Shanika Amarasinghe, Nathan S. Watson-Haigh, Matthew Gilliham, Stuart Roy, Ute Baumann

The evolutionary origin of CIPK16: a gene involved in enhanced salt tolerance

Molecular Phylogenetics and Evolution, 2016; 100:135-147

Published at: http://dx.doi.org/10.1016/j.ympev.2016.03.031

The Evolutionary Origin of CIPK16: A Gene Involved in Enhanced Salt Tolerance

Shanika Amarasinghe[1], Nathan S. Watson-Haigh[1], Matthew Gilliham[2], Stuart Roy[1], Ute Baumann[1]

1. Australian Centre for Plant Functional Genomics PMB 1 Glen Osmond SA 5064.

2. Australian Research Council Centre of Excellence in Plant Energy Biology, Department of Plant Science, School of Agriculture, Food and Wine, University of Adelaide, Glen Osmond SA 5064.

Correspondence:

Dr. Ute Baumann,
Australian Centre for Plant Functional Genomics
The University of Adelaide,
School of Agriculture, Food and Wine,
Glen Osmond, South Australia,
5064, AUSTRALIA.

Abstract

Calcineurin B-like protein interacting protein kinases (CIPKs) are key regulators of pre-transcriptional and post-translational responses to abiotic stress. *Arabidopsis thaliana* CIPK16 (*AtCIPK16*) was identified from a forward genetic screen as a gene that mediates lower shoot salt accumulation and improved salinity tolerance in Arabidopsis and transgenic barley. Here, we aimed to gain an understanding of the evolution of *AtCIPK16,* and orthologues of *CIPK16* in other plant species including barley, by conducting a phylogenetic analysis of terrestrial plant species. The resulting protein sequence based phylogenetic trees revealed a single clade that included AtCIPK16 along with two segmentally duplicated CIPKs, AtCIPK5 and AtCIPK25. No monocots had proteins that fell into this clade; instead the most closely related monocot proteins formed a group basal to the entire *CIPK16*, *5* and *25* clade. We also found that *AtCIPK16* contains a core *Brassicales* specific indel and a putative nuclear localisation signal, which are synapomorphic characters of *CIPK16* genes. In addition, we present a model that proposes the evolution of *CIPK16*, *5* and *25* clade.

Keywords (5-8)

CIPK16; *Arabidopsis thaliana*; barley; salt tolerance; phylogenetic analysis

1. Introduction

Salinity in soil impacts negatively on crop growth and is a significant limiting factor for agriculture, particularly in arid and semi-arid regions, with an estimated cost of US$27 billion due to lost crop production per year (Munns and Gilliham, 2015; Qadir et al., 2014). It has been estimated that on land irrigated for agriculture, which produces 40% of the world's calories, one-fifth of soils are salt-affected (FAOSTAT, 2014). The extent of this salt-affected irrigated agricultural land has been forecasted to increase by 4% every year (FAOSTAT, 2014; Pimentel et al., 2004). Crops with increased tolerance to salt, which provide higher yields under saline soil conditions, are needed to sustain future global food production (Munns and Gilliham, 2015). To this end, aspects of plant responses to salinity have to be understood before they can be manipulated by molecular assisted breeding or transgenesis (Roy et al., 2014).

Plants retain only around 5% of the water that they take up in transpiration, thus salt concentration in the transpiration stream needs to be in the order of $1/20^{th}$ of that in the soil to avoid the accumulation of salt in leaves to concentrations above that in the soil (Munns, 2005). As a result, all plants have developed mechanisms to exclude salt to a large degree: halophytes exclude ~92-95% of the salt in the soil solution and most crop plants exclude 96-99% (Munns, 2005; Munns and Tester, 2008). Plants achieve this by either minimizing the entry of salt into the leaves (i.e. the trait of shoot ion exclusion) or by tolerating the accumulation of salt in leaves by reducing the concentration of salt in the cytoplasm (i.e. the trait of tissue tolerance) by compartmentalizing of the salt in the cells of leaf sheath or leaf cell vacuole (Munns, 2005; Munns and Gilliham, 2015; Munns and Tester, 2008; Plett and Møller, 2010; Shabala, 2009). Both ion exclusion and tissue tolerance demand high amounts of energy for osmotic adjustment within the cytosol via organic solutes (Adem et al., 2014; Munns and Gilliham, 2015; Plett and Møller, 2010; Shabala, 2013). Wheat and rice have lower Na^+ and Cl^- concentrations in their leaves than the external solution as a consequence of ion exclusion mechanisms (Roy et al., 2014). Salt tolerant non-halophytes such as

barley, exclude less salt from leaves more clearly exhibit the trait of tissue tolerance (Colmer et al., 2005; Munns and Gilliham, 2015; Shabala, 2013).

In instances when plants are unable to eliminate the negative effects of salinity, they initially suffer due to the buildup of osmotic stress followed by salt-specific ionic stress (Munns, 2005; Munns and Tester, 2008; Rajendran et al., 2009). Immediately after exposure to salinity but before ions accumulate in the plant's shoot, plant growth rate is reduced (Rajendran et al., 2009; Tavakkoli et al., 2010). Over time as ions accumulate in the shoot, ion toxicity reduced plant growth rate even further. Reducing the severity of salinity stress in plants therefore, needs early detection of salinity stress and the activation of stress signaling mechanisms.

Many aspects of stress signaling are facilitated through secondary messengers such as calcium ions (Ca^{2+}) (Batistic et al., 2011). A 20 to 60 second long single or biphasic Ca^{2+} elevation in the cytosol is one of the initial cellular responses of a plant to high salinity (Choi et al., 2014; Tracy et al., 2008). The sensor molecules capturing these signals fall into four major categories, namely Calcineurin B-Like (CBL) proteins, calcium-dependent protein kinases (CDPKs), calmodulins (CaMs), and calmodulin-like proteins (CAMLs) (Weinl and Kudla, 2009). Among these, CBLs selectively interact with one or more protein kinases from the group named CBL interacting protein kinases (CIPKs) (Batistic and Kudla, 2004; Kim et al., 2000). The CIPKs have been catalogued as SNF1 (Sucrose non-fermenting 1)-related kinases and group 3 (SnRK3) proteins, according to their structural features and evolutionary associations (Hrabak et al., 2003). The general structure of all CIPK-type kinases includes a conserved N-terminal kinase domain, and a variable junction domain, which separates it from a unique C-terminal regulatory domain (Sanchez-Barrena et al., 2007; Weinl and Kudla, 2009). In common with many kinases, an activation loop lies between two conserved tri-peptide motifs (DFG…APE) in the kinase domain, which needs to be phosphorylated for the kinase to be activated (Nolen et al., 2004). While much of the regulatory domain sequence is divergent in these proteins, there exists a well conserved FISL/NAF domain, which mediates

the interaction with CBLs (Albrecht et al., 2001). Additionally, a conserved C-terminal protein–phosphatase interaction (PPi) domain mediates CIPK interaction with the 2C-type protein phosphatase (PP2C) group, via phosphorylation (Ohta et al., 2003; Sanchez-Barrena et al., 2013).

A *CIPK* from the model dicot *Arabidopsis thaliana*, named *AtCIPK16* is associated with Na^+ exclusion in plants (Roy et al., 2013). Transgenic Arabidopsis constitutively overexpressing *AtCIPK16* showed a significant reduction of shoot Na^+ in plants grown in elevated salt in both soil and hydroponics (Roy et al., 2013). Moreover, transgenic barley constitutively expressing *AtCIPK16* also exhibited decreased leaf Na^+ and increased salinity tolerance. This implies that *AtCIPK16* can be used as a tool in genetic engineering to improve salinity tolerance in crops. In Roy et al. (2013), *AtCIPK16* was identified using a Bay-0 x Shahdara mapping population. The Bay-0 accession allele of *AtCIPK16* contained a TATA box 65 bp upstream of the start codon and had higher gene expression under salt stress compared to Shahdara (which contained no TATA box in this region) (Roy et al., 2013). However, our understanding of the underlying mechanism of *CIPK16* mediated salt tolerance is still incomplete.

A study on how widespread the *CIPK16*-associated salinity tolerance mechanism is in the plant world could be an initial step in understanding the functional network associated with *CIPK16*. It also may lay the foundation for further experiments such as screening or editing the genes that boost salt tolerance in plants. A phylogenetic study on the prevalence of *CIPK16* in the plant kingdom would facilitate the discovery of *AtCIPK16* orthologues in crops important for global food production. Orthologues, by definition, would have a common ancestor and tend to have similar functionality (Fulton et al., 2006; Wu et al., 2006). Several phylogenetic analyses on *CIPK*s from different plant species have been conducted (Thoday-Kennedy et al., 2015 and references therein). For instance, a previous phylogenetic study on 25 *A. thaliana* CIPKs revealed that *AtCIPK16* resides in close proximity to two other segmentally duplicated *CIPK*s, namely *AtCIPK5* and *AtCIPK25* (Kolukisaoglu et al., 2004). However, to our

knowledge, a phylogenetic analysis detailing the prevalence of *CIPK16* across the plant kingdom has not been conducted so far.

The aim of the current study is to discover the origin of *CIPK16* and its closest relatives, *CIPK5* and *CIPK25*, using phylogenetic approaches, *in silico* protein analysis and known evolutionary relationships between terrestrial plants (**Error! Reference source not found.**) (Chase, 2004; S. Kagale et al., 2014; Soltis et al., 2011).

2. Materials and Methods

2.1. Molecular Phylogenetics of *CIPK16*

2.1.1. Sequence Retrieval

Protein and nucleotide sequences were retrieved from the sources detailed in Table 1. Species were selected based on the availability of the full genomic sequences (Cheng et al., 2013; Michael and Jackson, 2013). *Brassicaceae* species were targeted as they are closely related to *A. thaliana*. Sequences were retrieved through one of the following methods: 1) sequence similarity to *A. thaliana* CIPK gene/protein sequences; and, 2) keyword searches. Sequences retrieved by sequence similarity were performed either via the BLAST tool linked to online databases or by locally indexed databases using the NCBI BLAST+ tool (V 2.2.29). All blastn, blastp, tblastn, tblastx options were used in BLAST querying with an expectation value (e-value) ≤ $1x10^{-5}$. Default settings were used for querying with complete sequences. Settings were changed for queries with partial protein sequence in order to increase sensitivity; the short query option was deselected, the expect threshold was changed to 5 million, word size was changed to 2, and the compositional adjustments setting was set to “no adjustments”. Sequence retrieval by keyword searches used the terms “cipk”, “cbl interacting protein kinase”, “cbl interacting” and “calcineurin b like”. A fasta formatted sequence file for all the sequences used in this study is in the supplementary materials (S1).

2.1.2. Sequence Alignment

Protein multiple sequence alignments (MSAs) were generated using MUSCLE (default settings) implemented in Jalview (Edgar, 2004; Waterhouse et al., 2009). Manual alignment was carried out to improve the MSAs. Duplicates (defined by 99% identity or above) were removed from the multiple sequence alignments via the remove duplicates option in Jalview. Nucleotide alignments corresponding to the protein MSA were generated using Dialign 2.0 implemented in the RevTrans2.0 server, in order to correctly align DNA codons with corresponding amino acid residues (Morgenstern, 1999; Wernersson and Pedersen, 2003). Sequences were validated to be functional CIPK sequences by screening for the DF(G/D)L, APE motifs in the N-terminal kinase domain and the (N/T)AF motif in the C-terminal regulatory domain via a custom Perl script. Partial sequences and sequences without any of the DF(G/D)L, APE and (N/T)AF motifs were therefore excluded from the final refined alignment files. However, these sequences were not discarded but separated and manually examined.

