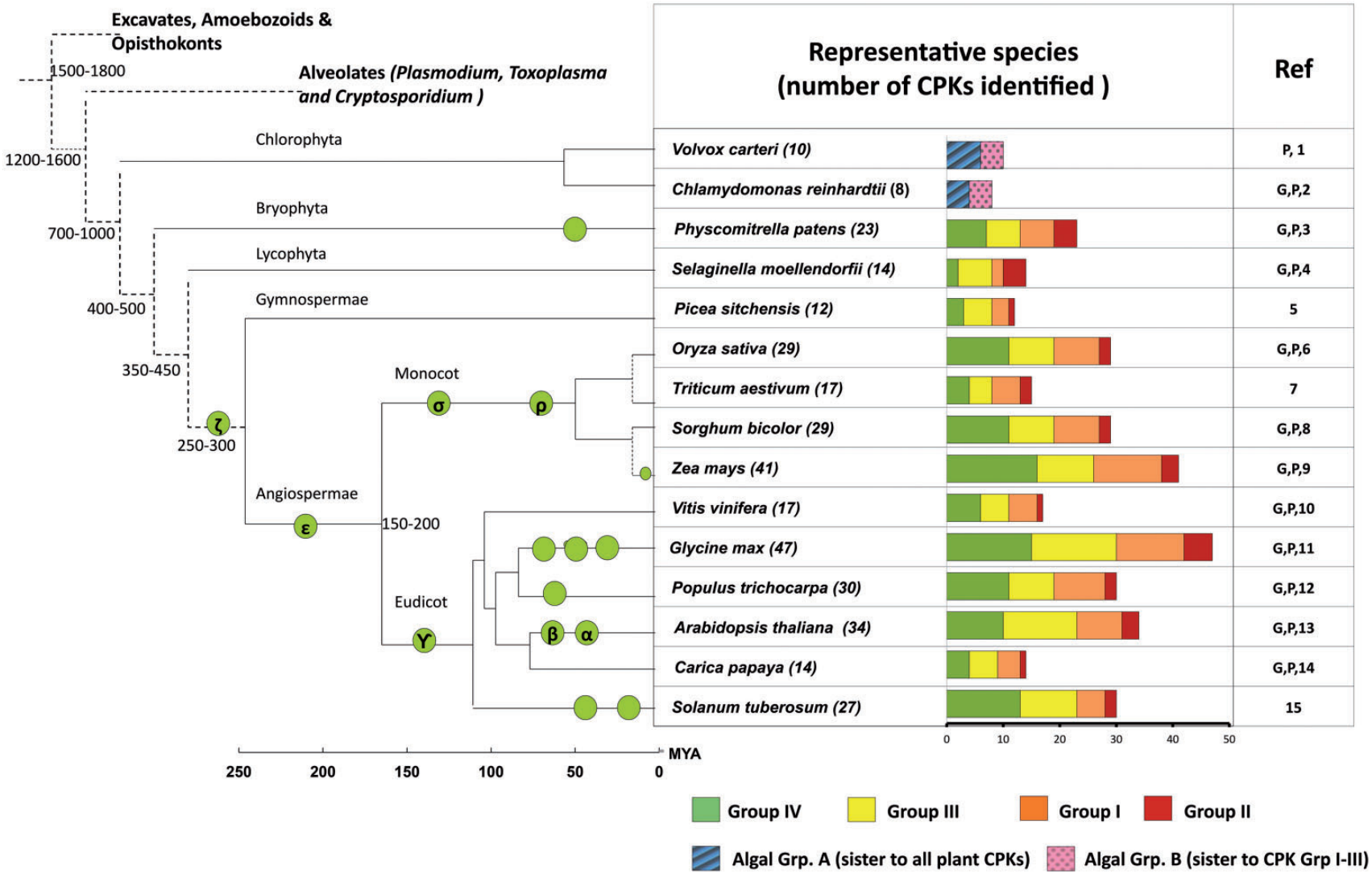


Fig. 2 Distribution of CPK evolutionary groups among the genome representatives of Plantae. Circles indicate whole-genome duplication (WGD) events that occurred within lineages, with some events designated with Greek symbols by previous literature. Branches in broken lines denote that the diversification ages are not drawn to scale. References for sequence mining: (G), Greenphyl Database (http://greenphyl.cirad.fr/v2/cgi-bin/index.cgi); (P), Phytozome Database (www.phytozome.net). Genome version: (1) *V. carteri* v1.0; (2) *C. reinhardtii* v4 and v4 u10.2; (3) *P. patens* v1.1 and COSMOSS annot v1.6; (4) *S. moellendorfii* v1.0; (5) *P. sitchensis* EST and TCs from DFCI; (6) *O. sativa* MSU release 6.0 and Asano et al. (2005); (7) *T. aestivum* EST and tentative consensus sequences from TIGR r2 and Li et al. (2008); (8) *S. bicolor* v1.4; (9) *Z. mays* v4 and v5a.59; (10) *V. vinifera* 8× and 12× genome assembly; (11) Glyma1.0; (12) *P. trichocarpa* v1.1 and v2.2; (13) TAIR r10 and Cheng et al. (2002); (14) Hawaii Papaya Genome Project v2007; (15) *S. tuberosum* PGSC, http://potatogenome.net.

domain, and occasionally due to the number of EF-hands in the CBD. Almost all the CPK sequences had four EF-hands. However, in some species, a small number of CPKs were found to have as few as one (e.g. AtCPK25) or up to five EF-hands (e.g. AtCPK22). The differences in length among CPK sequences may indicate the presence or absence of motifs that could affect localization and functional specificity.

The total numbers of CPK genes within the genomes examined were consistent with the pattern of genome duplication and polyploidization events that have occurred through plant evolution (summarized in **Fig. 2**). The divergence between red algae and Viridiplantae (green algae and land plants) occurred about 1,200–1,600 million years ago (MYA), while the split between green algae and land plants happened 700–1,000 MYA as estimated in several studies (Hedges et al. 2004, Yoon et al. 2004, Zimmer et al. 2007, Parfrey et al. 2010). The separation between non-vascular (Bryophytes) and vascular (Tracheophytes) plants took place around 400–900 MYA (Hedges 2002, Hedges et al. 2004, Taylor et al. 2005, Zimmer et al. 2007). From this point, whole-genome duplication (WGD) events have occurred in the Spermatophyte (seed plants) lineage: (i) the ancestral seed plant (ζ) and (ii) ancestral angiosperm (ε) WGD events (Jiao et al. 2011); (iii) the ancestral eudicot triplication event (γ) (Bowers et al. 2003, Jaillon et al. 2007, Tang et al. 2008, Jiao et al. 2011); and (iv) the ancestral monocot (σ) WGDs. Consistent with these events, green algae had the least number of CPKs (eight in *C. reinhardtii* and 10 in *V. carteri*), whereas angiosperms generally had more CPKs than other land plants, except for *V. vinifera*, *C. papaya* and *T. aestivum* (**Fig. 2**). It must be noted, however, that *C. papaya* and *T. aestivum* genomes have not yet been completely annotated. *Vitis vinifera*, on the other hand, has the least number of CPK genes among fully or nearly completely sequenced eudicot genomes (17 CPK genes), most probably because it has not undergone any WGD since the γ event (Jaillon et al. 2007).

There was considerable variation in the total number of CPK genes between plant families and species, due to family- or species-specific WGD events (represented by green circles in **Fig. 2**). For example, in monocots, *Z. mays* had the most CPK genes (41), probably owing to the grass lineage (ρ) WGDs (Paterson et al. 2010) and its recent genome doubling (Woodhouse et al. 2010). *Glycine max* had the most CPKs among eudicots, possibly due to several rounds of polyploidization within this species (Gill et al. 2009, Schmutz et al. 2010). Independent genome duplications in Fabaceae, Solanaceae, Brassicaceae (α and β events) and *Populus* (Tuskan et al. 2006, Gill et al. 2009, Magallon and Castillo 2009, Soltis et al. 2009, Jiao et al. 2011) also corresponded to the increase in CPK genes present in the representative species (47, 27, 34 and 30 CPK genes, respectively).

CPK evolution from algae to angiosperms

Phylogenetic analyses were undertaken using the amino acid sequences of only the conserved region, consisting of the PK domain, the AJ and the CBD (**Fig. 1**, black broken line). The N-VD and CT were excluded from multiple sequence alignments (**Supplementary File S1**) due to the extreme variability within these domains, causing disproportionate branches, inconsistent groupings and low bootstrap values. Using protist CPKs as the outgroup for all Viridiplantae CPKs, the general topology of the resulting Neighbor–Joining (NJ) and Maximum Likelihood (ML) trees appeared similar to that of Li et al. (2008b) and Boudsocq and Sheen (2013). A condensed view of the ML tree generated from the alignment of the conserved region is shown in **Fig. 3**, while the detailed topology for each evolutionary group is shown in **Supplementary Figs. S1–S4**. The NJ tree constructed is shown in **Supplementary Fig. S5**. To determine whether the N-VD and CT of the CPKs affect the evolutionary groupings, full-length CPK sequences were aligned using Multiple Sequence Comparison by Log-Expectation (MUSCLE) (Edgar 2004) (**Supplementary File S2**). Due to computational limitations, only an NJ tree was constructed; the general topology of the resulting tree is similar to that of the NJ and ML trees constructed from the conserved region of CPKs. Furthermore, separate ML trees of the full-length sequences for each of the four evolutionary groups were constructed and these also retain the same general topology described above. The NJ tree of all the full-length CPK sequences and the ML trees for each evolutionary groups are shown in **Supplementary Figs. S6–S10**.

As illustrated in **Fig. 3**, CPKs grouped into four major evolutionary groups (I–IV), with Groups II and III further divided into two subgroups (a and b) (Asano et al. 2005, Li et al. 2008b). As calculated from a distance matrix (Patristic Distances), the average branch lengths from the last common ancestor of all CPKs are as follows: Group I, 0.552; Group IIa, 0.585; Group IIb, 0.547; Group IIIa, 0.714; Group IIIb, 0.614; Group IV, 0.858 amino acid substitutions per site. Based on the branch pattern of the evolutionary tree, the Group IV lineage appeared to have diverged first from the last common ancestor (**Fig. 3**, branch *e4*, embryophyte group IV). Group III formed a clade separate from Groups I and II (bootstrap value of 100%), while the split between Groups I and II appeared to be the most recent (bootstrap value of 74%).

Despite being the earliest lineage from the last common ancestor of land plant CPKs, Group IV CPKs are the most divergent from this common ancestor. Group IV CPKs have the longest main branch (*e4*) and the highest average branch length (0.858), and the multiple sequence alignment shows many differences between Group IV CPKs and Group I–III CPKs, particularly within the CBD region (**Supplementary Files S1**, **S2**). On the other hand, Group II CPKs had the shortest main branch (*e2*) and, within this, Group IIb CPKs have the lowest average branch length (0.547). Furthermore, Group IIb CPKs also includes members from all lower plant genomes used in the analysis. This suggests that Group IIb CPKs are the most conserved from the last common ancestor of all CPKs. This may be important in determining target CPK genes for studying plant biotic and abiotic stress tolerance that is applicable to a broad range of plant species.

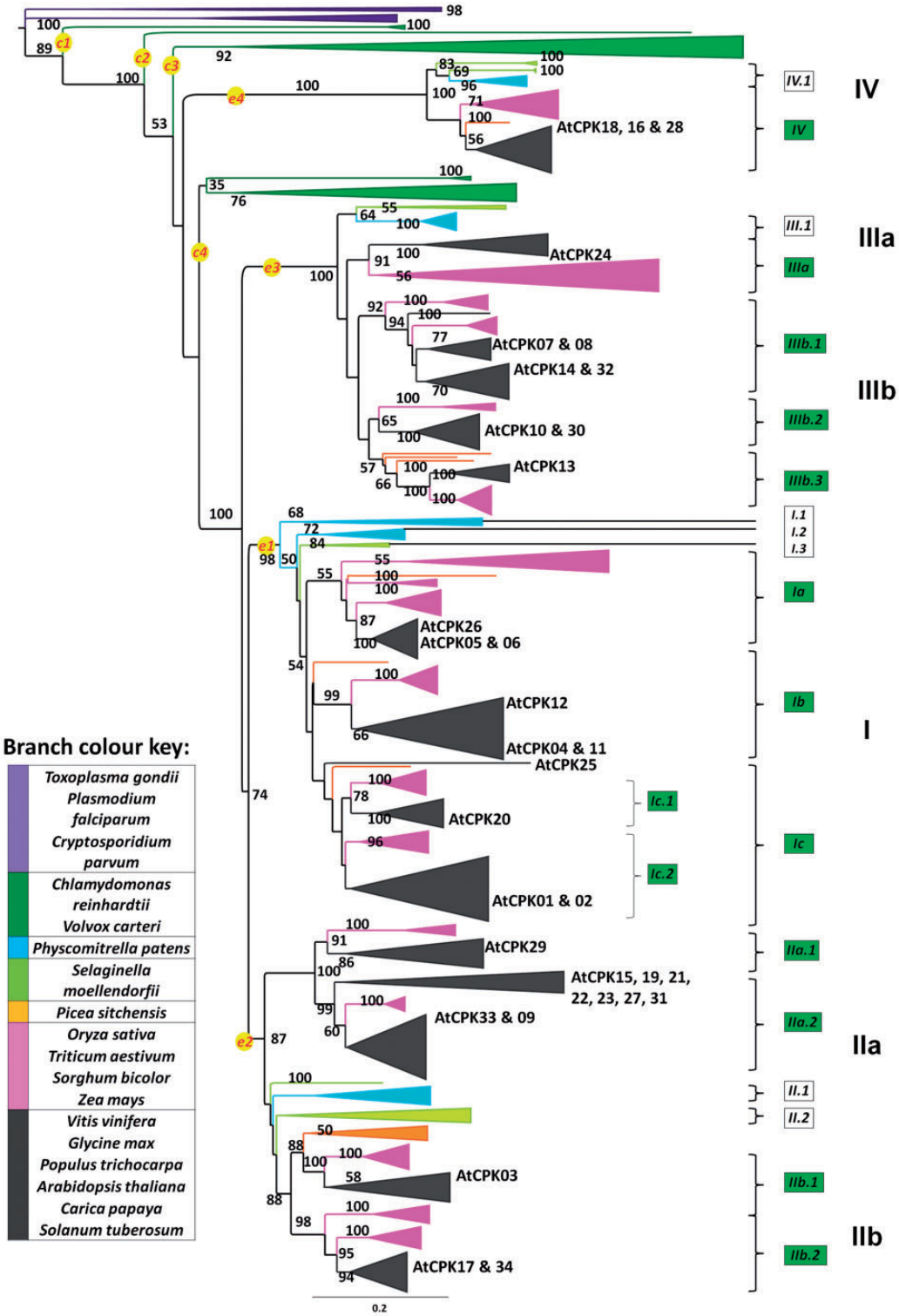


Fig. 3 Phylogenetic tree of CPKs from algae to higher plants. The tree was constructed by the Maximum Likelihood method, with 1,000 bootstrap replicates using GeneiousPro 5.6 software (Geneious version 5.6). A total of 352 plant (including green algae) and five apicomplexan (as outgroup) CPK protein sequences were included in this analysis. Branch colors match the species colors in the left box. Branching points as described in the text are indicated in red font encircled in yellow (c, chlorophytes/green algae; e, embryophytes/land plants). On the far right, evolutionary groupings are indicated in Roman numerals. Non-seed plant clusters are in white boxes while monocot–dicot clusters are in green boxes. The positions of Arabidopsis CPKs in the tree are also indicated. The Neighbour Joining tree shows similar general topology (**Supplementary File S2**).

All plant CPKs were well distributed among the four CPK evolutionary groups, with the exception of the algal CPKs. Green algae CPKs had three lineages separate from all plant CPKs (**Fig. 3**, branches *c1–c3*, chlorophyte group I–III) and a fourth lineage (*c4*) that clustered with Groups I–III but within distinct clades. Similar to angiosperms, non-flowering plants such as mosses, lycophytes and conifers had CPKs distributed throughout the four major groups. However, within each evolutionary grouping, CPK genes from bryophyte moss (*P. patens*) and a lycophyte (*S. moellendorfii*) formed early lineages distinct from gymnosperm and angiosperm lineages (*I.1-3*, *II.1-2*, *III.1* and *IV.1*). The separation of CPKs into evolutionary Groups I–IV is characteristic of land plants but not of green algae, while the separation into subgroups is only seen among seed plants. Non-seed land plant (*S. moellendorfii* and *P. patens*) CPKs form monophyletic groups of their own (e.g. Groups III.1 and IV.1) or form separate branches (Group I and II have several lineages of non-seed land plants, I.1, I.2, I.3, II.1 and II.2), which in all cases are basal to the angiosperm subgroups (**Fig. 3**).

Monocot and eudicot CPKs form several clusters within the evolutionary groups. Group IIIa and IV both had only one monocot–eudicot cluster (**Fig. 3**, clusters *IV* and *IIIa*), while Groups I, IIa, IIb and IIIb had 2–3 monocot–eudicot clusters (clusters *Ia*, *Ib* and *Ic.1-2*; *IIa.1-2*, *IIb.1-2*; *IIIb.1-3*). CPKs that belong to these monocot–eudicot clusters were highly similar in amino acid sequence (80–97% identities). Furthermore, no species- or family-specific clades were present within each monocot or eudicot cluster. This suggests that diversification of CPK genes occurred among ancestral angiosperms and are now shared by extant monocot and eudicot species. This corresponds to the increase in CPK gene numbers among angiosperms.

To estimate the timing of CPK diversification, a tree with a relaxed molecular clock was constructed using Bayesian Markov Chain Monte Carlo (MCMC) analysis in the program Bayesian Evolutionary Analysis Sampling Trees (BEAST) (Drummond and Rambaut 2007) (**Fig. 4**). Similar to the NJ and ML trees (**Fig. 3**; **Supplementary Figs. S1–S4**, **S6–S10**), this topology shows that CPKs from land plants were split into four evolutionary groups, while green algae CPKs formed a separate group with an earlier lineage. The Bayesian tree estimated that the diversification of CPKs into four major evolutionary groups occurred 268–340 MYA. There were very few differences in the overall topology between the Bayesian and the ML tree, particularly the separation between Groups I, II and III: in the ML tree (**Fig. 3**), Group III was a sister group to the common ancestor of Groups I and II, while in the Bayesian tree, Group II showed an earlier lineage from I and III. The common ancestor of all Group II CPKs appeared to have split into two lineages 340 MYA, which is the earliest amongst all the other CPK evolutionary groups.

Gene structure analysis of land plant CPKs: bryophyte, monocot and eudicot representatives

The gene structure analysis of *P. patens*, *O. sativa* and *A. thaliana* showed that the intron–exon patterns were similar between CPKs belonging to the same evolutionary group and taxon, but different between taxa (**Supplementary Fig. S10**). Within a clade, all AtCPKs had similar intron–exon patterns; but these were different from the intron–exon patterns of rice and moss CPKs, and vice versa. In most of the CPK genes examined, the first exon was long, followed by a series of shorter exons. In Arabidopsis, the evolutionary groups showed group-specific intron–exon patterns. For example, Group IV CPKs had a long initial exon followed by 10–12 very short exons, while Group I CPKs had a very long initial exon followed by 5–6 shorter exons. Group-specific patterns were also apparent in rice; however, the rice CPK intron–exon patterns were different from those of Arabidopsis. Notably, duplicated CPK gene pairs such as AtCPK4 and 11, AtCPK17 and 34, OsCPK25 and 26, OsCPK3 and 16, and OsCPK2 and 14 had highly similar intron–exon patterns, which may also impact on the functional similarities and/or redundancy between these genes. Only a few *P. patens* CPKs, mostly belonging to Group IV CPKs, had intron–exon information. All Group IV CPKs had more (between 10 and 12), but shorter exons, than Groups I–III which had 5–7 members. This supports the phylogenetic trees constructed using CPK protein sequences (**Fig. 3**), which shows that Group IV CPKs form a separate clade of earlier lineage.

Functional diversification of plant CPKs

The functional importance of CPK gene expansion and diversification events durng evolution is unclear. Why are there such a large number of CPK genes within a single plant species? Is the expansion of CPKs among plants highly related to their functional diversification? To address these questions, an extensive literature review and expression profile examination was undertaken to determine the functional diversification among plant CPKs. Most information was based on developmental and stress response studies of mRNA transcript accumulation of Arabidopsis, rice and wheat CPKs, with the addition of a few functional characterizations undertaken on individual CPKs from different species such as *V. vinifera*, (Yu et al. 2006), *Z. mays* (Estruch et al. 1994, Takezawa et al. 1996, Murillo et al. 2001, Szczegielniak et al. 2005), *S. tuberosum* (Raices et al. 2001, Kobayashi et al. 2007, Gargantini et al. 2009, Giammaria et al. 2011), *Solanum lycopersicum* (Rutschmann et al. 2002, Chang et al. 2009, Chang et al. 2011) and *Nicotiana tabacum* (Tai et al. 2009). The sequences of CPKs with reported function were obtained from publicly available sequence databases including GenBank (Benson et al. 2010) and Phytozome (Goodstein et al. 2012). An overview of the reported response and biological functions of CPKs is shown in **Fig. 5a–c** (detailed information is shown in **Supplementary Tables S2–S4** and **Supplementary Fig. S11**). To illustrate any correlation between sequence relationships of the CPKs for which published functional information is available and also to identify functional similarity between closely related sequences, the functional information was aligned to an unrooted NJ tree as shown in **Fig. 6**. The ML tree is shown in **Supplementary Fig. S12**, which has similar topology. It should be noted that CPKs in this tree are not necessarily