Additionally, SeqFIRE and GBlock tools were employed to identify the conserved regions of the alignments, which encompass important domains (Ajawatanawong et al., 2012; Talavera and Castresana, 2007). The “Conserved Block Module – single alignment mode” from SeqFire accepted protein MSAs in FASTA format. The default parameter settings were used for the SeqFire tool. The online Gblock server was used to extract the well-aligned and conserved sequence blocks from the MSAs. FASTA formatted protein sequences were input with all the options for “less stringent selection” enabled.

2.1.3. Phylogenetic Tree Computation

The refined alignment files were used for this step. The phylogenetic analysis was conducted in such a way that, initially, bi-species trees were created using 26 CIPKs of *A. thaliana* as the reference. After examining the 137 trees developed this way, sequences were sequentially joined with the 26 *A. thaliana* sequences to generate the final tree including all 47 terrestrial species used in our analyses (the known evolutionary relationships among these species are shown in **Error! Reference source not found.**).

MODELGENERATOR v. 0.85 was used to determine the best substitution model for each dataset (with and without outgroups) (Keane et al., 2006). We hypothesised that unknown substitution rate variations exist in the genes of our data sets. Therefore, we used the gamma distribution for modelling the rate variation (5 categories) (Yang, 1996; Yang and Rannala, 2012). The best model fit for the phylogenetic tree creation was based on Corrected Akaike Information Criterion (AICc), Akaike Information Criterion 2 (AIC2) and Bayesian Information Criterion (BIC) (S2).

Phylogenetic trees were generated using MEGA 6.06 software using a Maximum Likelihood approach (Mount, 2008; Tamura et al., 2013). To estimate how well the nodes of the ML tree were supported, 10,000 bootstrap trees were generated (Felsenstein, 1985). The DOLLOP program from the PHYLIP package implemented in T-REX (http://www.trex.uqam.ca/index.php?action=phylip&app=dollop) was used to determine the minimum gene set for ancestral nodes of the phylogenetic tree (Boc et al., 2012; Felsenstein, 1996). The generated parsimony tree (Newick format) was used as the input to Ancestor v 1.1 in order to predict the ancestral sequences (Diallo et al., 2010). These ancestral sequences were used as queries for further BLAST searches.

2.2. Identification of Unique Sequence Features

The Prosite (http://prosite.expasy.org/) and Pfam (http://pfam.xfam.org/) web resources were used to extract known important residues, motifs and domains of AtCIPK16 and its homologues (Finn et al., 2014; Sigrist et al., 2013). CIPK homologous sequences were examined using ScanProsite available through Prosite (v. 20.124) with the option “high sensitivity”. We queried the Pfam database (v.27.0) using protein sequences with the default e-value threshold of $1x10^{-6}$ (Finn et al., 2014).

To identify potential nuclear localisation signals (NLS) within AtCIPK16 and its homologues, we submitted protein sequences to cNLS Mapper (http://nls-mapper.iab.keio.ac.jp/) in FASTA format (Kosugi et al., 2009). The following parameters were used; a cut-off score of 2.0; long bipartite NLSs were searched in the entire region of the proteins. Structural (e.g. secondary structure) and biochemical (e.g. solvent

accessibility, subcellular localisation) features were predicted using PredictProtein and NetSurfP (Petersen et al., 2009; Rost et al., 2004). Default parameters were used.

2.3. Intron-Exon Architecture Analysis of CIPK16 Orthologues

To visualise and compare the intron-exon structure of CIPK16 orthologues we used GSDraw, available in PIECE (http://wheat.pw.usda.gov/piece/) (Wang et al., 2013). The input files contained the genomic nucleotide sequences and the cDNA sequences (S3). PIECE is a comparative genomics database named for Plant Intron and Exon Comparison and Evolution studies.

2.4. *AtCIPK16* Diversity Among *A. thaliana* Accessions

VCF files and BAM files were obtained from http://1001genomes.org for the purpose of identifying SNPs within the vicinity of AtCIPK16 (at2g25090.1) +/- 2500bp (up to but not including any neighbouring genes). For the identification of SNPs, we used VCF files for 696 accessions made available under the GMINordborg2011, MPICWang2013 and Salk projects. SNPeff was used to predict the effect (e.g. synonymous, non-synonymous, start codon gain/loss, stop codon gain/loss and frameshifts) of the identified SNPs (Cingolani et al., 2012).

Roy et al., 2013 have previously reported a 10bp deletion in the promoter region of *AtCIPK16* in Bay-0. For this reason, we examined accessions for which BAM files were available since VCF files are typically generated by SNP identification pipelines that ignore indel information. Furthermore, we restricted the selection of BAM files to those accessions for which reads had been mapped to the reference Col-0 (i.e. Shahdara, Bay-0, Sakata, Ri-0, Oy-0, Jea, blh-1 and Alc-0 under the JGIHazelWood 2008/11 projects). BAM files were visualised in IGV and alignments padded with gaps to reduce mismatches and achieve perfect gapped alignments.

3. Results

3.1. Molecular Phylogenetics of CIPK 16, 5 and 25 Protein Sequences

Computation of phylogenetic trees allowed us to predict evolutionary relationships between genes. In the first instance, we computed phylogenetic trees for CIPK families from different *Brassicaceae* and *Cleome* species; the evolutionary relationships between these species is shown in Figure 1. These species include *C. sativa*, *C. rubella*, *A. alpina*, *B. stricta*, *B. oleraceae*, *E. salsugineum*, *S. parvula*, *L. alabamica*, *A. arabicum* and *T. hassleriana*. The individual unrooted protein sequence derived phylogenetic trees for these species are provided in supplementary materials (S4-S13). The summary of the gene/protein tree for all studied *Brassicales* species is shown in **Error! Reference source not found.** and the fully expanded tree is in S14.

The number of representatives for the CIPK16, 5 and 25 clade ([CIPK16/5/25]) varies among *Brassicales* species (**Error! Reference source not found.**). We were able to identify a complete or a partial sequence in all core *Brassicales* (*Brassicaceae* and *Cleomaceae*), which clustered with AtCIPK16 (Figure 1), with the exception of *L. alabamica*, Within the *Brassicaceae* we were also able to identify orthologues for both AtCIPK5 and AtCIPK25. However, the only homologous sequence we identified in *T. hassleriana*, (a single representative of *Cleomaceae*) was placed at the base of the CIPK5/CIPK25 clade. For dicot species outside the core *Brassicales* (*C. papaya, T. cacao and G. raimondii*), and monocots, we found only homologues which form groups basal to [CIPK16/5/25] (Figure 2 and Figure 3). We could not identify any AtCIPK16, 5 or 25 orthologues in "non-core *Brassicales*" dicots (NCBs) (Figure 3). A fully expanded tree for Figure 3 is available as a supplementary figure (S15). The basal angiosperm *Amborella trichopoda* is the most distant species to *A. thaliana* that possesses a gene that clusters in the basal group for [CIPK16/5/25] (Figure 3). We were unable to identify close homologues to [CIPK16/5/25] in terrestrial plant species outside of angiosperms (data not shown).

3.2. Unique Characteristics of CIPK16s

Comparing MSAs and the computed phylogenetic trees revealed unique regions of CIPK16 orthologues. One such significant character is a unique indel (MMPEGLGGRRG) that exists in the activation loop of the kinase domain of CIPK16 orthologues (ALI) (**Error! Reference source not found.**). ALI-CIPK16 was not present in any other gene we studied. Additionally, it was not present in any sequence in any of the sequence databases we used for our study (Table 1). ALI lies between the conserved regions of the activation loop. A fragment 100% identical to ALI was present in the manually curated database of *B. oleraceae* scaffolds (Scaffold000171 FRAGMENT 1092155:1092254). This sequence was only partial and did not contain the C terminal NAF domain and the PPi domain (S16).

Another distinguishing feature is a putative nuclear localization signal in the junction domain of CIPK16 orthologues (JDNLS). According to cNLS server predictions, AtCIPK16 has monopartite and bipartite nuclear localization signals (NLS) with the sequence spanning from 300 to 308 ($PPTKKKKKD_{308}$) (**Error! Reference source not found.**). A score of 6.5 assigned by the server for this signal suggests that AtCIPK16 can be partially located to the nucleus. Proteins from other CIPK clades did not possess an NLS in the junction domain (**Error! Reference source not found.**). However, all CIPKs possessed a bipartite signal (a score equal to or less than 5.5) with a tendency to be directed to the cytoplasm.

3.3. Intron-Exon Architecture of *CIPK16* Orthologues

The intron-exon study conducted on *AtCIPK16* orthologues from members of the *Brassicaceae* and *Cleomaceae* shows that they all possess two exons separated by an intron (**Error! Reference source not found.**). Exon 1 length varies among species from 692 to 709 nucleotides, whereas Exon 2 length varies from 685 to 742 nucleotides. The indels in exon 1 and exon 2 were analysed separately by a multiple sequence alignment of the DNA sequences (S16). We see the presence of many transitions and transversions compared to the consensus sequence within both exons of the analysed species (S16). The intron lengths of the *AtCIPK16* orthologues are quite variable (Figure 7). They vary from 350bp in *T. hassleriana* to 2048 bp in *A. lyrata*. All introns except the ones from *B.napus*, *S.parvula*, and *C.sativa* are

phase 2 introns (i.e. they interrupt the reading frame of a gene by inserting a sequence between the second and third nucleotide of a codon). *B.napus*, *S.parvula*, and *C.sativa* contain phase 1 introns (i.e. they interrupt the reading frame of a gene by inserting a sequence between the first and second nucleotide of a codon). *AtCIPK5* and *AtCIPK25* orthologues are intron-less and therefore are not shown.

3.4. *AtCIPK16* Diversity Among *A. thaliana* Accessions

From the analysis of VCF files from 696 *A. thaliana* accessions, we identified 359 positions harboring SNPs within the vicinity of *AtCIPK16*. Of these, 195 (54.3%) were upstream, 4 (1.1%) in the 5'-UTR, 17 (4.7%) in the CDS of exon 1, 59 (16.4%) in the intron, 22 (6.1%) in the CDS of exon 2, 10 (2.8%) in the 3'-UTR and 52 (14.5%). Twenty-two of the 39 SNPs that fell within the coding region are silent (synonymous) while 17 cause a change in an amino acid (non-synonymous) (S17).

Of the 8 accessions for which we had access to BAM files, we identified two (Bay-0 and blh-1) which contained a 10bp deletion within the promoter region (65 bases upstream of the ATG) of *AtCIPK16* compared to the Col-0 reference (**Error! Reference source not found.**). This deletion has previously been reported only in Bay-0 (Roy et al., 2013), and results in the creation of a TATA box.

4. Discussion

AtCIPK16 promotes sodium exclusion and salt tolerance (Roy et al., 2013). Understanding the pervasiveness of *CIPK16* in the plant kingdom would lay the foundation to better understanding its mode of action in plants. Already identified CIPKs from *A. thaliana*, predicted ancestral versions of the AtCIPKs and keywords were used to mine for CIPK sequences from the plant sequence databases. We carried out a molecular phylogenetic analysis of the multigene CIPK family in terrestrial plants to investigate potential processes in evolution that may have given rise to the modern day CIPK proteins (Soltis and Soltis, 2003). Additional *in-silico* protein analysis approaches were used to identify unique regions in primary protein structures, intron-exon architecture and variation within the sequences of *AtCIPK16* in different accessions of Arabidopsis to strengthen the phylogenetic inferences.

In order to generate the phylogenetic trees, we gathered protein sequences from all fully sequenced species to minimise the impact of missing data and evolutionary pressure on domain identification in AtCIPK16 orthologues and misinterpretation of the analysis (Haudry et al., 2013; Kagale et al., 2014).

4.1. Synapomorphic Characters Define core *Brassicales* restricted *CIPK16*s

Comparison of the phylogenetic data and MSAs show that the CIPK protein sequences and nucleotide sequences of *Brassicaceae* CIPK16 orthologues have a highly conserved synapomorphic character (**Error! Reference source not found.**, **Error! Reference source not found.**). Indel ALI is one of these, although this sequence lacks one amino acid in the *Cleomaceae* species *T. hassleriana* and is slightly dissimilar to those of the *Brassicaceae* species (**Error! Reference source not found.**). It is noteworthy that this unique insertion was not found in any other dicot or monocot species that has been fully sequenced. Therefore, we hypothesise that ALI can be used as a unique of CIPK16 orthologues within the *Brassicales*. The partial sequence we discovered in *B. oleraceae* supports this hypothesis, although given that ambiguity of partial sequences tends to introduce false relationships among species in a phylogenetic analysis, the *B. oleraceae* sequence was excluded when generating phylogenetic trees.

The other important highly conserved character noted was the junction domain nuclear localisation signal (JDNLS) (**Error! Reference source not found.**). It is present in all CIPK16 orthologues from *Brassicaceae* except that of the basal species *A. arabicum*. This raises the question of its functional importance for the localization of a CIPK16 orthologue in the cell. However, this requires further experimental validation.

It is clear from our study that *CIPK16* is a lineage-specific gene for core *Brassicales*. The consistency in intron-exon studies supports the CIPK16 orthologues (**Error! Reference source not found.**). The most parsimonious explanation for *CIPK16s* to be core *Brassicales* specific is that *CIPK16* arose as a result of a gene duplication event after the speciation of this group of plants. Genes that are duplicated can evolve through the acquisition of new or specialised functions at the expression or protein level

(neofunctionalization), the retention of ancestral functionality or to escape from adaptive conflict (EAC) (Blanc and Wolfe, 2004; Des Marais and Rausher, 2008; Moghe et al., 2014). The identification of non-synonymous SNPs (S17) among the 696 accessions we analysed warrant closer examination to ascertain whether they are associated with higher or lower tolerance to salt. Interestingly, we see that Bay-0 and blh-1 accessions share a common TATA box positioned 65 bp upstream of *CIPK16* (**Error! Reference source not found.**). This is important as Bay-0 has shown higher *CIPK16* gene expression in response to salt stress compared to Shahdara in a previous study (Roy et al., 2013). Whether the presence of the TATA box confers a similar increase in CIPK16 expression in blh-1 needs to be experimentally determined.

Prior research on *Brassicaceae* gene evolution revealed that the majority of lineage-specific genes from *A. thaliana* are stress responsive (Donoghue et al., 2011). AtCIPK16 has been shown to interact with *shaker-type* K^+ channels in *A. thaliana* (AKT1), which keeps the cellular Na^+/K^+ ratio low under low K^+ stress and confers salt tolerance when overexpressed (Lee et al., 2007; Roy et al., 2013). Apart from AKT1, AtCIPK16 has shown interactions with CBL1, CBL9 and protein phosphatase 2C type proteins (Lan et al., 2011; Lee et al., 2007). Moreover, there is experimental evidence showing AtCIPK5, one of AtCIPK16's closest relatives, interacts with CBL1, CBL3, CBL4 and CBL9 (Kim et al., 2000; Kolukisaoglu et al., 2004; Schlücking et al., 2013). However, very little, if anything is known about the functionality of AtCIPK16s other closest relative AtCIPK25. There is evidence on [CIPK16/5/25] homologues' from species such as Chickpea and rice being responsive to plant abiotic stress (Meena et al., 2015; Yoon et al., 2009). Nevertheless, we believe that CIPK16, 5 and 25 and identified homologues of [CIPK16/5/25] should be further pursued to analyse their function in order to help us understand the drivers of *CIPK16* evolution.

4.2. The Evolution of *CIPK16, 5* and *25* Clade

From our analysis, we are able to propose an evolutionary model for *CIPK16* (**Error! Reference source not found.**). We considered the whole *CIPK 16, 5* and *25* clade in explaining the evolution of *CIPK16* as well as sister taxa (Kolukisaoglu et al., 2004). It has been shown that a recent paleopolyploidization event (At-α) took place, which was restricted to *Brassicaceae* (Barker et al., 2010; Schranz and Mitchell-Olds, 2006). To support this fact, our study shows that segmental duplication (SD) of intron-less CIPKs in *Brassicaceae* are confined to that group. This includes the SD, which gave rise to *CIPK5* and *CIPK25*. We could not find evidence that *Cleomaceae* experienced an independent genome duplication (Cs-α) as suggested previously (Schranz and Mitchell-Olds, 2006). However, our results indicate that *CIPK16* existed before the speciation of *Cleomaceae* and therefore before Cs-α. It can be assumed that the WGD event that took place 124.6 ± 2.57 Mya (At-β) gave rise to the ancestral version of the *CIPK16, 5* and *25* clade from a single ancestral state (**Error! Reference source not found.**). This is consistent with previous work, which states that the paleopolyploidization event At-β is shared between *Brassicaceae* and *Cleomaceae* (Barker et al., 2010). According to this hypothesis, and supported by our study, the ancestral version of *CIPK16* and *CIPK5* and *25* therefore had to evolve after the rise of non-core *Brassicales* species. This agrees with the previous work which showed species of *Carica* do not share At-β (Barker et al., 2010; S. Kagale et al., 2014) (**Error! Reference source not found.**).

The fact that NCBs and monocots have no CIPK16, CIPK5 or CIPK25 orthologues suggests they must possess an ancestral version of [CIPK16/5/25] or the gene itself has been made redundant by evolution (Pérez-Pérez et al., 2009). The most basal species in our phylogenetic analysis to contain a sequence that clusters with [CIPK16/5/25] is the angiosperm *A. trichopoda.* This suggests that the earliest ancestor of [CIPK16/5/25] evolved after the diversification of angiosperms.

4.3. Continuing *CIPK16* Research for Salinity Stress

It is clear from our phylogenetic analysis that AtCIPK16 does not have a clear orthologue in important crops such as barley or wheat. However, our finding that the last common ancestor of [CIPK16/5/25]

gave rise to *CIPK16* after the divergence of dicots and monocots (more specifically after the diversification of core *Brassicales*, a subgroup of dicots), and the previous finding that overexpressing *AtCIPK16* confers salt tolerance in the monocot barley, poses further questions (Roy et al., 2013). Do the conserved elements ALI and the JDNLS have functional importance in CIPK16s? Would it be possible that the functionality of CIPK16, 5 and 25 result from functional partitioning of the ancestral genes due to selective pressure? If so, are the functionalities of CIPK16, 5 and 25 still retained in seemingly ancestral versions we see in NCBs and monocots? This study therefore, highlights the necessity to explore the functionality of *AtCIPK16* in *A. thaliana* and cereals such as barley.

Glossary

CIPK: Calcineurin B-Like Protein Interacting Protein Kinase

AtCIPK16: *Arabidopsis thaliana* CIPK16

AtCIPK5: *Arabidopsis thaliana* CIPK5

AtCIPK25: *Arabidopsis thaliana* CIPK25

Indel: Insertion/Deletion

Ca^{2+}: Calcium ion

CBL: Calcineurin B-Like

CDPK: Calcium-dependent protein kinase

CaM: Calmodulin

CAML: Calmodulin like

SNF1: Sucrose non-fermenting 1

SnRK3: SNF1-related kinases group 3

FISL/NAF: NAF domain

PPi: Protein–phosphatase interaction domain

PP2C: 2C-type protein phosphatase

MSA: Multiple sequence alignment

AIC: Akaike Information Criterion

AIC2: Akaike Information Criterion 2

BIC: Bayesian Information Criterion

ML: Maximum Likelihood

NLS: Nuclear localisation signal

cDNA: complimentary DNA

SD: Segmental duplication

Abbreviations

[CIPK16/5/25]: CIPK16, 5 and 25 clade

NCBs: none core *Brassicales* dicots

ALI: Activation loop indel in CIPK16

JDNLS: Junction domain nuclear localisation signal in CIPK16

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Acknowledgement

Authors would like to thank the Australian Centre for Plant Functional Genomics for funding the research of this project. Authors would also like to thank Mr. Ashan Hettiarachchige for generating the final high-resolution images according to the manuscript guidelines.

Author and Contributors

Study concepts/study design or data acquisition or data analysis/interpretation: all authors; manuscript drafting or manuscript revision for important intellectual content: all authors; manuscript final version approval: all authors; literature research: SA; phylogenetic and in-silico analysis: SA and NWH; and manuscript editing: all authors. Agreement to be accountable for all aspects of the work: all authors.

Tables

Table 1

Sequence acquired species	Web Resource	References	Sequence Access method	Identification method
Arabidopsis thaliana	TAIR (http://www.arabidopsis.org/)	(Lamesch et al., 2012)	Online from TAIR10	BLAST (blastn, blastp), keyword search
Arabis alpina ***Boechera stricta***	NCBI (www.ncbi.nlm.nih.gov/)	(Coordinators, 2013)	Online	BLAST (blastn, blastp, tblastn, tblastx)
Leavensworthia alabamica	CoGe (https://genomevolution.org/CoGe/)	(Lyons and Freeling, 2008)	FTP download	BLAST (blastn, blastp, tblastn, tblastx)
Aethionema arabicum	CoGe (https://genomevolution.org/CoGe/)	(Haudry et al., 2013)	FTP download	BLAST (blastn, blastp, tblastn, tblastx)
A. lyrata	Phytozome (www.phytozome.net/)	(Nordberg et al., 2014)	Online	BLAST (blastn, blastp, tblastn, tblastx)
Raphanus sativus	Raphanus sativus Genome Database (http://radish.kazusa.or.jp/)	(Kitashiba et al., 2014)	Online	BLAST (blastn, blastp, tblastn)
Capsella rubella	Phytozome (www.phytozome.net/)	(Goodstein et al., 2012)	Online	BLAST (blastn, blastp, tblastn, tblastx)

Schrenkiella parvula	thellungiella.org (http://thellungiella.org/)	(Dassanayake et al., 2011)	FTP download	BLAST (blastn, blastp, tblastn, tblastx)
Eutrema salsugineum	thellungiella.org (http://thellungiella.org/)	(Yang et al., 2013)	Online	BLAST (blastn, blastp, tblastn, tblastx)
Sisymbrium irio	Brassicadb (http://brassicadb.org/brad/)	(Haudry et al., 2013)	FTP download	BLAST (blastn, blastp, tblastn, tblastx)
Brassica rapa (v 1.5) ***Brassica napus (v 1.0)***	Brassicadb (http://brassicadb.org/brad/)	(Cheng et al., 2011)	Online	BLAST (blastn, blastp, tblastn, tblastx)
Brassica oleraceae (v 1.0)	Brassicadb (http://brassicadb.org/brad/)	(Cheng et al., 2011)	FTP download	BLAST (blastn, blastp, tblastn, tblastx)
Camelina sativa	camelinadb (http://www.camelinadb.ca/)	(Sateesh Kagale et al., 2014)	FTP download	BLAST (blastn, blastp, tblastn, tblastx)
Tarenaya hassleriana	CoGe (https://genomevolution.org/CoGe/)	(Cheng et al., 2013)	Online	BLAST (blastn, blastp, tblastx)
Carica papaya	NCBI (www.ncbi.nlm.nih.gov/)	(Coordinators, 2013)	Online	BLAST (blastn, blastp, tblastn, tblastx)
Theobroma cacao	Cacao Genome Database (www.cacaogenomedb.org/)	(Argout et al., 2011)	FTP download	BLAST (blastn, blastp, tblastn, tblastx)
Gossypium raimondii	Phytozome (www.phytozome.net/)	(Goodstein et al., 2012)	Online	BLAST (blastn, blastp, tblastn, tblastx)