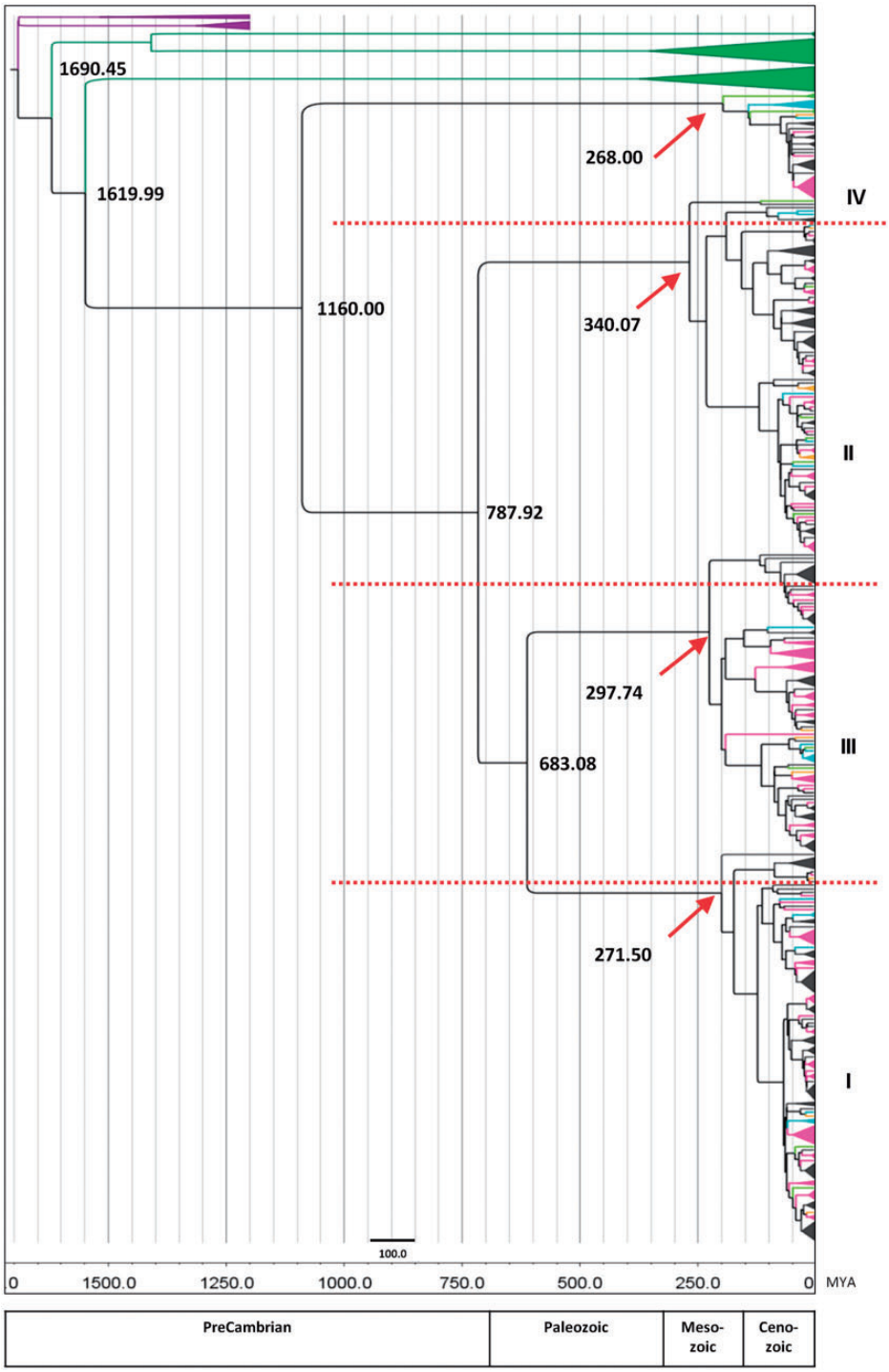


Fig. 4 Bayesian evolutionary tree of CPKs with a relaxed molecular clock. The tree was constructed using Bayesian Evolutionary Analysis Sampling Trees (BEAST) software (Drummond and Rambaut 2007), with calibration points at 1,700 MYA (Hedges 2002) for apicomplexan and 1,400 MYA for green algae most recent common ancestor (MRCA) (Yoon et al. 2004). The tree prior used was Yule model. Evolutionary groupings are indicated in Roman numerals. The main stem age for each of the four evolutionary CPK groups is marked by the red arrow. Branch colors match the species colors in **Fig. 3**.

included in the broadly sampled phylogenetic tree (**Fig. 3**; **Supplementary Figs. S1–S4**) as some of the CPKs with functional information come from genomes with incomplete or no data available.

CPKs appear to respond to different developmental and stress stimuli, but it must be noted that not all CPKs have been tested against all types of responses (**Fig. 5a–c**; **Supplementary Table S4**). Most abiotic stress studies have

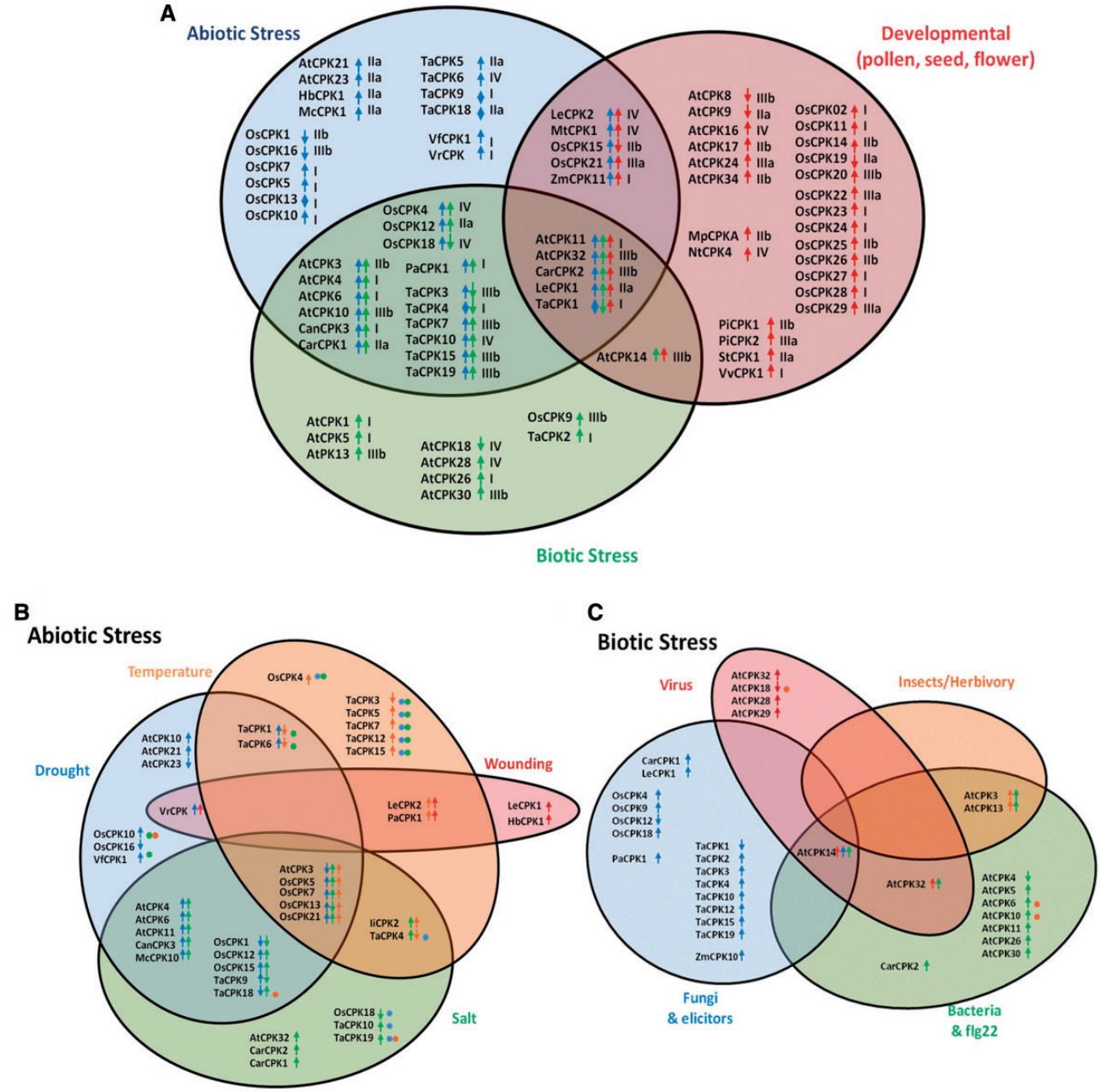
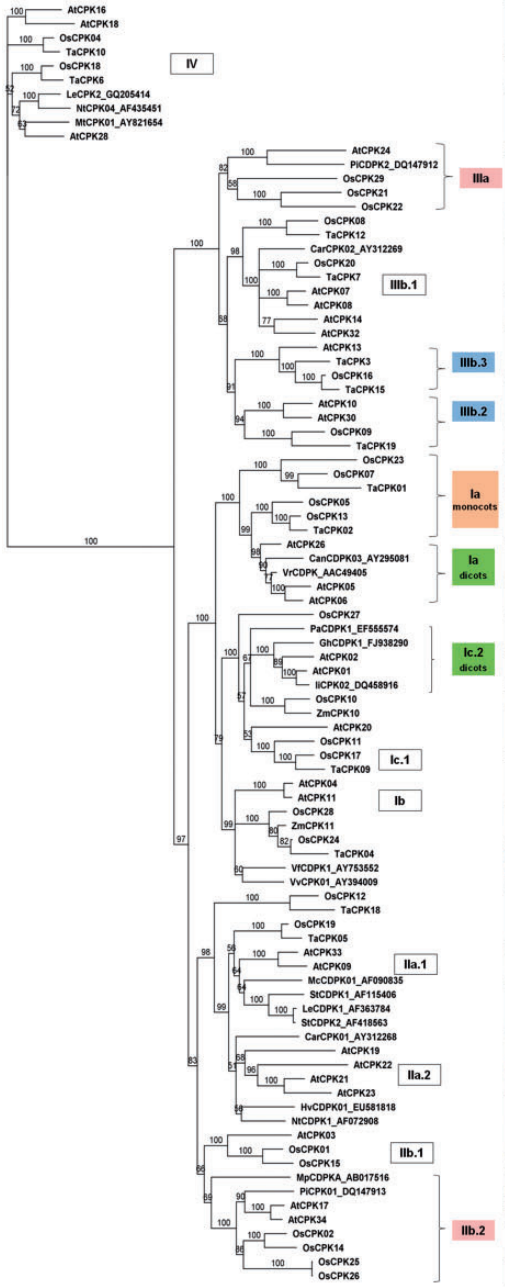


Fig. 5 Overview of plant CPK functional information based on the literature. (a) Developmental, biotic and abiotic stress responses. (b) Abiotic stress response. (c) Biotic stress responses. Information includes transcript and protein accumulation, enzyme activity, gene knockout and overexpression experiments. The upward-pointing arrows indicate up-regulation, while downward-pointing arrows indicate down-regulation in response to a certain type of stress. Diamonds indicate up- and down-regulation under different types of abiotic stress. Dots indicate no change in CPK accumulation in response to a specific stress. The colors of arrows and dots correspond to the font color of stress. See **Supplementary Tables S3** and **S4** for detailed information and citations.

focused on drought and salt conditions, with very few studies having focused on extreme temperatures and wounding (**Fig. 5b**). The majority of biotic stress data have come from microarray experiments in Arabidopsis and semi-quantitative reverse transcription–PCR (RT–PCR) carried out on wheat RNA (Li et al. 2008a). There is very little information with regards to virus infection and responses to herbivory (**Fig. 5c**).

It was expected that the NJ and ML trees would reveal functional similarities between related sequences. However, based on current available data, there was no correlation between functional response and phylogenetic grouping among the six major CPK evolutionary groups, except for a few small clusters within the subgroups (**Fig. 6**). Each group (major or subgroup) had CPK genes involved in developmental and stress



Plant	IIIa and IIb.2 – Mostly Developmental	Expression/ Function
Arabidopsis	AtCPK24 (IIIa) AtCPK 17 & 34 (IIb.2)	Abundant in flower & pollen Pollen tube growth
Petunia	PiCPK2 (IIIa) PiCPK1 (IIb.2)	Pollen tube extension (PiCPK2) Pollen tube growth (PiCPK1)
Rice	OsCPK21, 22 & 29 (IIIa) OsCPK2, 14, 25 & 26 (IIb.2)	Abundant in panicle, developing seed & stamen Cold & dessication (OsCPK21) Seed development (OsCPK2)
Maize	GRMZM2g097533, GRMZM2g332660, GRMZM080871 and GRMZM2g158721 (IIIa) GRMZM2g472311, GRMZM2g365815, GRMZM2g340224, and GRMZM2g167276 (IIb.2)	Unknown
Wheat	TaCPK13	Unknown
Sorghum	Sb08g007660, Sb07g025560, Sb03g043700 (IIIa) Sb05g001380, Sb05g002110, Sb09024100 and Sb03g037570 (IIb.2)	Unknown
Potato	StCPK25 (IIIa) StCPK17, 18 & 19 (IIb.2)	Unknown
Soybean	Glyma12g05730 and Glyma11g13740 (IIIa) Glyma2g44720, Glyma20g08140, Glyma07g36000 (IIb.2)	Unknown
Poplar	POPTR_0007s02120 (IIIa) POPTR_0009s07330 and POPTR_0001s28150 (IIb.2)	Unknown
Grape	GSVIVP00001926001 (IIIa) GSVIVP00022286001 (IIb.2)	Unknown
Papaya	evm.TU.contig_33100 (IIIa) evm.TU.supercontig_157.56 (IIb.2)	Unknown
Plant	**IIIb2 and IIIb.3 – Fungal response**	**Expression/ Function**
Arabidopsis	AtCPK10 & 30 (IIIb.2) AtCPK13 (IIIb.3)	Upregulated by *Erysiphe sp.* infection (AtCPK10)
Rice	OsCPK9 (IIIb.2) OsCPK3 and 16 (IIIb.3)	Upregulated by *Magnaporthe sp.* infection (OsCPK9)
Wheat	TaCPK19 (IIIb.2) TaCPK3 and 15 (IIIb.3)	Upregulated by *Blumeria graminis* infection
Maize	GRMZM2G030673, GRMZM2G088361 and GRMZM2G311220 (IIIb.3)	Unknown
Sorghum	Sb01g011630 (IIIb.2) Sb03g038870 and Sb09g022960 (IIIb.3)	Unknown
Potato	StCPK26 (IIIb.2)	Unknown
Soybean	Glyma17g10410, Glyma05g01470, Glyma06g20170 and Glyma04g34440 (IIIb.2) Glyma18g43160 and Glyma07g18310 (IIIb.3)	Unknown
Poplar	POPTR_0015s07740 and POPTR_0012s07360 (IIIb.2) POPTR_0016s12460 and POPAL_831925 (IIIb.3)	Unknown
Grape	GSVIVP00000071001 (IIIb.2) GSVIVP00038883001 (IIIb.3)	Unknown
Papaya	evm.TU.supercontig_17.194 (IIIb.2)	Unknown
Plant	**Ia (monocots)- Cold stress**	**Expression/ Function**
Rice	OsCPK7 & 13 OsCPK5, 6 & 23	Up-regulated in cold stress (7&13) Cold tolerance (OsCPK13) Unknown (OsCPK5, 6 & 23)
Wheat	TaCPK1 & 2	Up-regulated in cold stress (TaCPK1 only)
Maize	GRMZM2g040743, GRMZM2g032852, GRMZM2g347047, GRMZM2g081310, GRMZM2g321239, and GRMZM2g314396	Unknown
Sorghum	Sb6g026530, Sb04g031570, Sb04g038450, Sb01g048570	Unknown
Plant	**Ia & Ic.2 (dicots) – Bacterial response**	**Expression/ Function**
Arabidopsis	AtCPK5, 5 & 26 (Ia) AtCPK1 & 2 (Ic.2)	Flg22 signalling (all) Resistance to *Fusarium oxysporum* and *P. syringae* (AtCPK1)
Capsicum anuum	CanCDPK3 (Ia)	Upregulated by *Xanthomonas axonopodis* infection
Phalaenopsis amabilis	PaCPK1 (Ic.2)	Upregulated by *Erwinia chrysanthemi*
Potato	StCPK4 & 5 (Ia) StCPK 6 & 7 (Ic.2)	Unknown
Soybean	Glyma20g00320 and Glyma02g48160 (Ia) Glyma20g017020 and Glyma10g23620 (Ic.2)	Unknown
Poplar	POPTR_0004s21710 and Popal_803966 (Ia) POPTR_0010s25090 and POPTR_0008s01530 (Ic.2)	Unknown
Grape	GSVIVP00036285001 (Ia) GSVIVP00002665001 (Ic.2)	Unknown
Papaya	evm.TU.supercontig__60.3 (Ia) evm.TU.supercontig__33.122 (Ic.2)	Unknown