Vitis vinifera	Genoscope (www.genoscope.cns.fr/externe/GenomeBrowser/Vitis/)	(Jaillon et al., 2007)	Online	BLAST (blastn, blastp, tblastn, tblastx)
Musa acuminate malaccensis	Banana Genome Hub (banana-genome.cirad.fr/)	(Droc et al., 2013)	Online	BLAST (blastn,blastp, tblastn, tblastx)
Fragaria vesca ***Malus x domestica*** ***Prunus persica*** ***Pyrus communis***	Genome Database for Rosaceae (www.rosaceae.org/)	(Jung et al., 2014)	Online	BLAST (blastn,blastp, tblastn, tblastx)
Brachypodium distachyon	Brachypodium database moved to Phytozome	(Goodstein et al., 2012)	Online	BLAST (blastn,blastp, tblastn, tblastx)
Oryza sativa	Rice Genome Annotation Project (http://rice.plantbiology.msu.edu/)	(Ouyang et al., 2007)	Online	BLAST (blastn,blastp, tblastn, tblastx)
Triticum aestivum	IWGSC (www.wheatgenome.org/)	(International Wheat Genome Sequencing Consortium (IWGSC), 2014)	Online	BLAST (blastn,blastp, tblastn, tblastx)
Hordeum vulgare	BARLEX from IPK (www.ipk-gatersleben.de/en/)	(Colmsee et al., 2015)	Online	BLAST (blastn,blastp, tblastn, tblastx)
Hordeum vulgare	MIPS (http://mips.helmholtz-muenchen.de/plant/barley/)	(Nussbaumer et al., 2013)	Online	BLAST (blastn,blastp, tblastn, tblastx)
Amborella trichopoda	http://www.amborella.org/	(Albert et al., 2013)	FTP download	BLAST (blastn,blastp, tblastn, tblastx)

Picea abies	The cogenie.org (http://congenie.org/)	(Nystedt et al., 2013)	Online	BLAST (blastn,blastp, tblastn, tblastx)
Generic	UniProt (www.uniprot.org/)	(Consortium, 2015)	Online	BLAST (blastn,blastp, tblastn, tblastx), keyword search
Generic	PlantGDB (www.plantgdb.org/)	(Duvick et al., 2008)	Online	BLAST (blastn,blastp, tblastn, tblastx)
Generic	EnsamblePlants (plants.ensembl.org/)	(Cunningham et al., 2015)	Online	BLAST (blastn,blastp, tblastn, tblastx)

Table Legends

Table 1 Species and resources used from which protein and nucleotide sequences were identified for this study

Figure Legends

Figure 1 A cladogram showing a summary of known relationships (Chase, 2004; S. Kagale et al., 2014; Soltis et al., 2011) among the species used in this study together with higher level taxonomic designations to which we commonly refer.

Figure 2 Molecular Phylogenetic analysis for Brassicales-Malvales CIPKs used in this study by Maximum Likelihood method (summarised view)

The evolutionary history was inferred by using the Maximum Likelihood method based on the Le_Gascuel_2008 model (Le and Gascuel, 2008). The tree with the highest log likelihood (-65950.5450) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbour-Joining method to a matrix of pairwise distances estimated using a JTT model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 1.0539)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 408 amino acid sequences. There were a total of 698 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura et al., 2013). Intron harbouring groups' branches are coloured olive colour. The Intron less groups' branches are coloured in dark purple. The CIPK16, 5 and 25 clade from the intron less group is expanded highlighted by taxon names of colour turquoise. The other CIPK nodes are collapsed down and named for clarity of presentation. The fully expanded tree is available as supplementary materials (S14). Sequences from KC310466.1, AAU87882.1, AAU87884.1, KC991147.1, AGO32663.1, KC991149.1 and AEX07321.2 were included in the analyses as the most closely related non-CIPK-type protein kinases.

Figure 3 Summary of the Molecular Phylogenetic analysis for CIPK16/5/25 group CIPKs used in this study by Maximum Likelihood method (summarised view)

The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model (Jones et al, 1992). The tree with the highest log likelihood (-32815.7688) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. A discrete Gamma distribution was used to model evolutionary rate differences among sites (5 categories (+G, parameter = 0.9058)). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 113 amino acid sequences. There were a total of 504 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura et al., 2013). The nodes are collapsed down and named for clarity of presentation. The CIPK16, 5 and 25 nodes are highlighted by the colour turquoise. Nodes representing homologues to [CIPK16/5/25] from dicots and monocots are coloured dark red and green respectively. The tree is rooted on other A. thaliana CIPKs. The fully expanded tree is available as supplementary materials (S15).

Figure 4 Multiple Sequence Alignment (MSA) of the Activation Loop Domain of CIPK Proteins

The alignment was developed from the complete sequences of CIPK proteins using MUSCLE algorithm incorporated in Jalview application. The MSA showing the indel (ALI) of proteins in CIPK16 clade (shown by a red box). Other A. thaliana CIPKs (AtCIPKs) were used to support the fact that ALI is only present in the CIPK16 clade proteins. ALI lies between the conserved regions of the activation loop (i.e. between DFGLSAL and SSDDLLHTRCGTPAYVAPE). For easy referencing in this text we would number the activation loop from D1FGLSAL…AYVAPE37.The colours represent the conservation of the alignment, green being highly conserved and blue being less. AtCIPK16 is underlined in dark blue. The conservation histogram and normalised consensus logo is shown beneath the MSA.

Figure 5 Multiple Sequence Alignment (MSA) of the Junction domain of CIPK Proteins

The alignment was developed from the complete sequences of CIPK proteins using MUSCLE algorithm incorporated in Jalview application. The Nuclear Localization Signal (NLS) unique to CIPK16 junction domain is shown in yellow (JDNLS). The NLS was predicted by the cNLS Mapper (http://nls-mapper.iab.keio.ac.jp/) (Kosugi et al., 2009). CIPK16 orthologues from A. arabicum and T. hassleriana did not predict a NLS in this region. Other A. thaliana CIPKs were used to show the variability within this region. JDNLS lies in the junction domain in middle of kinase domain and the regulatory domain. AtCIPK16 is underlined by a dark blue line. The consensus logo for the NLS in CIPK16s, conservation histogram and normalised consensus logo is shown beneath the MSA.

Figure 6 An IGV screenshot of the promoter region of AtCIPK16 previously shown to contain a deletion in Bay-0 (Roy et al. 2013).

Read alignments for both Bay-0 and blh-1 accessions are shown together with their corresponding coverage tracks and indicate both accessions contain a conserved 10bp deletion (position highlighted by the red bar) relative to the Col-0 reference (mismatching bases not shown). The read alignments of the original BAM files were modified slightly to pad the alignments of reads spanning the deletion. The drop in coverage seen in Bay-0 is indicative of a non-gapped alignment tool being used to generate the BAM file. The BAM file of blh-1 already contained gapped read alignments, indicating a gapped aligner was used to generate the BAM file. The effect of the deletion in these two accessions is the creation of a TATA box (indicated by the blue bars spanning the deletion below the sequence track).

Figure 7 Intron-exon analysis of CIPK16 clade proteins using PIECE web tool (Wang et al., 2013):

The species names are shown in a cladogram with the respective intron-exon architecture of the CIPK proteins. A base pair scale is shown on top. The Exons are shown as black bars and the introns are shown in light brown. The length of each exon and intron are shown above each region. The phase of

the intron is shown below each intron in light brown. We deliberately left the intron-exon structure for *L. alabamica* blank as we were unable to find a *CIPK16* orthologue for this species.

Figure 8 Proposed evolutionary model of CIPK 16, 5 and 25

The model of evolution of CIPK16, 5 and 25 from their last common ancestor in angiosperms. Each oval represents a gene. The rectangular box represents a group of plants with a common ancestor. The ovals with no fill colour represents inferred ancestral states. The coloured ovals represent the present day proteins from different groups of terrestrial plants mentioned below each group. Previously recorded evolutionary milestones are mentioned appropriately. A cladogram shows the known evolutionary relationships among the groups.

References

Adem, G.D., Roy, S.J., Zhou, M., Bowman, J.P., Shabala, S., 2014. Evaluating contribution of ionic, osmotic and oxidative stress components towards salinity tolerance in barley. BMC Plant Biol.14, 113. doi:10.1186/1471-2229-14-113

Ajawatanawong, P., Atkinson, G.C., Watson-Haigh, N.S., MacKenzie, B., Baldauf, S.L., 2012. SeqFIRE: a web application for automated extraction of indel regions and conserved blocks from protein multiple sequence alignments. Nucleic Acids Res. gks561. doi:10.1093/nar/gks561

Albert, V.A., Barbazuk, W.B., dePamphilis, C.W., Der, J.P., Leebens-Mack, J., Ma, H., Palmer, J.D., Rounsley, S., Sankoff, D., Schuster, S.C., Soltis, D.E., Soltis, P.S., Wessler, S.R., Wing, R.A., Albert, V.A., Ammiraju, J.S.S., Barbazuk, W.B., Chamala, S., Chanderbali, A.S., dePamphilis, C.W., Der, J.P., Determann, R., Leebens-Mack, J., Ma, H., Ralph, P., Rounsley, S., Schuster, S.C., Soltis, D.E., Soltis, P.S., Talag, J., Tomsho, L., Walts, B., Wanke, S., Wing, R.A., Albert, V.A., Barbazuk, W.B., Chamala, S., Chanderbali, A.S., Chang, T.-H., Determann, R., Lan, T.,

Soltis, D.E., Soltis, P.S., Arikit, S., Axtell, M.J., Ayyampalayam, S., Barbazuk, W.B., Burnette, J.M., Chamala, S., Paoli, E.D., dePamphilis, C.W., Der, J.P., Estill, J.C., Farrell, N.P., Harkess, A., Jiao, Y., Leebens-Mack, J., Liu, K., Mei, W., Meyers, B.C., Shahid, S., Wafula, E., Walts, B., Wessler, S.R., Zhai, J., Zhang, X., Albert, V.A., Carretero-Paulet, L., dePamphilis, C.W., Der, J.P., Jiao, Y., Leebens-Mack, J., Lyons, E., Sankoff, D., Tang, H., Wafula, E., Zheng, C., Albert, V.A., Altman, N.S., Barbazuk, W.B., Carretero-Paulet, L., dePamphilis, C.W., Der, J.P., Estill, J.C., Jiao, Y., Leebens-Mack, J., Liu, K., Mei, W., Wafula, E., Altman, N.S., Arikit, S., Axtell, M.J., Chamala, S., Chanderbali, A.S., Chen, F., Chen, J.-Q., Chiang, V., Paoli, E.D., dePamphilis, C.W., Der, J.P., Determann, R., Fogliani, B., Guo, C., Harholt, J., Harkess, A., Job, C., Job, D., Kim, S., Kong, H., Leebens-Mack, J., Li, G., Li, L., Liu, J., Ma, H., Meyers, B.C., Park, J., Qi, X., Rajjou, L., Burtet-Sarramegna, V., Sederoff, R., Shahid, S., Soltis, D.E., Soltis, P.S., Sun, Y.-H., Ulvskov, P., Villegente, M., Xue, J.-Y., Yeh, T.-F., Yu, X., Zhai, J., Acosta, J.J., Albert, V.A., Barbazuk, W.B., Bruenn, R.A., Chamala, S., Kochko, A. de, dePamphilis, C.W., Der, J.P., Herrera-Estrella, L.R., Ibarra-Laclette, E., Kirst, M., Leebens-Mack, J., Pissis, S.P., Poncet, V., Schuster, S.C., Soltis, D.E., Soltis, P.S., Tomsho, L., 2013. The Amborella Genome and the Evolution of Flowering Plants. Science 342, 1241089. doi:10.1126/science.1241089

Albrecht, V., Ritz, O., Linder, S., Harter, K., Kudla, J., 2001. The NAF domain defines a novel protein-protein interaction module conserved in Ca2+-regulated kinases. EMBO J. 20, 1051–1063. doi:10.1093/emboj/20.5.1051

Argout, X., Salse, J., Aury, J.-M., Guiltinan, M.J., Droc, G., Gouzy, J., Allegre, M., Chaparro, C., Legavre, T., Maximova, S.N., Abrouk, M., Murat, F., Fouet, O., Poulain, J., Ruiz, M., Roguet, Y., Rodier-Goud, M., Barbosa-Neto, J.F., Sabot, F., Kudrna, D., Ammiraju, J.S.S., Schuster, S.C., Carlson, J.E., Sallet, E., Schiex, T., Dievart, A., Kramer, M., Gelley, L., Shi, Z., Bérard, A., Viot, C., Boccara, M., Risterucci, A.M., Guignon, V., Sabau, X., Axtell, M.J., Ma, Z., Zhang, Y., Brown, S., Bourge, M., Golser, W., Song, X., Clement, D., Rivallan, R., Tahi, M., Akaza, J.M., Pitollat, B.,

Gramacho, K., D'Hont, A., Brunel, D., Infante, D., Kebe, I., Costet, P., Wing, R., McCombie, W.R., Guiderdoni, E., Quetier, F., Panaud, O., Wincker, P., Bocs, S., Lanaud, C., 2011. The genome of Theobroma cacao. Nat. Genet. 43, 101–108. doi:10.1038/ng.736

Barker, M.S., Vogel, H., Schranz, M.E., 2010. Paleopolyploidy in the Brassicales: Analyses of the Cleome Transcriptome Elucidate the History of Genome Duplications in Arabidopsis and Other Brassicales. Genome Biol. Evol. 1, 391–399. doi:10.1093/gbe/evp040

Batistic, O., Kim, K.-N., Kleist, T., Kudla, J., Luan, S., 2011. The CBL–CIPK Network for Decoding Calcium Signals in Plants, in: Luan, S. (Ed.), Coding and Decoding of Calcium Signals in Plants. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 235–258.

Batistic, O., Kudla, J., 2004. Integration and channeling of calcium signaling through the CBL calcium sensor/CIPK protein kinase network. Planta 219, 915–924. doi:10.1007/s00425-004-1333-3

Blanc, G., Wolfe, K.H., 2004. Functional Divergence of Duplicated Genes Formed by Polyploidy during Arabidopsis Evolution. Plant Cell 16, 1679–1691. doi:10.1105/tpc.021410

Boc, A., Diallo, A.B., Makarenkov, V., 2012. T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. Nucleic Acids Res. 40, W573–W579. doi:10.1093/nar/gks485

Chase, M.W., 2004. Monocot relationships: an overview. Am. J. Bot. 91, 1645–1655. doi:10.3732/ajb.91.10.1645

Cheng, F., Liu, S., Wu, J., Fang, L., Sun, S., Liu, B., Li, P., Hua, W., Wang, X., 2011. BRAD, the genetics and genomics database for Brassica plants. BMC Plant Biol. 11, 136. doi:10.1186/1471-2229-11-136

Cheng, S., van den Bergh, E., Zeng, P., Zhong, X., Xu, J., Liu, X., Hofberger, J., de Bruijn, S., Bhide, A.S., Kuelahoglu, C., Bian, C., Chen, J., Fan, G., Kaufmann, K., Hall, J.C., Becker, A., Brautigam, A., Weber, A.P.M., Shi, C., Zheng, Z., Li, W., Lv, M., Tao, Y., Wang, J., Zou, H., Quan, Z., Hibberd, J.M., Zhang, G., Zhu, X.-G., Xu, X., Schranz, M.E., 2013. The Tarenaya hassleriana

Genome Provides Insight into Reproductive Trait and Genome Evolution of Crucifers[W][OPEN]. Plant Cell 25, 2813–2830. doi:10.1105/tpc.113.113480

Choi, W.-G., Toyota, M., Kim, S.-H., Hilleary, R., Gilroy, S., 2014. Salt stress-induced Ca2+ waves are associated with rapid, long-distance root-to-shoot signaling in plants. Proc. Natl. Acad. Sci. 111, 6497–6502. doi:10.1073/pnas.1319955111

Cingolani, P., Platts, A., Wang, L.L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w ^{1118} ; iso-2; iso-3. Fly (Austin) 6, 80–92. doi:10.4161/fly.19695

Colmer, T.D., Munns, R., Flowers, T.J., 2005. Improving salt tolerance of wheat and barley: future prospects. Aust. J. Exp. Agric. 45, 1425–1443. doi:10.1071/EA04162

Colmsee, C., Beier, S., Himmelbach, A., Schmutzer, T., Stein, N., Scholz, U., Mascher, M., 2015. BARLEX – the Barley Draft Genome Explorer. Mol. Plant 8, 964–966. doi:10.1016/j.molp.2015.03.009

Consortium, T.U., 2015. UniProt: a hub for protein information. Nucleic Acids Res. 43, D204–D212. doi:10.1093/nar/gku989

Coordinators, N.R., 2013. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. 41, D8–D20. doi:10.1093/nar/gks1189

Cunningham, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S., Gil, L., Girón, C.G., Gordon, L., Hourlier, T., Hunt, S.E., Janacek, S.H., Johnson, N., Juettemann, T., Kähäri, A.K., Keenan, S., Martin, F.J., Maurel, T., McLaren, W., Murphy, D.N., Nag, R., Overduin, B., Parker, A., Patricio, M., Perry, E., Pignatelli, M., Riat, H.S., Sheppard, D., Taylor, K., Thormann, A., Vullo, A., Wilder, S.P., Zadissa, A., Aken, B.L., Birney, E., Harrow, J., Kinsella, R., Muffato, M., Ruffier, M., Searle, S.M.J., Spudich, G., Trevanion, S.J.,

Yates, A., Zerbino, D.R., Flicek, P., 2015. Ensembl 2015. Nucleic Acids Res. 43, D662–D669. doi:10.1093/nar/gku1010

Dassanayake, M., Oh, D.-H., Haas, J.S., Hernandez, A., Hong, H., Ali, S., Yun, D.-J., Bressan, R.A., Zhu, J.-K., Bohnert, H.J., Cheeseman, J.M., 2011. The genome of the extremophile crucifer Thellungiella parvula. Nat. Genet. 43, 913–918. doi:10.1038/ng.889

Des Marais, D.L., Rausher, M.D., 2008. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. Nature 454, 762–765. doi:10.1038/nature07092

Diallo, A.B., Makarenkov, V., Blanchette, M., 2010. Ancestors 1.0: a web server for ancestral sequence reconstruction. Bioinforma. Oxf. Engl. 26, 130–131. doi:10.1093/bioinformatics/btp600

Donoghue, M.T., Keshavaiah, C., Swamidatta, S.H., Spillane, C., 2011. Evolutionary origins of Brassicaceae specific genes in Arabidopsis thaliana. BMC Evol. Biol. 11, 47. doi:10.1186/1471-2148-11-47

Droc, G., Larivière, D., Guignon, V., Yahiaoui, N., This, D., Garsmeur, O., Dereeper, A., Hamelin, C., Argout, X., Dufayard, J.-F., Lengelle, J., Baurens, F.-C., Cenci, A., Pitollat, B., D'Hont, A., Ruiz, M., Rouard, M., Bocs, S., 2013. The Banana Genome Hub. Database 2013, bat035. doi:10.1093/database/bat035

Duvick, J., Fu, A., Muppirala, U., Sabharwal, M., Wilkerson, M.D., Lawrence, C.J., Lushbough, C., Brendel, V., 2008. PlantGDB: a resource for comparative plant genomics. Nucleic Acids Res. 36, D959–D965. doi:10.1093/nar/gkm1041

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. doi:10.1093/nar/gkh340

FAOSTAT, 2014. Food and Agriculture Organization of the United Nations [WWW Document]. Food Agric. Organ. Fo U. N. URL http://www.fao.org/economic/ess/en/#.U1XSWvmSymw (accessed 4.22.14).

Felsenstein, J., 1996. Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. Methods Enzymol. 266, 418–427. doi:10.1016/S0076-6879(96)66026-1

Felsenstein, J., 1985. Confidence Limits on Phylogenies: An Approach Using the Bootstrap. Evolution 39, 783–791. doi:10.2307/2408678

Finn, R.D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E.L.L., Tate, J., Punta, M., 2014. Pfam: the protein families database. Nucleic Acids Res. 42, D222–D230. doi:10.1093/nar/gkt1223

Fulton, D.L., Li, Y.Y., Laird, M.R., Horsman, B.G., Roche, F.M., Brinkman, F.S., 2006. Improving the specificity of high-throughput ortholog prediction. BMC Bioinformatics 7, 270. doi:10.1186/1471-2105-7-270

Goodstein, D.M., Shu, S., Howson, R., Neupane, R., Hayes, R.D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N., Rokhsar, D.S., 2012. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res. 40, D1178–D1186. doi:10.1093/nar/gkr944

Haudry, A., Platts, A.E., Vello, E., Hoen, D.R., Leclercq, M., Williamson, R.J., Forczek, E., Joly-Lopez, Z., Steffen, J.G., Hazzouri, K.M., Dewar, K., Stinchcombe, J.R., Schoen, D.J., Wang, X., Schmutz, J., Town, C.D., Edger, P.P., Pires, J.C., Schumaker, K.S., Jarvis, D.E., Mandáková, T., Lysak, M.A., van den Bergh, E., Schranz, M.E., Harrison, P.M., Moses, A.M., Bureau, T.E., Wright, S.I., Blanchette, M., 2013. An atlas of over 90,000 conserved noncoding sequences provides insight into crucifer regulatory regions. Nat. Genet. 45, 891–898. doi:10.1038/ng.2684

Hrabak, E.M., Chan, C.W.M., Gribskov, M., Harper, J.F., Choi, J.H., Halford, N., Kudla, J., Luan, S., Nimmo, H.G., Sussman, M.R., Thomas, M., Walker-Simmons, K., Zhu, J.-K., Harmon, A.C., 2003. The Arabidopsis CDPK-SnRK Superfamily of Protein Kinases. Plant Physiol. 132, 666–680. doi:10.1104/pp.102.011999

International Wheat Genome Sequencing Consortium (IWGSC), 2014. A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. Science 345, 1251788. doi:10.1126/science.1251788