Fig. 6 Plant CPKs with reported function and identified orthologous groups. The Neighbour Joining tree was constructed using CPKs for which function has been reported, with 100 bootstrap replicates using GeneiousPro 5.6 software (Geneious version 5.6). Maximum Likelihood tree shows similar topology, shown in **Supplementary Fig. S12**. The table highlights potential orthologs within clades that seem to have a functional pattern. Clades are highlighted based on the type of response. See **Supplementary Tables S2–S4** for detailed information and citations.

responses (**Fig. 5**). Likewise, no particular organ- or cell type-specific clades were observed, except that each evolutionary group has small clades that are preferentially expressed in floral tissues (particularly the stamen), which indicates developmental function(s) (**Fig. 5**; **Supplementary Fig. S11**). In many cases, closely related CPKs within an orthologous group from the same species did not show similarity in function despite high amino acid sequence similarity. For example,

within Group Ib, AtCPK4 and 11 responded to various biotic and abiotic stresses, while AtCPK12 did not seem to respond to any stress, even though it had the highest sequence similarity to them among all AtCPKs. However, the intron–exon structures of AtCPK12 have a few differences compared with those of AtCPK4 and 11 (**Supplementary Fig. S13**). Another example within Group IV, AtCPK18, showed a decrease in transcript accumulation upon viral infection (Babu et al. 2008), while its closest homolog, AtCPK16, did not. Rather, AtCPK16 appeared to function primarily in pollen development as it is exclusively abundant in pollen (Swarbreck et al. 2008). These two genes also have different intron–exon patterns (**Supplementary Fig. S13**). An interesting relationship was also observed between likely paralogs OsCPK1 and 15 (Group IIb.1), wherein both CPKs responded to drought and salt, but one was up-regulated (OsCPK15) whereas the other was down-regulated (OsCPK1) (Ray et al. 2007). These two genes, despite being the most similar in terms of amino acid sequence, have distinct intron–exon patterns.

In contrast, there were small orthologous CPK groups that seemed to respond to similar types of signals (**Fig. 6**). A number of CPKs within Groups IIb.2 and IIIa appeared to be developmentally associated, while at least five small clades in Group I and IIIb showed potential orthologous functions in stress reponse: Group Ia (monocots only), IIIb.2, IIIb.3, and small eudicot clades in Ia and Ic.2 (**Fig. 6**). Most of the other CPK evolutionary subgroups (Ib, Ic.1, IIa, IIb.1, IIIb.1 and IV) responded to different developmental and biotic and/or abiotic stress signals with no apparent pattern. The orthologous CPK groups mentioned are described briefly in the following sections.

Group IIb.2 and IIIa are mostly involved in development. Several CPKs were known to have a role in development of reproductive structures. Proteins with such a role were identified in Groups IIb.2 and IIIa (**Fig. 6**). All CPKs in Group IIb.2 with demonstrated function appeared to be involved in development. The transcripts of AtCPK17 and 34 (Swarbreck et al. 2008), OsCPK2, 14, 25 and 26 (Ye et al. 2009), and PiCPK1 (Yoon et al. 2006) were almost undetectable in vegetative tissues but were highly abundant in pollen, flower or seed tissues (**Supplementary Fig. S11**, **Supplementary Table S2**). Arabidopsis *cpk17* and *cpk34* double mutants had 350-fold reduced pollen transmission efficiency and 3-fold reduction in pollen tube growth rate (Myers et al. 2009). Transient overexpression of AtCPK34 green fluorescent protein (GFP)-fused proteins in tobacco pollen resulted in depolarization of pollen tube growth (Zhou et al. 2009). Based on their position in the phylogenetic tree (**Supplementary Fig. S2**), OsCPK2 and 14 (85% pairwise amino acid identity with each other) and OsCPK25 and 26 (99.6% pairwise amino acid identity) were potential orthologs of AtCPK17 and 34. However, there was no published functional information available for these tandem rice CPKs, except that OsCPK2 appeared to be important in light-responsive signaling involved in seed development (Morello et al. 2000). Another Group IIb.2 CPK, PiCPK1, from petunia, appeared to have similar functions to AtCPK17 and 34. When overexpressed, PiCPK1 resulted in stunted pollen tubes with almost spherical tips and inhibited pollen germination and tube growth (Yoon et al. 2006).

In addition to Group IIb.2, Group IIIa CPKs appeared to have developmental roles. AtCPK24, the only Arabidopsis CPK within this subgroup, had high transcript accumulation in floral and pollen tissues, but had very low accumulation throughout the rest of the plant. AtCPK24 was also shown to be important in pollen development as it appeared to connect pathways between the vegetative nucleus and generative cell and as its overexpression resulted in reduced pollen tube elongation. Rice has three CPK genes within this subgroup (OsCPK21, 22 and 29), which were all predominantly abundant in panicle, stamen and seed development. No functional studies with regards to pollen development have been done on these rice CPKs; therefore, a similar experiment to that of AtCPK24 may be valuable to verify their function. However, OsCPK21 was also found to respond to cold and desiccation stress (2- and 3-fold increase in transcript accumulation, respectively), but not OsCPK22 and 29 (Ray et al. 2007). In petunia, transient overexpression of PiCPK2 showed inhibition of pollen tube extension but no effect in growth polarity or germination rates, resulting in short tubes with normal morphology (Yoon et al. 2006).

The remaining CPKs in Groups IIb.2 and IIIa (**Fig. 6**) have no specific information available but may also be important in development. These CPKs include seven genes in sorghum, eight genes in maize (four in IIIa and three in IIb.2), one gene in wheat (TaCPK13 in IIb.2), four genes in potato (one in IIIa and three in IIb.2), five genes in soybean (two in IIIa and three in IIb.2), three genes in poplar (one in IIIa and two in IIb.2), two genes in grape (one each in IIIa and in IIb.2), and two genes in papaya (one each in IIIa and in IIb.2). Understanding the function of these CPKs may help elucidate the involvement of CPKs in development and identify CPK sequence motifs or patterns that are important in development, particularly in reproductive structures such as pollen.

Group 1a monocot CPKs appear to be involved in cold stress responses. Within Group Ia, a number of monocot CPKs were shown to respond to cold stress. OsCPK13 transcripts were reported to be up-regulated by cold stress; its transcripts and protein both showed high accumulation in cold-tolerant rice varieties and this protein conferred cold tolerance when overexpressed (Abbasi et al. 2004, Komatsu et al. 2007, Ray et al. 2007). Similarly, OsCPK7 transcript accumulation increased in response to cold, based on Northern blot analysis (Saijo et al. 2000). The closest ortholog in wheat, TaCPK1, has also been shown to be up-regulated by cold stress, when measured by semi-quantitative RT–PCR (Li et al. 2008a). In contrast, TaCPK2, the closest paralog of TaCPK1 and a potential ortholog of OsCPK13, did not show any changes in the same cold stress study. Further analysis determining cold stress responses by other monocot CPKs within this group (**Fig. 6**) may verify

this observation, such as the three genes in maize and the four genes in sorghum listed in **Fig. 6**. One conifer CPK, Picea_TC127192, belongs this group (see **Supplementary Fig. S1**). Being the only gymnosperm sequence within this group, functional studies of this CPK may be important to provide further evidence on the cold sensitivity of this entire group, and in determining potential motifs that influence CPK involvement to cold stress among seed plants.

Group IIIb.2 and IIIb.3 CPKs respond to fungal infection. Several Group IIIb.2 and IIIb.3 CPKs appeared to respond to fungal infection, although most information was based only on transcript accumulation studies. In Group IIIb.2, three CPKs showed increased transcript abundance in response to different fungal infections: AtCPK10 to *Erysiphe* sp. (Swarbreck et al. 2008), TaCPK19 to *Blumeria graminis* (Li et al. 2008a) and OsCPK9 to *Magnaporthe* sp. (Asano et al. 2005). However, in Arabidopsis, AtCPK30, the paralog of AtCPK10 (86% amino acid similarity), showed no change in transcript accumulation in the same study. In Group IIIb.3, positive responses to *B. graminis* were also shown by two paralogous wheat CPKs, TaCPK3 and 15. In rice, OsCPK3 did not seem to change transcript accumulation in response to *Magnaporthe* sp. As the studies mentioned involved different types of fungi, more evidence may be required to confirm significant association of these CPKs with fungal infection. It may be valuable to study the responses of CPKs within Group IIIb.2 and IIIb.3 in different seed plants (**Fig. 6**) against certain fungi that have wide host ranges.

Group Ia and Ic.2 in eudicots respond to bacterial infection. Members of small eudicot clades within Group Ia and Ic.2 may have positive roles in immune responses to bacteria, as AtCPK5, 6 and 26 and CanCDPK3 (Group Ia) and AtCPK1 and PaCDPK1 (Group Ic.2) appeared to be important in flg22 signalling and bacterial infections (Tsai et al. 2007, Boudsocq et al. 2010, Coca and San Segundo 2010). However, the closest paralog of AtCPK1, AtCPK2 (81% similarity in amino acid), did not appear to respond to bacteria or fungi as shown in the same studies. Other plant CPKs within this group have been studied in relation to salt stress and gibberellic acid (IiCPK2) and ethylene biosynthesis (GhCPK1), but have not been examined in relation to bacterial infection. Further analysis of closely related CPKs within Group Ia and Ic.2 among different eudicot species may substantiate the potential importance of these groups in plant bacterial infections.

Discussion

Is the sequence evolution of CPKs correlated with functional diversification?

Our broadly sampled phylogenetic analysis provides insights regarding the evolution of CPKs from green algae to higher plants. By using protist sequences as the outgroup and including CPK sequences from representative green algae and basal groups of land plants (non-vascular and non-seed-bearing plants), we present how CPK genes have evolved in Viridiplantae. Is the evolution of CPKs correlated with functional diversification among plants? The following sections address this question, with specific focus on protists and green algae, basal land plant groups (bryophyte and lycophyte mosses) and higher plant groups (gymnosperms and angiosperms).

CPK diversification is distinct between protists, green algae and land plants

Based on sequence similarity and intron–exon structure, CPKs originated from the fusion of protist genes encoding a CaMK and a CaM (Harmon et al. 2000, Zhang and Choi 2001, Harper et al. 2004), but it is unclear how this ancestral CPK gene diversified into multiple gene family members among plants. Apicomplexan protists generally have only about 7–10 CPK genes, while plants have up to 50, depending on genomic complexity. Despite being in the same gene family and having the same domains, protist CPK sequences are highly distinct from plant CPKs, with an average of 70% difference in amino acid sequence according to Zhang and Choi (2001). Previous authors noted that independent CPK gene expansion events occurred between protists and plants (Nagamune and Sibley 2006, Billker et al. 2009). Apicomplexan CPKs divide into four groups, Api1–Api4 (Nagamune and Sibley 2006), which are different from the major evolutionary groups observed in plants: two groups are sister groups to plant CPKs (Api1 and Api2, which include the protist CPK sequences used in our phylogenetic analysis); Api3 and Api4 are more similar to animal CaMK (Nagamune and Sibley 2006). In our analysis, none of the CPKs from any plant or green algae clustered with the protist CPKs. With the assumption that protist and plant CPKs have come from a common ancestral CPK, our data support the hypothesis that protist and plant CPK diversification into multiple gene family groups were independent of each other.

The diversification of plant CPKs into the evolutionary Groups I–IV was not observed in green algae. Similar to the protists, green algae CPKs also have four major groups (**Fig. 3**, *c1* to *c4*), but only two of these (*c3* and *c4*) appear to be related closely to land plant CPKs. However, this clustering of algal CPKs separate from land plant CPKs could also be an artifact of long-branch attraction; there is only 53% and 36% bootstrap support for branches *c3* and *c4*, respectively. Algal CPKs are highly divergent from each other, in the same way as they are distinct from land plant CPKs because algal lineages had a longer time to diverge than land plants. Since most sequences from algae are hypothetical proteins derived from the genome sequence only, stronger evidence of their functional existence must be gathered from transcriptome and protein-based studies, as well as from additional green algae genomes. Our analysis provides an indication that green algae and land plant CPKs had a common ancestral gene, but the

diversifications of CPKs between these taxa were independent of one another.

There was an expansion of the CPK gene family during plant terrestrial transition and/or adaptation

In contrast to green algae CPKs, basal land plant CPKs (bryophyte moss and lycophytes) were distributed among the four major evolutionary groups. This suggests that the diversification of CPKs into the four groups present in extant plants may have been essential to the transition into or adaptation of plants to terrestrial life. This premise is supported by molecular clock analyses and comparison of functional information between green algae CPKs and land plants.

The timing estimated by the Bayesian molecular clock analysis was consistent with the hypothesis of CPK diversification having occurred with the transition or adaptation to terrestrial life (**Fig. 4**). The diversification time estimates were between 270 and 340 MYA, which are later than the split between green algae and land plants (700–900 MYA) (Hedges et al. 2004, Zimmer et al. 2007, Parfrey et al. 2011) and the first appearance of land plants (estimated between 400 and 700 MYA) (Raven and Edwards 2001, Gensel 2008), but is close to the point when land plants diverged into vascular and non-vascular plants (350–400 MYA) (Kenrick and Crane 1997, Yoon et al. 2004). It must be noted, however, that the timing was estimated using a relaxed molecular clock and therefore the exact time points of CPK diversification are not known. Plant CPK diversification also appeared to coincide with the Eutracheophytic epoch (256–398 MYA), which is characterized by a dramatic increase in spore and vascular plant diversity (a characteristic feature among land plants), as shown by fossil evidence (Gray 1985, Kenrick and Crane 1997). Therefore, based on molecular clock analysis, land plant CPK genes appear to have undergone sequence evolution around the periods of plant terrestrial transition and/or adaptation.

The phylogenetic tree (**Fig. 3**) and molecular clock analysis (**Fig. 4**) of CPK genes do not reflect speciation timings. Although gene duplication events and speciation can be correlated, these can have different timings and diversification rates (Lanfear et al. 2010). For large gene families such as the CPKs, the rate of molecular evolution for individual genes can be dependent on their biological function within a species (Warren et al. 2010). For example, each of the four major CPK groups included sequences from the moss *P. patens*, which may indicate that the last common ancestor of land plants had four CPK genes; but the main stem ages show some differences (268.00, 340.07, 297.74 and 271.50 MYA for Groups IV, II, III and I, respectively, as illustrated in **Fig. 2**). The duplication of CPKs into four different genes may have occurred within the genome of this common ancestor, whereas the diversification of CPKs does not appear to have happened before its speciation but rather later on after the appearance of land plants. The diversification happened sequentially, which would have been important for the evolution of land plants during the Eutracheophytic epoch. Sequential divergence among groups or classes within a gene family has also been reported for floral MADS-box genes (Nam et al. 2003) and prolamin genes (Xu and Messing 2008). Although molecular clock timings are not precise because they are based on assumptions, the inferences from them are helpful for providing general insights about gene or species evolution.

Whether coinciding with initial transition or later adaptation to terrestrial life, there is a high likelihood that CPK diversification played a role in plant terrestrial adaptation. Calcium signaling is vital in adaptive physiological processes, particularly in maintaining homeostasis and responding to stresses imposed by the environment (McLaughlin and Wimmer 1999). The transition from aquatic to terrestrial habitat brought about new physical conditions and challenges to plants, such as desiccation, reduced access to water and nutrition, abrupt temperature changes and closer contact with microorganisms. Starting from green algae, CPKs appeared to be important in maintaining cellular homeostasis. Among extant green algae, CPKs were shown to function in copper acclimatization (Contreras-Porcia et al. 2011), organization and contraction of F-actin (Sugiyama et al. 2000), activation of microsomal proteins during osmotic stress (Yuasa and Muto 1992) and cytoplasmic streaming (McCurdy and Harmon 1992). In comparison, similar functions were observed in CPKs among non-vascular plants, although there may be some additional functional specializations. One study of moss showed a CPK to be up-regulated by nutrient starvation (Mitra and Johri 2000). In the liverwort *Marchantia polymorpha*, a CPK gene was described as having a splice variant that is preferentially accumulated in the liverwort's male sexual organ (Nishiyama et al. 1999). On the other hand, in higher plants such as angiosperms, CPKs were involved in similar but more complex physiological processes, particularly in development and stress. This included pollen tube formation, hormone-regulated stomatal movement, seed development, and cellular defense pathways such as microbe-associated molecular pattern (MAMP) signaling and mitogen-activated protein kinase (MAPK) activation (Boudsocq et al. 2010, Boudsocq and Sheen 2013). In addition, structural changes unique among terrestrial plants include well-developed spore/pollen-bearing organs, stomates and water-conducting systems (Kenrick and Crane 1997); CPKs are usually abundant in these types of tissues across various types of seed plants. It appears that gene expansion and sequence diversification of CPKs into four evolutionary groups may have occurred in parallel with the increase in physiological and structural complexity among land plants as an adaptation to terrestrial life.

The CPK gene family in seed plants has undergone expansion in number and function but maintained sequence conservation.

The CPK gene family has expanded greatly from four genes in the land plant ancestor and <10 genes among extant green

algae, to about 10–20 genes among lower land plants and approximately 30–40 genes among angiosperms. From the ancestral genes of the four evolutionary groups, CPK genes have undergone several duplication events, which include WGD from seed plant, monocot, eudicot and plant family ancestors, as well as species-specific WGD events and individual gene duplications. In the phylogenetic analysis presented, the bryophyte and lycophyte CPKs were distributed among the four evolutionary groups, but did not cluster with any of the subgroups, except for Group IIb CPKs. Smaller clades that we assigned into clusters (i.e. Ia, Ib, Ic.1 and Ic.2 described in **Fig. 3**) only included CPKs from monocots and eudicots, and occasionally conifers. Each of the representative angiosperm species had at least one CPK gene within each of the 13 monocot–eudicot clades, which correlates to the increase in number of CPKs among angiosperms. This pattern of gene family expansion observed in CPKs among various plants is similar to those reported for other plant gene families involved in both development and stress responses, such as the prohibitin (PHB) gene family (Di et al. 2010), xylogucan endo-transglycosylase/hydrolase (XTH) genes (Eklöf and Brumer 2010), the Wuschel-related homeobox (WOX) gene family (Zhang et al. 2010) and the rapid alkalinization factors (RALF) gene family (Cao and Shi 2012). The massive expansion of CPK genes among angiosperms therefore is primarily a passive effect of polyploidization events that occurred from the last common ancestors of seed plants, angiosperms, monocot and eudicots.

CPK sequences among all the land plants included in our analysis are highly conserved, particularly in the PK domain, the AJ and the CBD. The average amino acid pairwise identity within these regions between all 357 CPK sequences used in the analysis (including the five protist CPKs) was 55%. Among all land plant CPKs, the average pairwise identity was 58%. The pairwise identities of CPKs in our analysis are similar to that of the highly conserved Hsp70 gene family, with 45% identity between protists, animals and plants (Boorstein et al. 1994, Daugaard et al. 2007, Murphy 2013) and to eIF2 alpha in plants, which have >50% identity between animals, yeasts and plants (Immanuel et al. 2012). Moreover, the ratio of non-synonymous (Ka) to synonymous (Ks) nucleotide substitution rates (Ka/Ks ratio) among Arabidopsis CPKs in these regions demonstrates evolutionary pressure for these sequences to be maintained. Among all the 34 AtCPKs, the Ka/Ks ratios ranged from 0.09 to 0.5 (data not shown). This indicates that these CPK genes are under purifying (stabilizing) selection. Even though there has been significant expansion of CPK genes, the evolutionary pressure to maintain high sequence conservation in the PK domain, the AJ and the CBD contributes to various examples of functional similarity, redundancy and overlap among many CPKs. A Ka/Ks calculation could not be undertaken within the N-VD and CT domain; as noted earlier these regions are highly variable (in both character and length). Sequence evolution that brought about distinctive function and localization in many extant CPKs may have occurred mostly within the N-VD and CT.

Several processes drive the preservation of duplicated genes among extant organisms, including functional retention, pseudogenization (non-functionalization), neofunctionalization and subfunctionalization (Konrad et al. 2011). Duplication of genes results in increased gene dosage, which needs to be balanced to retain or improve species fitness. Dosage balance imposes selective pressure for genes to retain, lose, gain or modify function and localization (Ray et al. 2007, Konrad et al. 2011). Some duplicated genes retain full function if higher dosage increases fitness. On the other hand, some duplicated genes may lose all functionality and yet be retained in the genome, resulting in pseudogenes (Hughes 1994, Konrad et al. 2011). Neofunctionalization refers to the acquisition of novel functions by a duplicated gene, while subfunctionalization involves the complementary loss and retention of some ancestral functions so that both duplicated genes are retained (Hughes 1994, Konrad et al. 2011). Among CPKs, no specific functional differences or pattern were observed between major CPK evolutionary groups. Some CPKs respond exclusively to certain types of developmental, biotic or abiotic stress, or to a specific combination of these. As mentioned in the Results, some closely related CPKs have the same function (functional retention), while some have opposing expression patterns or totally different functions (functional divergence).

CPK functional diversification events may be ancestral as they are shared by a wide range of taxa from monocots and dicots (development and osmotic pressure response), while some may be recent, as they are unique to a species or shared among closely related taxa (such as cold, fungal and bacterial response). These events depend on the environmental constraints that the plants have been exposed to and may have arisen on several occasions. For example, ancestral lineages of modern plant species may have gone through several rounds of adaptation to temperature during successive ice ages. As the evolutionary groupings can be considered as the outcome of duplications among ancient genes, we can hypothesize that the ancestral CPK genes of plants had multiple functions, both in the development of reproductive structures and in the maintenance of cellular homeostasis; however, due to multiple duplication events, these have subfunctionalized into either developmental or osmotic stress response and neofunctionalized in response to terrestrial life challenges and changing environments such as temperature, drought, infections and herbivory.