Jaillon, O., Aury, J.-M., Noel, B., Policriti, A., Clepet, C., Casagrande, A., Choisne, N., Aubourg, S., Vitulo, N., Jubin, C., Vezzi, A., Legeai, F., Hugueney, P., Dasilva, C., Horner, D., Mica, E., Jublot, D., Poulain, J., Bruyère, C., Billault, A., Segurens, B., Gouyvenoux, M., Ugarte, E., Cattonaro, F., Anthouard, V., Vico, V., Del Fabbro, C., Alaux, M., Di Gaspero, G., Dumas, V., Felice, N., Paillard, S., Juman, I., Moroldo, M., Scalabrin, S., Canaguier, A., Le Clainche, I., Malacrida, G., Durand, E., Pesole, G., Laucou, V., Chatelet, P., Merdinoglu, D., Delledonne, M., Pezzotti, M., Lecharny, A., Scarpelli, C., Artiguenave, F., Pè, M.E., Valle, G., Morgante, M., Caboche, M., Adam-Blondon, A.-F., Weissenbach, J., Quétier, F., Wincker, P., 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449, 463–467. doi:10.1038/nature06148

Jung, S., Ficklin, S.P., Lee, T., Cheng, C.-H., Blenda, A., Zheng, P., Yu, J., Bombarely, A., Cho, I., Ru, S., Evans, K., Peace, C., Abbott, A.G., Mueller, L.A., Olmstead, M.A., Main, D., 2014. The Genome Database for Rosaceae (GDR): year 10 update. Nucleic Acids Res. 42, D1237–D1244. doi:10.1093/nar/gkt1012

Kagale, S., Koh, C., Nixon, J., Bollina, V., Clarke, W.E., Tuteja, R., Spillane, C., Robinson, S.J., Links, M.G., Clarke, C., Higgins, E.E., Huebert, T., Sharpe, A.G., Parkin, I.A.P., 2014. The emerging biofuel crop Camelina sativa retains a highly undifferentiated hexaploid genome structure. Nat. Commun. 5, 3706. doi:10.1038/ncomms4706

Kagale, S., Robinson, S.J., Nixon, J., Xiao, R., Huebert, T., Condie, J., Kessler, D., Clarke, W.E., Edger, P.P., Links, M.G., Sharpe, A.G., Parkin, I.A.P., 2014. Polyploid Evolution of the Brassicaceae during the Cenozoic Era. Plant Cell 26, 2777–2791. doi:10.1105/tpc.114.126391

Keane, T.M., Creevey, C.J., Pentony, M.M., Naughton, T.J., Mclnerney, J.O., 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol. Biol. 6, 29. doi:10.1186/1471-2148-6-29

Kim, K.-N., Cheong, Y.H., Gupta, R., Luan, S., 2000. Interaction Specificity of Arabidopsis Calcineurin B-Like Calcium Sensors and Their Target Kinases. Plant Physiol. 124, 1844–1853. doi:10.1104/pp.124.4.1844

Kitashiba, H., Li, F., Hirakawa, H., Kawanabe, T., Zou, Z., Hasegawa, Y., Tonosaki, K., Shirasawa, S., Fukushima, A., Yokoi, S., Takahata, Y., Kakizaki, T., Ishida, M., Okamoto, S., Sakamoto, K., Shirasawa, K., Tabata, S., Nishio, T., 2014. Draft Sequences of the Radish (Raphanus sativus L.) Genome. DNA Res. dsu014. doi:10.1093/dnares/dsu014

Kolukisaoglu, Ü., Weinl, S., Blazevic, D., Batistic, O., Kudla, J., 2004. Calcium Sensors and Their Interacting Protein Kinases: Genomics of the Arabidopsis and Rice CBL-CIPK Signaling Networks. Plant Physiol. 134, 43–58. doi:10.1104/pp.103.033068

Kosugi, S., Hasebe, M., Tomita, M., Yanagawa, H., 2009. Systematic identification of cell cycle-dependent yeast nucleocytoplasmic shuttling proteins by prediction of composite motifs. Proc. Natl. Acad. Sci. U. S. A. 106, 10171–10176. doi:10.1073/pnas.0900604106

Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M., Karthikeyan, A.S., Lee, C.H., Nelson, W.D., Ploetz, L., Singh, S., Wensel, A., Huala, E., 2012. The Arabidopsis Information Resource (TAIR): improved gene annotation and new tools. Nucleic Acids Res. 40, D1202–D1210. doi:10.1093/nar/gkr1090

Lan, W.-Z., Lee, S.-C., Che, Y.-F., Jiang, Y.-Q., Luan, S., 2011. Mechanistic Analysis of AKT1 Regulation by the CBL–CIPK–PP2CA Interactions. Mol. Plant 4, 527–536. doi:10.1093/mp/ssr031

Lee, S.C., Lan, W.-Z., Kim, B.-G., Li, L., Cheong, Y.H., Pandey, G.K., Lu, G., Buchanan, B.B., Luan, S., 2007. A protein phosphorylation/dephosphorylation network regulates a plant potassium channel. Proc. Natl. Acad. Sci. 104, 15959–15964. doi:10.1073/pnas.0707912104

Lyons, E., Freeling, M., 2008. How to usefully compare homologous plant genes and chromosomes as DNA sequences. Plant J. 53, 661–673. doi:10.1111/j.1365-313X.2007.03326.x

Meena, M.K., Ghawana, S., Dwivedi, V., Roy, A., Chattopadhyay, D., 2015. Expression of chickpea CIPK25 enhances root growth and tolerance to dehydration and salt stress in transgenic tobacco. Plant Physiol. 683. doi:10.3389/fpls.2015.00683

Michael, T.P., Jackson, S., 2013. The First 50 Plant Genomes. Plant Genome 6, 0. doi:10.3835/plantgenome2013.03.0001in

Moghe, G.D., Hufnagel, D.E., Tang, H., Xiao, Y., Dworkin, I., Town, C.D., Conner, J.K., Shiu, S.-H., 2014. Consequences of Whole-Genome Triplication as Revealed by Comparative Genomic Analyses of the Wild Radish Raphanus raphanistrum and Three Other Brassicaceae Species. Plant Cell Online tpc.114.124297. doi:10.1105/tpc.114.124297

Morgenstern, B., 1999. DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. Bioinforma. Oxf. Engl. 15, 211–218. doi:10.1093/bioinformatics/15.3.211

Mount, D.W., 2008. The Maximum Likelihood Approach for Phylogenetic Prediction. Cold Spring Harb. Protoc. 2008, pdb.top34. doi:10.1101/pdb.top34

Munns, R., 2005. Genes and salt tolerance: bringing them together. New Phytol. 167, 645–663. doi:10.1111/j.1469-8137.2005.01487.x

Munns, R., Tester, M., 2008. Mechanisms of Salinity Tolerance. Annu. Rev. Plant Biol. 59, 651–681.doi:10.1146/annurev.arplant.59.032607.092911

Munns, R., Gilliham, M., 2015. Salinity tolerance of crops – what is the cost? New Phytol. 668–673. doi:10.1111/nph.13519

Nolen, B., Taylor, S., Ghosh, G., 2004. Regulation of Protein Kinases: Controlling Activity through Activation Segment Conformation. Mol. Cell 15, 661–675. doi:10.1016/j.molcel.2004.08.024

Nordberg, H., Cantor, M., Dusheyko, S., Hua, S., Poliakov, A., Shabalov, I., Smirnova, T., Grigoriev, I.V., Dubchak, I., 2014. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. Nucleic Acids Res. 42, D26–31. doi:10.1093/nar/gkt1069

Nussbaumer, T., Martis, M.M., Roessner, S.K., Pfeifer, M., Bader, K.C., Sharma, S., Gundlach, H., Spannagl, M., 2013. MIPS PlantsDB: a database framework for comparative plant genome research. Nucleic Acids Res. 41, D1144–D1151. doi:10.1093/nar/gks1153

Nystedt, B., Street, N.R., Wetterbom, A., Zuccolo, A., Lin, Y.-C., Scofield, D.G., Vezzi, F., Delhomme, N., Giacomello, S., Alexeyenko, A., Vicedomini, R., Sahlin, K., Sherwood, E., Elfstrand, M., Gramzow, L., Holmberg, K., Hällman, J., Keech, O., Klasson, L., Koriabine, M., Kucukoglu, M., Käller, M., Luthman, J., Lysholm, F., Niittylä, T., Olson, Å., Rilakovic, N., Ritland, C., Rosselló, J.A., Sena, J., Svensson, T., Talavera-López, C., Theißen, G., Tuominen, H., Vanneste, K., Wu, Z.-Q., Zhang, B., Zerbe, P., Arvestad, L., Bhalerao, R., Bohlmann, J., Bousquet, J., Garcia Gil, R., Hvidsten, T.R., de Jong, P., MacKay, J., Morgante, M., Ritland, K., Sundberg, B., Lee Thompson, S., Van de Peer, Y., Andersson, B., Nilsson, O., Ingvarsson, P.K., Lundeberg, J., Jansson, S., 2013. The Norway spruce genome sequence and conifer genome evolution. Nature 497, 579–584. doi:10.1038/nature12211

Ohta, M., Guo, Y., Halfter, U., Zhu, J.-K., 2003. A novel domain in the protein kinase SOS2 mediates interaction with the protein phosphatase 2C ABI2. Proc. Natl. Acad. Sci. U. S. A. 100, 11771–11776. doi:10.1073/pnas.2034853100

Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., Thibaud-Nissen, F., Malek, R.L., Lee, Y., Zheng, L., Orvis, J., Haas, B., Wortman, J., Buell, C.R., 2007. The TIGR Rice Genome Annotation Resource: improvements and new features. Nucleic Acids Res. 35, D883–D887. doi:10.1093/nar/gkl976

Pérez-Pérez, J.M., Candela, H., Micol, J.L., 2009. Understanding synergy in genetic interactions. Trends Genet. 25, 368–376. doi:10.1016/j.tig.2009.06.004

Petersen, B., Petersen, T.N., Andersen, P., Nielsen, M., Lundegaard, C., 2009. A generic method for assignment of reliability scores applied to solvent accessibility predictions. BMC Struct. Biol. 9, 51. doi:10.1186/1472-6807-9-51

Pimentel, D., Berger, B., Filiberto, D., Newton, M., Wolfe, B., Karabinakis, E., Clark, S., Poon, E., Abbett, E., Nandagopal, S., 2004. Water Resources: Agricultural and Environmental Issues. BioScience 54, 909. doi:10.1641/0006-3568(2004)054[0909:WRAAEI]2.0.CO;2

Plett, D., Møller, I.S., 2010. Na+ transport in glycophytic plants: what we know and would like to know. Plant Cell Environ. 33, 612–626. doi:10.1111/j.1365-3040.2009.02086.x

Qadir, M., Quillérou, E., Nangia, V., Murtaza, G., Singh, M., Thomas, R. j., Drechsel, P., Noble, A. d., 2014. Economics of salt-induced land degradation and restoration. Nat. Resour. Forum 38, 282–295. doi:10.1111/1477-8947.12054

Rajendran, K., Tester, M., Roy, S.J., 2009. Quantifying the three main components of salinity tolerance in cereals. Plant Cell Environ. 32, 237–249. doi:10.1111/j.1365-3040.2008.01916.x

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., Mesirov, J.P., 2011. Integrative Genomics Viewer. Nat. Biotechnol. 29, 24–26. doi:10.1038/nbt.1754

Rost, B., Yachdav, G., Liu, J., 2004. The PredictProtein server. Nucleic Acids Res. 32, W321–W326. doi:10.1093/nar/gkh377