The hypothesis given above may be tested using various approaches; one such approach could be examining the sequence and functional divergence among the most conserved members of this gene family. Group IIb CPKs seem to be most highly conserved compared with the others, as this group has the shortest branches on the phylogenetic tree, on average. Moreover, within this subgroup, there are moss, lycophyte and conifer members. In Arabidopsis, there appears to be a functional distinction between the members of this CPK group. AtCPK17 and 34 (Group IIb.2) function primarily in pollen development, while AtCPK3 (Group IIb.1) is involved

in various biotic and abiotic responses. AtCPK17 and 34 transcripts were undetectable in plant vegetative tissues across all developmental stages and stress treatments (Swarbreck et al. 2008) except in flower and pollen where they were extremely abundant. Conversely, AtCPK3 has moderate to high abundance in vegetative tissue but very low abundance in flowers and pollen. Further analysis of these genes and their orthologs in other species and observation of their effect when ectopically expressed in the organs of their subgroup counterpart may help elucidate the diversification of function among CPKs.

Can we predict CPK functions based on homology?

The extensive expansion of the CPK gene family during plant evolution has resulted in multiple CPK genes within genomes, belonging to different evolutionary groups. Both phylogenetic and gene structure analyses show that CPK sequences are highly conserved, with no obvious functional patterns among major evolutionary groups. The lack of pattern with respect to function and evolutionary history makes it challenging to find true functional orthologs of CPKs between species, and to predict the function of newly identified CPK sequences within a genome.

Our phylogenetic analysis and review of expression and functional information of CPKs presents a detailed view of potential CPK orthologs among agriculturally important plant species (**Figs. 3**, **6**; **Supplementary Figs. S1–S4**). Close CPK homologs with highly similar functions are present among the monocot–eudicot clades. Our meta-analysis identified several potentially orthologous groups: two that function in development, Group IIb.2 and IIIa; one that is mainly responsive to cold stress, Group Ia (monocots only); two that show response to fungal infection, Group IIIb.2 and IIIb.3; and two small eudicot clades within Group Ia and Ic.2 that respond to bacteria. There is a need to verify functional orthology within these groups by examining the function of closely related CPKs in other plant species.

Gene structure analyses of bryophyte, monocot and dicot genome representatives support the phylogenetic trees constructed. CPK sequences within a genome showed similar intron–exon patterns when they were part of the same evolutionary group. Highly homologous CPKs have very similar intron–exon patterns, although this is not true in all cases. One example of this is AtCPK4 and 11 compared with AtCPK12. These three sequences have high homology in terms of protein sequence. AtCPK4 and 11 have very similar intron–exon patterns as well, but the AtCPK12 gene structure show some differences from these two. In terms of function, AtCPK4 and 11 show functional similarity, but AtCPK12 does not. These findings indicate the potential use of gene structure and phylogenetic analysis together to predict functional specificity between paralogs. Highly homologous sequences with similar gene structure will most probably show similarity in function. However, this approach may not be appropriate to find functional orthologs, as the intron–exon patterns may greatly vary between different genomes.

It must be noted that our phylogenetic analysis involved only the conserved regions (PK domain, AJ and CBD) of CPKs. As we did not identify any functional patterns using these domains, the functional specificity of a particular CPK may be partly due to short motifs within these regions but is probably more largely due to the hypervariable N-VD and CT domain (Hrabak et al. 2003). Evolutionary analysis that includes these regions, however, makes multiple alignment and tree construction of all the CPKs from different species difficult due to their extreme sequence variability. Functional divergence may also be due to sequence evolution within the promoter regions; evolution of the three-dimensional structures of the proteins which are not evident in the primary protein sequences; or co-factors that modify function.

The difficulty in identifying functional patterns and orthologs may also reflect the scarcity of functional information about CPKs, particularly in biotic stress responses (**Fig. 5c**). There is very little information on the role of CPKs in pathogen defense, particularly responses associated with herbivory and/or viral infection. In addition, most information regarding CPK function is based on transcript accumulation, so there is a need to support or validate these data through protein activity, interaction-based and mutation-based experiments. Functional information on moss, lycophyte and gymnosperm CPKs is also scarce despite their importance in elucidating CPK diversification from green algae to higher plants.

Conclusion

The evolution of CPKs appears to have occurred in parallel with the terrestrial transition of plants. CPK evolution is characterized by expansion of this gene family from green algae to higher plants, with the diversification of CPKs into four major groups only seen among land plants. The amount and diversity of CPKs among seed plants arose from ancestral and genome-specific WGD events, as well as gene-specific duplication and deletions. From green algae to higher plants, CPK function is primarily in signaling cascades involved in osmotic pressure and cytoplasmic movements. These functions diversified with land plant evolution in response to osmotic, developmental, nutritional and immunological challenges imposed by the new and constantly evolving terrestrial environment.

Despite gene family expansion, parts of plant CPK gene sequences appear to be highly conserved, which could explain redundancy in function between and within its evolutionary groups. Even in certain closely related CPKs within a genome, few obvious functional patterns were found within the conserved regions of their encoded proteins. We suggest that CPK gene explosion among higher plants is largely a result of the polyploidization events that occurred during plant evolution. CPKs have subfunctionalized and neofunctionalized into different developmental and stress responses. The sequence evolution of the PK domain, AJ and CBD, upon which most of the CPK evolutionary analyses are based, is not sufficient for

functional classification. What, then, defines functional specificity and similarity among CPKs? How can we predict their function in response to a stimulus or stress response? Is it influenced only by very few differences in the amino acid sequence, by short motifs or predominantly by differences in gene regulatory factors? Functional prediction by homology of the primary sequences may be insufficient in searching for CPK orthologs. Further research examining stress-specific motifs within the protein sequences, in combination with protein structural studies, promoter region analysis and targeted functional studies of the orthologous CPK groups, will be important to elucidate a more obvious link between the functional and sequence diversity among CPKs.

Materials and Methods

Mining of CPK sequences

To simplify the search without compromising sensitivity and specificity, only five representative CPK genes were used as query protein sequences. These included sequences from each of the four major evolutionary groups of AtCPKs: AtCPK1, 21, 8 and 16 and a consensus sequence of the 34 Arabidopsis CPKs. AtCPK1, 21, 8 and 16 have the highest percentage pairwise identity within Arabidopsis CPK groups I, II, III and IV, respectively. BLASTp and tBLASTn searches were undertaken using default parameters to identify putative CPK sequences from the selected genomes: *V. carteri* (v1.0), *C. reinhardtii* (v4 and v4 u10.2), *P. patens* (v1.1), *S. moellendorfii* (v1.0), *Picea sitchensis* [expressed sequence tag (EST) data, Dana Farber Cancer Institute (DFCI) http://compbio.dfci.harvard.edu/cgi-bin/tgi], *O. sativa* [Michigan State University (MSU) release 6.0], *T. aestivum* [EST data, The Institute for Genomic Research (TIGR) r2], *S. bicolour* (Sbi1.4 assembly), *Z. mays* (v4 and v5a.59), *A. thaliana* [The Arabidopsis Information Resource (TAIR) r10], *V. vinifera* (8× and 12× assembly), *G. max* (v1.0), *P. trichocarpa* (v1.1 and 1.2), *C. papaya* (Hawaii Papaya Genome Project v2007) and *S. tuberosum* [Potato Genome Sequencing Consortium (PGSC), http://potatogenome.net]. Hits with significant similarity were classified as CPKs using three criteria: (i) a cut-off BLAST score of at least 250 and an E-value of e^{-2} or less; (ii) presence of the five CPK domains: N-VD, PK, AJ, CAM-like and CT variable domain; and (iii) having 1–5 non-degenerate (functional) EF-hands within the calmodulin-like domain. All of the CPK-like sequences detected in the BLAST searches that have degenerate EF-hands as determined by InterProScan were excluded from this study. Complete information including accession numbers and/or gene ID for each CPK sequence retrieved during this search is provided in **Supplementary Table S1**. When alternative splicing variants were present, only one protein sequence was chosen (the one with the longest sequence) to be included in the analysis.

Notes regarding nomenclature

CPKs are also known as 'calmodulin-like domain protein kinases'. Based on the nomenclature proposed by Hrabak et al. (2003), their collective name is abbreviated to CDPKs, while the names for their genes and proteins are indicated by the first letter of the genus (in upper case) and species (in lower case), followed by the abbreviation 'CPK' and a number (e.g. AtCPK1). Most authors have followed this nomenclature style; however, some authors have used 'CDPK' in naming sequences (e.g. TgCDPK1) (Kugelstadt et al. 2007, Billker et al. 2009, Jaworski et al. 2010, Wernimont et al. 2010, Wernimont et al. 2011). In this paper, for consistency, the abbreviation 'CPK' was used for all sequences but the assigned numbers were retained (e.g. OsCDPKs to OsCPKs). For genomes for which published CPKs were available, we continued the CPK nomenclature. For genomes where no CPKs have been published previously, the CPK genes/proteins are designated with the genome/locus ID (e.g. Sb6g026530), to avoid potential confusion with any concurrent research to identify CPKs from these genomes.

The 'calcium-binding domain', which we abbreviate as 'CBD', includes only the domain with calcium-binding EF-hands. We have referred to the autoinhibitory junction (AJ) as a separate domain. The CBD is not equivalent to the 'CDPK-activation domain', which was designated as 'CAD' by Wernimont et al. (2010). CAD includes the AJ domain as an extension of the calcium-binding domain.

Multiple sequence alignment and phylogenetic analysis

Multiple sequence alignments and phylogenetic analyses were carried out using GeneiousPro 5.6 (Geneious version 5.6). A total of 352 plant CPK protein sequences were aligned using the ClustalW program (Larkin et al. 2007). Five apicomplexan CPKs consisting of TgCPK1 (ToxoDB ID 162.m00001), TgCPK3 (ToxoDB ID 541.m00134), PfCPK3 (PlasmoDB ID PFC0420w), CpCPK1 (CryptoDB ID cgd3_920) and CpCPK3 (CryptoDB ID cgd5_820) were included in the alignments and used as outgroups. Poorly aligned regions were manually removed and the alignments used for phylogenetic analyses only included the PK domain, AJ and the calmodulin-like domain (**Fig. 1**). For **Fig. 3**, Distance (Jukes–Cantor model) and Likelihood [Whelan and Goldman (WAG) model] trees were constructed using the NJ (Geneious Tree Builder) and ML (PhyML) methods, respectively, with 1,000 bootstrap replicates. Consensus trees of the two methods show similar topology but only the ML tree is shown in **Fig. 3**. Trees were viewed and colored using FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

Bayesian (MCMC analysis in the software BEAST (Drummond and Rambaut 2007) was used to construct a tree with a relaxed molecular clock. Calibration points were set at 1,700 MYA for apicomplexan (Hedges 2002, Hedges et al. 2004) and 1,400 MYA for green algae (Yoon et al. 2004) most recent common ancestors (MRCAs). The Tree Prior used was the Yule model.

For **Fig. 6**, CPK sequences for which published sequence information was available were aligned with ClustalW using

GeneiousPro 5.6. The NJ tree was constructed using Geneious Tree Builder, with default parameters.

Gene structure analysis

The intron–exon organizations of Arabidopsis, rice and bryophyte CPKs were illustrated using the online tool Gene Structure Display Server (http://gsds.cbi.pku.edu.cn/). The corresponding cDNA and unspliced gene sequences of these CPKs were obtained from Phytozome (http://www.phytozome.net/).

Analysis of CPK gene expression in Arabidopsis

For CPK expression and function, two approaches were used. First, Arabidopsis CPK transcript accumulation levels were analysed using Affymetrix 22K microarray data available in TAIR (Swarbreck et al. 2008) and the online platform Genevestigator V3 (https://www.genevestigator.com/gv/index.jsp). Secondly, experimental data on specific CPK responses to biotic and abiotic responses, hormones, developmental signals and other genes were collated from the literature.

Supplementary data

Supplementary data are available at PCP online.

Funding

This work was supported by the School of Applied Sciences, Auckland University of Technology, New Zealand; Vice Chancellor's Scholarship, Auckland University of Technology, New Zealand; The Agricultural and Marketing Research and Development Trust (AGMARDT) Post-Doctoral Fellowship, New Zealand; The New Zealand Institute for Plant & Food Research Limited, New Zealand.

Acknowledgments

The authors would like to thank Dr. Richard Newcomb, Dr. Revel Drummond and Tracey Immanuel for their valuable suggestions on this manuscript, and Dr. Alexei Drummond for his help with the molecular clock analysis.

Disclosures

The authors have no conflicts of interest to declare.

References

Abbasi, F., Onodera, H., Toki, S., Tanaka, H. and Komatsu, S. (2004) OsCDPK13, a calcium-dependent protein kinase gene from rice, is induced by cold and gibberellin in rice leaf sheath. *Plant Mol. Biol.* 55: 541–552.

Asano, T., Hakata, M., Nakamura, H., Aoki, N., Komatsu, S., Ichikawa, H. et al. (2011) Functional characterisation of OsCPK21, a calcium-dependent protein kinase that confers salt tolerance in rice. *Plant Mol. Biol.* 75: 179–191.

Asano, T., Tanaka, N., Yang, G., Hayashi, N. and Komatsu, S. (2005) Genome-wide identification of the rice calcium-dependent protein kinase and its closely related kinase gene families: comprehensive analysis of the CDPKs gene family in rice. *Plant Cell Physiol.* 46: 356–366.

Asano, T., Wakayama, M., Aoki, N., Komatsu, S., Ichikawa, H., Hirochika, H. et al. (2010) Overexpression of a calcium-dependent protein kinase gene enhances growth of rice under low-nitrogen conditions. *Plant Biotechnol.* 27: 369–373.

Babu, M., Griffiths, J., Huang, T.-S. and Wang, A. (2008) Altered gene expression changes in Arabidopsis leaf tissues and protoplasts in response to Plum pox virus infection. *BMC Genomics* 9: 325.

Baillie, B.K., Belda-Baillie, C.A. and Maruyama, T. (2000) A calcium-dependent protein kinase functions in wound healing in Ventricaria ventricosa (chlorophyta). *J. Phycol.* 36: 1145–1152.

Banks, J.A., Nishiyama, T., Hasebe, M., Bowman, J.L., Gribskov, M., dePamphilis, C. et al. (2011) The selaginella genome identifies genetic changes associated with the evolution of vascular plants. *Science* 332: 960–963.

Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. and Sayers, E.W. (2010) GenBank. *Nucleic Acids Res.* 38: D46–D51.

Billker, O., Lourido, S. and Sibley, L.D. (2009) Calcium-dependent signaling and kinases in apicomplexan parasites. *Cell Host Microbe* 5: 612–622.

Boorstein, W.R., Ziegelhoffer, T. and Craig, E.A. (1994) Molecular evolution of the Hsp70 multigene family. *J. Mol. Evol.* 38: 1–17.

Boudsocq, M. and Sheen, J. (2013) CDPKs in immune and stress signaling. *Trends Plant Sci.* 18: 13–40.

Boudsocq, M., Willmann, M., McCormack, M., Lee, H., Shan, L., He, P. et al. (2010) Differential immune signalling via calcium sensor protein kinases. *Nat. Lett.* 464: 418–423.

Bowers, J.E., Chapman, B.A., Rong, J. and Paterson, A.H. (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422: 433–438.

Broad Institute of Harvard and MIT (2010) *Phytophthora infestans* Sequencing Project. http://www.broadinstitute.org.

Cao, J. and Shi, F. (2012) Evolution of the RALF gene family in plants: gene duplication and selection patterns. *Evol. Bioinform.* 8: 271–292.

Chang, W., Fu, G., Chen, X., Zhu, J. and Zhang, Z. (2011) Biochemical characterization of a calcium-sensitive protein kinase LeCPK2 from tomato. *Indian J. Biochem. Biophys.* 48: 148–153.

Chang, W.J., Su, H.S., Li, W.J. and Zhang, Z.L. (2009) Expression profiling of a novel calcium-dependent protein kinase gene, LeCPK2, from tomato (Solanum lycopersicum) under heat and pathogen-related hormones. *Biosci. Biotechnol. Biochem.* 73: 2427–2431.

Cheng, S.-H., Willmann, M., Chen, H.-C. and Sheen, J. (2002) Calcium signaling through protein kinases: the *Arabidopsis* calcium-dependent protein kinase gene family. *Plant Physiol.* 129: 469–485.

Coca, M. and San Segundo, B. (2010) AtCPK1 calcium-dependent protein kinase mediates pathogen resistance in Arabidopsis. *Plant J.* 63: 526–540.

Contreras-Porcia, L., Dennett, G., Gonzalez, A., Vergara, E., Medina, C., Correa, J.A. et al. (2011) Identification of copper-induced genes in the marine alga Ulva compressa (Chlorophyta). *Mar. Biotechnol.* 13: 544–556.

Daugaard, M., Rohde, M. and Jaattela, M. (2007) The heat shock protein 70 family: highly homologous proteins with overlapping and distinct functions. *FEBS Lett.* 581: 3702–3710.

DeFalco, T.A., Bender, K.W. and Snedden, W.A. (2010) Breaking the code: Ca^{2+} sensors in plant signalling. *Biochem. J.* 425: 27–40.

Di, C., Xu, W., Su, Z. and Yuan, J.S. (2010) Comparative genome analysis of PHB gene family reveals deep evolutionary origins and diverse gene function. *BMC Bioinformatics* 11(Suppl. 6), S22.

Drummond, A.J. and Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* 7: 214.

Du, H., Wang, Y.-B., Xie, Y., Liang, Z., Jiang, S.-J., Zhang, S.-S. et al. (2013) Genome-wide identification and evolutionary and expression analyses of MYB-related genes in land plants. *DNA Res.* 20: 437–448.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32: 1792–1797.

Eklöf, J.M. and Brumer, H. (2010) The XTH gene family: an update on enzyme structure, function, and phylogeny in xyloglucan remodeling. *Plant Physiol.* 153: 456–466.

Engstrom, E.M. (2011) Phylogenetic analysis of GRAS proteins from moss, lycophyte and vascular plant lineages reveals that GRAS genes arose and underwent substantial diversification in the ancestral lineage common to bryophytes and vascular plants. *Plant Signal. Behav.* 6: 850–854.

Estruch, J.J., Kadwell, S., Merlin, E. and Crossland, L. (1994) Cloning and characterization of a maize pollen-specific calcium-dependent calmodulin-independent protein kinase. *Proc. Natl Acad. Sci. USA* 91: 8837–8841.

Fukuzawa, H., Kubo, T. and Yamano, T. (2008) Genome of a green alga, Chlamydomonas reinhardtii, lights up key functions of plant and animal cells. *Tanpakushitsu kakusan koso* 53: 1133–1143.

Gargantini, P.R., Giammaria, V., Grandellis, C., Feingold, S.E., Maldonado, S. and Ulloa, R.M. (2009) Genomic and functional characterization of StCDPK1. *Plant Mol. Biol.* 70: 153–172.

Gensel, P.G. (2008) The earliest land plants. *Annu. Rev. Ecol. Evol. Syst.* 39: 459–477.

Geneious version 5.6 created by Biomatters, available from http://www.geneious.com/.

Giammaria, V., Grandellis, C., Bachmann, S., Gargantini, P.R., Feingold, S.E., Bryan, G. et al. (2011) StCDPK2 expression and activity reveal a highly responsive potato calcium-dependent protein kinase involved in light signalling. *Planta* 233: 593–609.

Gill, N., Findley, S., Walling, J.G., Hans, C., Ma, J., Doyle, J. et al. (2009) Molecular and chromosomal evidence for allopolyploidy in soybean. *Plant Physiol.* 151: 1167–1174.

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J. et al. (2012) Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* 40: D1178–D1186.

Grabarek, Z. (2006) Structural basis for diversity of the EF-hand calcium-binding proteins. *J. Mol. Biol.* 359: 509–525.

Gray, J. (1985) The microfossil record of early land plants: advances in understanding of early terrestrialization, 1970–1984. *Philos. Trans. R. Soc. B: Biol. Sci.* 309: 167–195.

Hardie, D.G. (1999) Plant protein serine/threonine kinases: classification and functions. *Annu. Rev. Plant Biol.* 50: 97–131.

Harmon, A.C., Gribskov, M., Gubrium, E. and Harper, J.F. (2001) The CDPK superfamily of protein kinases. *New Phytol.* 151: 175–183.

Harmon, A.C., Gribskov, M. and Harper, J.F. (2000) CDPKs—a kinase for every Ca2+ signal?. *Trends Plant Sci.* 5: 154–159.

Harper, J.R., Breton, G. and Harmon, A. (2004) Decoding Ca(2+) signals through plant protein kinases. *Annu. Rev. Plant Biol.* 55: 263–288.

Hashimoto, K. and Kudla, J. (2011) Calcium decoding mechanisms in plants. *Biochimie* 93: 2054–2059.

Hedges, S.B. (2002) The origin and evolution of model organisms. *Nat. Rev. Genet.* 3: 838–849.

Hedges, S.B., Blair, J.E., Venturi, M.L. and Shoe, J.L. (2004) A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evol. Biol.* 4: 2.

Hrabak, E., Chan, C., Gribskov, M., Harper, J., Choi, J., Halford, N. et al. (2003) The Arabidopsis CDPK-SnRK superfamily of protein kinases. *Plant Physiol.* 132: 666–680.

Hughes, A.L. (1994) The evolution of functionally novel proteins after gene duplication. *Proc. R. Soc. B: Biol. Sci.* 256: 119–124.

Immanuel, T.M., Greenwood, D.R. and MacDiarmid, R.M. (2012) A critical review of translation initiation factor eIF2 alpha kinases in plants—regulating protein synthesis during stress. *Funct. Plant Biol.* 39: 717–735.

Jaillon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A. et al. (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449: 463–467.

Jaworski, K., Pawełek, A., Szmidt-Jaworska, A. and Kopcewicz, J. (2010) Expression of calcium-dependent protein kinase gene (PnCDPK1) is affected by various light conditions in Pharbitis nil seedlings. *J. Plant Growth Regul.* 29: 316–327.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E. et al. (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature* 473: 97–100.

Kenrick, P. and Crane, P.R. (1997) The origin and early evolution of plants on land. *Nature* 389: 33–39.

Kiselev, K.V., Turlenko, A.V. and Zhuravlev, Y.N. (2010) Structure and expression profiling of a novel calcium-dependent protein kinase gene PgCDPK1a in roots, leaves, and cell cultures of Panax ginseng. *Plant Cell Tiss. Organ Cult.* 103: 197–204.

Klimecka, M. and Muszyńska, G. (2007) Structure and functions of plant calcium-dependent protein kinases. *Acta Biochim. Pol.* 54: 219–233.

Kobayashi, M., Ohura, I., Kawakita, K., Yokota, N., Fujiwara, M., Shimamoto, K. et al. (2007) Calcium-dependent protein kinases regulate the production of reactive oxygen species by potato NADPH oxidase. *Plant Cell* 19: 1065–1080.

Komatsu, S., Yang, G., Khan, M., Onodera, H., Toki, S. and Yamaguchi, M. (2007) Over-expression of calcium-dependent protein kinase 13 and calreticulin interacting protein 1 confers cold tolerance on rice plants. *Mol. Genet. Genomics* 277: 713–723.

Konrad, A., Teufel, A.I., Grahnen, J.A. and Liberles, D.A. (2011) Toward a general model for the evolutionary dynamics of gene duplicates. *Genome Biol. Evol.* 3: 1197–1209.

Kugelstadt, D., Winter, D., Pluckhahn, K., Lehmann, W.D. and Kappes, B. (2007) Raf kinase inhibitor protein affects activity of Plasmodium falciparum calcium-dependent protein kinase 1. *Mol. Biochem. Parasitol.* 151: 111–117.

Lanfear, R., Welch, J.J. and Bromham, L. (2010) Watching the clock: studying variation in rates of molecular evolution between species. *Trends Ecol. Evol.* 25: 495–503.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H. et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947–2948.

Li, A., Wang, X., Leseberg, C.H., Jia, J. and Mao, L. (2008a) Biotic and abiotic stress responses through calcium-dependent protein kinase (CDPK) signaling in wheat (Triticum aestivum L.). *Plant Signal. Behav.* 3: 654–656.