Roy, S.J., Huang, W., Wang, X.J., Evrard, A., Schmöckel, S.M., Zafar, Z.U., Tester, M., 2013. A novel protein kinase involved in Na+ exclusion revealed from positional cloning. Plant Cell Environ. 36, 553–568. doi:10.1111/j.1365-3040.2012.02595.x

Roy, S.J., Negrão, S., Tester, M., 2014. Salt resistant crop plants. Curr. Opin. Biotechnol., Food biotechnology - Plant biotechnology 26, 115–124. doi:10.1016/j.copbio.2013.12.004

Sanchez-Barrena, M.J., Fujii, H., Angulo, I., Martinez-Ripoll, M., Zhu, J.-K., Albert, A., 2007. The Structure of the C-Terminal Domain of the Protein Kinase AtSOS2 Bound to the Calcium Sensor AtSOS3. Mol. Cell 26, 427–435. doi:10.1016/j.molcel.2007.04.013

Sanchez-Barrena, M.J., Martinez-Ripoll, M., Albert, A., 2013. Structural Biology of a Major Signaling Network that Regulates Plant Abiotic Stress: The CBL-CIPK Mediated Pathway. Int. J. Mol. Sci. 14, 5734–5749. doi:10.3390/ijms14035734

Shabala, S., 2013. Learning from halophytes: physiological basis and strategies to improve abiotic stress tolerance in crops. Ann. Bot. mct205. doi:10.1093/aob/mct205

Schlücking, K., Edel, K.H., Köster, P., Drerup, M.M., Eckert, C., Steinhorst, L., Waadt, R., Batistič, O., Kudla, J., 2013. A New β-Estradiol-Inducible Vector Set that Facilitates Easy Construction and Efficient Expression of Transgenes Reveals CBL3-Dependent Cytoplasm to Tonoplast Translocation of CIPK5. Mol. Plant 6, 1814–1829. doi:10.1093/mp/sst065

Schranz, M.E., Mitchell-Olds, T., 2006. Independent Ancient Polyploidy Events in the Sister Families Brassicaceae and Cleomaceae. Plant Cell Online 18, 1152–1165. doi:10.1105/tpc.106.041111

Sigrist, C.J.A., de Castro, E., Cerutti, L., Cuche, B.A., Hulo, N., Bridge, A., Bougueleret, L., Xenarios, I., 2013. New and continuing developments at PROSITE. Nucleic Acids Res. 41, D344–347. doi:10.1093/nar/gks1067

Soltis, D.E., Smith, S.A., Cellinese, N., Wurdack, K.J., Tank, D.C., Brockington, S.F., Refulio-Rodriguez, N.F., Walker, J.B., Moore, M.J., Carlsward, B.S., Bell, C.D., Latvis, M., Crawley, S., Black, C., Diouf, D., Xi, Z., Rushworth, C.A., Gitzendanner, M.A., Sytsma, K.J., Qiu, Y.-L., Hilu, K.W., Davis, C.C., Sanderson, M.J., Beaman, R.S., Olmstead, R.G., Judd, W.S., Donoghue, M.J., Soltis, P.S., 2011. Angiosperm phylogeny: 17 genes, 640 taxa. Am. J. Bot. 98, 704–730. doi:10.3732/ajb.1000404

Soltis, D.E., Soltis, P.S., 2003. The Role of Phylogenetics in Comparative Genetics. Plant Physiol. 132, 1790–1800. doi:10.1104/pp.103.022509

Talavera, G., Castresana, J., 2007. Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. Syst. Biol. 56, 564–577. doi:10.1080/10635150701472164

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol. Biol. Evol. 30, 2725–2729. doi:10.1093/molbev/mst197

Tavakkoli, E., Rengasamy, P., McDonald, G.K., 2010. The response of barley to salinity stress differs between hydroponic and soil systems. Funct. Plant Biol. 37, 621. doi:10.1071/FP09202

Thoday-Kennedy, Jacobs, A.K., Roy, S.J., 2015. The role of the CBL–CIPK calcium signalling network in regulating ion transport in response to abiotic stress. Plant Growth Regul. 76, 3–12. doi:10.1007/s10725-015-0034-1

Tracy, F.E., Gilliham, M., Dodd, A.N., Webb, A. a. R., Tester, M., 2008. NaCl-induced changes in cytosolic free Ca2+ in Arabidopsis thaliana are heterogeneous and modified by external ionic composition. Plant Cell Environ. 31, 1063–1073. doi:10.1111/j.1365-3040.2008.01817.x

Wang, Y., You, F.M., Lazo, G.R., Luo, M.-C., Thilmony, R., Gordon, S., Kianian, S.F., Gu, Y.Q., 2013. PIECE: a database for plant gene structure comparison and evolution. Nucleic Acids Res. 41, D1159–D1166. doi:10.1093/nar/gks1109

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J., 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189–1191. doi:10.1093/bioinformatics/btp033

Weinl, S., Kudla, J., 2009. The CBL–CIPK Ca2+-decoding signaling network: function and perspectives. New Phytol. 184, 517–528. doi:10.1111/j.1469-8137.2009.02938.x

Wernersson, R., Pedersen, A.G., 2003. RevTrans: multiple alignment of coding DNA from aligned amino acid sequences. Nucleic Acids Res. 31, 3537–3539. doi:10.1093/nar/gkg609

Wu, F., Mueller, L.A., Crouzillat, D., Pétiard, V., Tanksley, S.D., 2006. Combining Bioinformatics and Phylogenetics to Identify Large Sets of Single-Copy Orthologous Genes (COSII) for

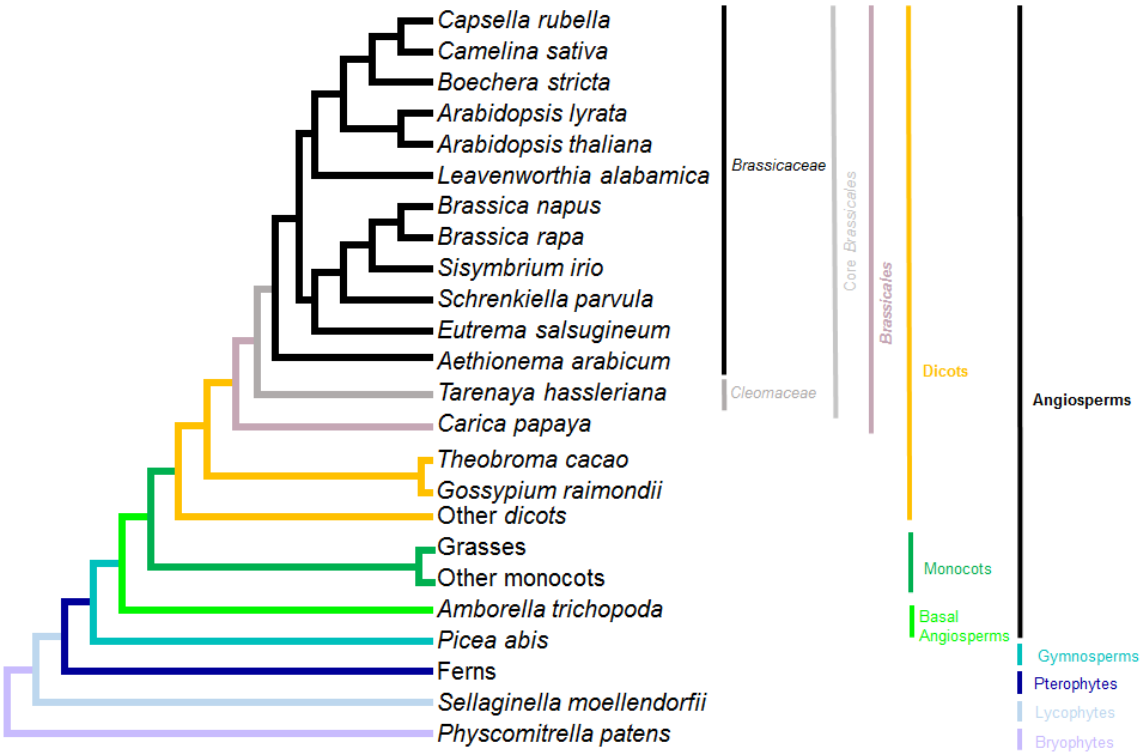
Comparative, Evolutionary and Systematic Studies: A Test Case in the Euasterid Plant Clade. Genetics 174, 1407–1420. doi:10.1534/genetics.106.062455

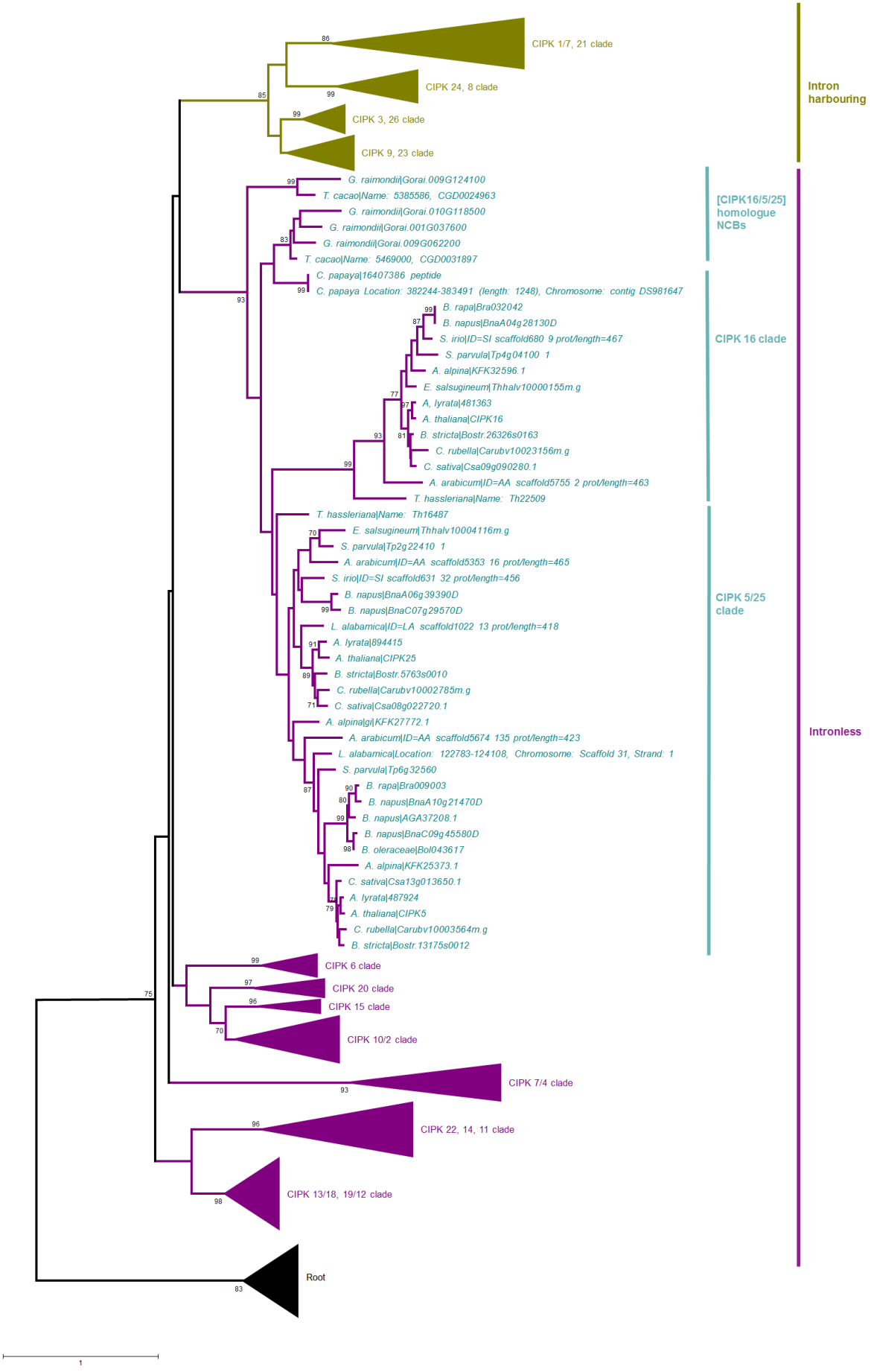
Yang, R., Jarvis, D.E., Chen, H., Beilstein, M.A., Grimwood, J., Jenkins, J., Shu, S., Prochnik, S., Xin, M., Ma, C., Schmutz, J., Wing, R.A., Mitchell-Olds, T., Schumaker, K.S., Wang, X., 2013. The Reference Genome of the Halophytic Plant Eutrema salsugineum. Front. Plant Sci. 4. doi:10.3389/fpls.2013.00046