Li, A.L., Zhu, Y.F., Tan, X.M., Wang, X., Wei, B., Guo, H.Z. et al. (2008b) Evolutionary and functional study of the CDPK gene family in wheat (Triticum aestivum L.). *Plant Mol. Biol.* 66: 429–443.

Magallon, S. and Castillo, A. (2009) Angiosperm diversification through time. *Amer. J. Bot.* 96: 349–365.

McCurdy, D.W. and Harmon, A.C. (1992) Calcium-dependent protein kinase in the green alga Chara. *Planta* 188: 54–61.

McLaughlin, S.B. and Wimmer, R. (1999) Tansley Review No. 104. Calcium physiology and terrestrial ecosystem processes. *New Phytol.* 142: 373–417.

Ming, R., Yu, Q., Moore, P.H., Paull, R.E., Chen, N.J., Wang, M.L. et al. (2012) Genome of papaya, a fast growing tropical fruit tree. *Tree Genet. Genomes* 8: 445–462.

Mitra, D. and Johri, M.M. (2000) Enhanced expression of a calcium-dependent protein kinase from the moss Funaria hygrometrica under nutritional starvation. *J. Biosci.* 25: 331–338.

Morello, L., Frattini, M., Giani, S., Christou, P. and Breviario, D. (2000) Overexpression of the calcium-dependent protein kinase OsCDPK2 in transgenic rice is repressed by light in leaves and disrupts seed development. *Transgen. Res.* 9: 453–462.

Murillo, I., Jaeck, E., Cordero, M.J. and San Segundo, B. (2001) Transcriptional activation of a maize calcium-dependent protein kinase gene in response to fungal elicitors and infection. *Plant Mol. Biol.* 45: 145–158.

Murphy, M.E. (2013) The HSP70 family and cancer. *Carcinogenesis* 34: 1181–1188.

Myers, C., Romanowsky, S.M., Barron, Y.D., Garg, S., Azuse, C.L., Curran, A. et al. (2009) Calcium-dependent protein kinases regulate polarized tip growth in pollen tubes. *Plant J.* 59: 528–539.

Nagamune, K. and Sibley, L.D. (2006) Comparative genomic and phylogenetic analyses of calcium ATPases and calcium-regulated proteins in the Apicomplexa. *Mol. Biol. Evol.* 23: 1613–1627.

Nam, J., dePamphilis, C.W., Ma, H. and Nei, M. (2003) Antiquity and evolution of the MADS-box gene family controlling flower development in plants. *Mol. Biol. Evol.* 20: 1435–1447.

Nishiyama, R., Mizuno, H., Okada, S., Yamaguchi, T., Takenaka, M., Fukuzawa, H. et al. (1999) Two mRNA species encoding calcium-dependent protein kinases are differentially expressed in sexual organs of Marchantia polymorpha through alternative splicing. *Plant Cell Physiol.* 40: 205–212.

Parfrey, L.W., Grant, J., Tekle, Y.I., Lasek-Nesselquist, E., Morrison, H.G., Sogin, M.L. et al. (2010) Broadly sampled multigene analyses yield a well-resolved eukaryotic tree of life. *Syst. Biol.* 59: 518–533.

Parfrey, L.W., Lahr, D.J.G., Knoll, A.H. and Katz, L.A. (2011) Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc. Natl Acad. Sci. USA* 108: 13624–13629.

Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H. et al. (2009) The Sorghum bicolor genome and the diversification of grasses. *Nature* 457: 551–556.

Paterson, A.H., Freeling, M., Tang, H. and Wang, X. (2010) Insights from the comparison of plant genome sequences. *Annu. Rev. Plant Biol.* 61: 349–372.

Potato Genone Sequencing Consortium, Xu, X., Pan, S., Cheng, S., Zhang, B., Mu, D., et al. (2011) Genome sequence and analysis of the tuber crop potato. *Nature* 475: 189–195.

Raices, M., Chico, J.M., Tellez-Inon, M.T. and Ulloa, R.M. (2001) Molecular characterization of StCDPK1, a calcium-dependent protein kinase from Solanum tuberosum that is induced at the onset of tuber development. *Plant Mol. Biol.* 46: 591–601.

Ralph, S.G., Chun, H.J.E., Kolosova, N., Cooper, D., Oddy, C., Ritland, C.E. et al. (2008) A conifer genomics resource of 200,000 spruce (Picea spp.) ESTs and 6,464 high-quality, sequence-finished full-length cDNAs for Sitka spruce (Picea sitchensis). *BMC Genomics* 9: 484.

Raven, J.A. and Edwards, D. (2001) Roots: evolutionary origins and biogeochemical significance. *J. Exp. Bot.* 52: 381–401.

Ray, S., Agarwal, P., Arora, R., Kapoor, S. and Tyagi, A.K. (2007) Expression analysis of calcium-dependent protein kinase gene family during reproductive development and abiotic stress conditions in rice (Oryza sativa L. ssp. indica). *Mol. Genet. Genomics* 278: 493–505.

Rensing, S.A., Lang, D., Zimmer, A.D., Terry, A., Salamov, A., Shapiro, H. et al. (2008) The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. *Science* 319: 64–69.

Rouard, M., Guignon, V., Aluome, C., Laporte, M.A., Droc, G., Walde, C. et al. (2011) GreenPhylDB v2.0: comparative and functional genomics in plants. *Nucleic Acids Res.* 39: D1095–D1102.

Rutschmann, F., Stalder, U., Piotrowski, M., Oecking, C. and Schaller, A. (2002) LeCPK1, a calcium-dependent protein kinase from tomato. Plasma membrane targeting and biochemical characterization. *Plant Physiol.* 129: 156–168.

Saijo, Y., Hata, S., Kyozuka, J., Shimamoto, K. and Izui, K. (2000) Over-expression of a single Ca2+-dependent protein kinase confers both cold and salt/drought tolerance on rice plants. *Plant J.* 23: 319–327.

Schmutz, J., Cannon, S.B., Schlueter, J., Ma, J., Mitros, T., Nelson., W. et al. (2010) Genome sequence of the palaeopolyploid soybean. *Nature* 463: 178–183.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S. et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* 326: 1112–1115.

Soltis, D.E., Albert, V.A., Leebens-Mack, J., Bell, C.D., Paterson, A.H., Zheng, C. et al. (2009) Polyploidy and angiosperm diversification. *Amer. J. Bot.* 96: 336–348.

Sugiyama, K.-i., Mori, I.C., Takahashi, K., Muto, S. and Shihira-Ishikawa, I. (2000) A calcium-dependent protein kinase functions in wound healing in *Ventricaria ventricosa* (Chlorophyta). *J. Phycol.* 36: 1145–1152.

Swarbreck, D., Wilks, C., Lamesch, P., Berardini, T.Z., Garcia-Hernandez, M., Foerster, H. et al. (2008) The Arabidopsis Information Resource (TAIR): gene structure and function annotation. *Nucleic Acids Res.* 36: D1009–D1014.

Szczegielniak, J., Klimecka, M., Liwosz, A., Ciesielski, A., Kaczanowski, S., Dobrowolska, G. et al. (2005) A wound-responsive and phospholipid-regulated maize calcium-dependent protein kinase. *Plant Physiol.* 139: 1970–1983.

Tai, S.S., Liu, G.S., Sun, Y.H. and Chen, J. (2009) Cloning and expression of calcium-dependent protein kinase (CDPK) gene family in common tobacco (Nicotiana tabacum). *Agric. Sci. China* 8: 1448–1457.

Takezawa, D., Patil, S., Bhatia, A. and Poovaiah, B.W. (1996) Calcium-dependent protein kinase genes in corn roots. *J. Plant Physiol.* 149: 329–335.

Tang, H., Wang, X., Bowers, J.E., Ming, R., Alam, M. and Paterson, A.H. (2008) Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. *Genome Res.* 18: 1944–1954.

Taylor, T.N., Kerp, H. and Hass, H. (2005) Life history biology of early land plants: deciphering the gametophyte phase. *Proc. Natl Acad. Sci. USA* 102: 5892–5897.

Tsai, T.M., Chen, Y.R., Kao, T.W., Tsay, W.S., Wu, C.P., Huang, D.D. et al. (2007) PaCDPK1, a gene encoding calcium-dependent protein

kinase from orchid, Phalaenopsis amabilis, is induced by cold, wounding, and pathogen challenge. *Plant Cell Rep.* 26: 1899–1908.

Tuskan, G.A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U. et al. (2006) The genome of black cottonwood, Populus trichocarpa (Torr. & Gray). *Science* 313: 1596–1604.

Warren, A.S., Anandakrishnan, R. and Zhang, L. (2010) Functional bias in molecular evolution rate of Arabidopsis thaliana. *BMC Evol. Biol.* 10: 125.

Wernimont, A.K., Amani, M., Qiu, W., Pizarro, J.C., Artz, J.D., Lin, Y.H. et al. (2011) Structures of parasitic CDPK domains point to a common mechanism of activation. *Proteins* 79: 803–820.

Wernimont, A.K., Artz, J.D., Finerty, P., Lin, Y.H., Amani, M., Allali-Hassani, N. et al. (2010) Structures of apicomplexan calcium-dependent protein kinases reveal mechanism of activation by calcium. *Nat. Struct. Mol. Biol.* 17: 596–601.

Woodhouse, M.R., Schnable, J.C., Pedersen, B.S., Lyons, E., Lisch, D., Subramaniam, S. et al. (2010) Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. *PLoS Biol.* 8: e1000409.

Xu, J.-H. and Messing, J. (2008) Organization of the prolamin gene family provides insight into the evolution of the maize genome and gene duplications in grass species. *Proc. Natl Acad. Sci. USA* 105: 14330–14335.

Ye, S., Wang, L., Xie, W., Wan, B., Li, X. and Lin, Y. (2009) Expression profile of calcium-dependent protein kinase (CDPKs) genes during the whole lifespan and under phytohormone treatment conditions in rice (Oryza sativa L. ssp. indica). *Plant Mol. Biol.* 70: 311–325.

Yoon, G.M., Dowd, P.E., Gilroy, S. and McCubbin, A.G. (2006) Calcium-dependent protein kinase isoforms in Petunia have distinct functions in pollen tube growth, including regulating polarity. *Plant Cell* 18: 867–878.

Yoon, H.S., Hackett, J.D., Ciniglia, C., Pinto, G. and Bhattacharya, D. (2004) A molecular timeline for the origin of photosynthetic eukaryotes. *Mol. Biol. Evol.* 21: 809–818.

Yu, X.C., Li, M.J., Gao, G.F., Feng, H.Z., Geng, X.Q., Peng, C.C. et al. (2006) Abscisic acid stimulates a calcium-dependent protein kinase in grape berry. *Plant Physiol.* 140: 558–579.

Yuasa, T. and Muto, S. (1992) Ca^{2+}-dependent protein kinase from the halotolerant green alga Dunaliella tertiolecta: partial purification and Ca^{2+} dependent association of the enzyme to the microsomes. *Arch. Biochem. Biophys.* 296: 175–182.

Zhang, N., Zeng, L., Shan, H. and Ma, H. (2012) Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. *New Phytol.* 195: 923–937.

Zhang, X., Zong, J., Liu, J., Yin, J. and Zhang, D. (2010) Genome-wide analysis of WOX gene family in rice, sorghum, maize, arabidopsis and poplar. *J. Integr. Plant Biol.* 52: 1016–1026.

Zhang, X.S. and Choi, J.H. (2001) Molecular evolution of calmodulin-like domain protein kinases (CDPKs) in plants and protists. *J. Mol. Evol.* 53: 214–224.

Zhou, L., Fu, Y. and Yang, Z. (2009) A genome-wide functional characterization of Arabidopsis regulatory calcium sensors in pollen tubes. *J. Integr. Plant Biol.* 51: 751–761.

Zimmer, A., Lang, D., Richardt, S., Frank, W., Reski, R. and Rensing, S.A. (2007) Dating the early evolution of plants: detection and molecular clock analyses of orthologs. *Mol. Genet. Genomics* 278: 393–402.

Zuo, R., Hu, R., Chai, G., Xu, M., Qi, G., Kong, Y. and Zhou, G. (2013) Genome-wide identification, classification, and expression analysis of CDPK and its closely related gene families in poplar (Populus trichocarpa). *Mol. Biol. Rep.* 40: 2645–2662.