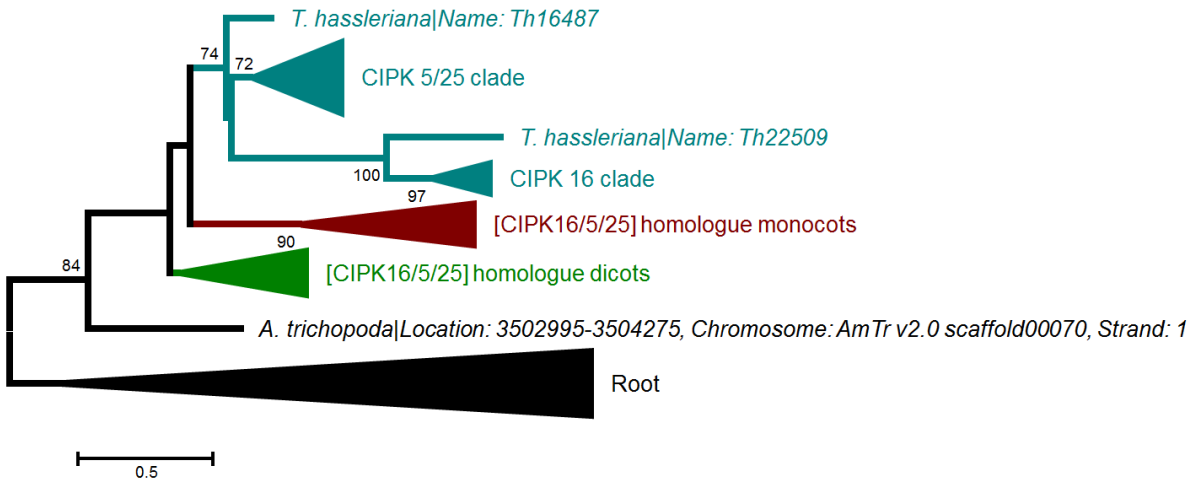
Yang, Z., 1996. Among-site rate variation and its impact on phylogenetic analyses. Trends Ecol. Evol. 11, 367–372. doi:10.1016/0169-5347(96)10041-0

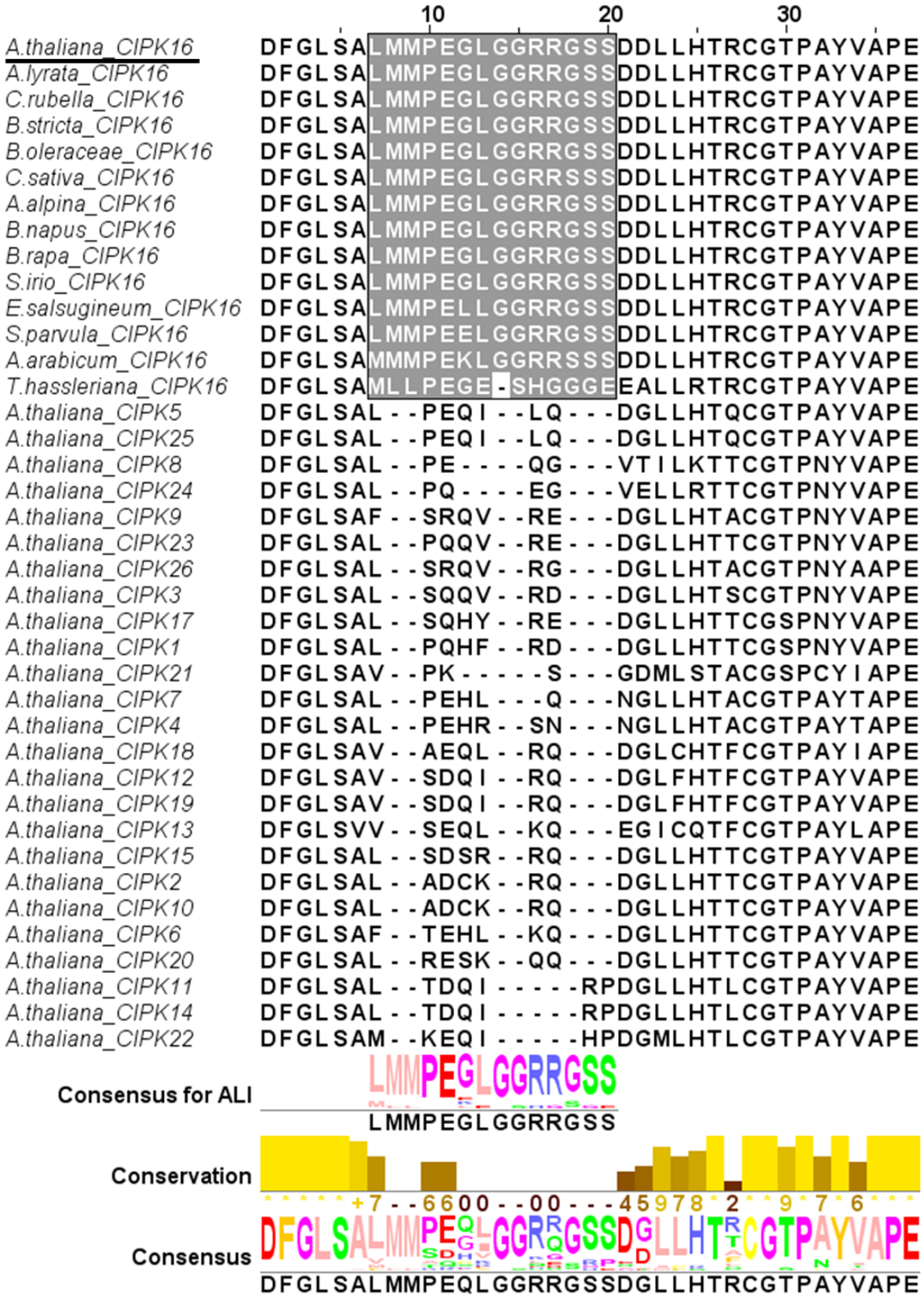
Yang, Z., Rannala, B., 2012. Molecular phylogenetics: principles and practice. Nat. Rev. Genet. 13, 303–314. doi:10.1038/nrg3186

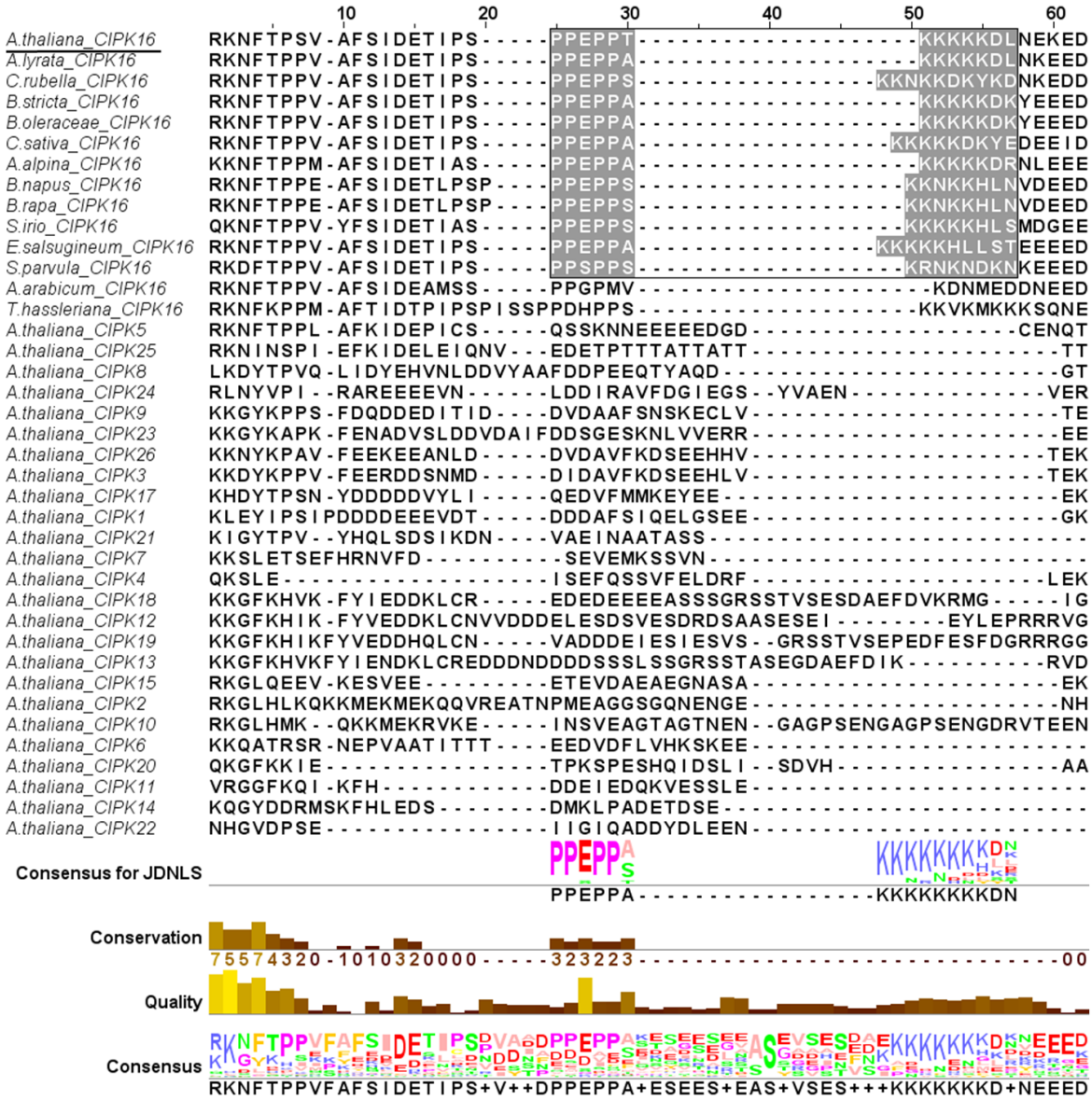
Yoon, S., Park, J., Ryu, M., Yoon, I.S., Kim, K.-N., 2009. Calcineurin B-like proteins in rice can bind with calcium ion and associate with the Arabidopsis CIPK family members. Plant Sci. 177, 577–583. doi:10.1016/j.plantsci.2009.08.011

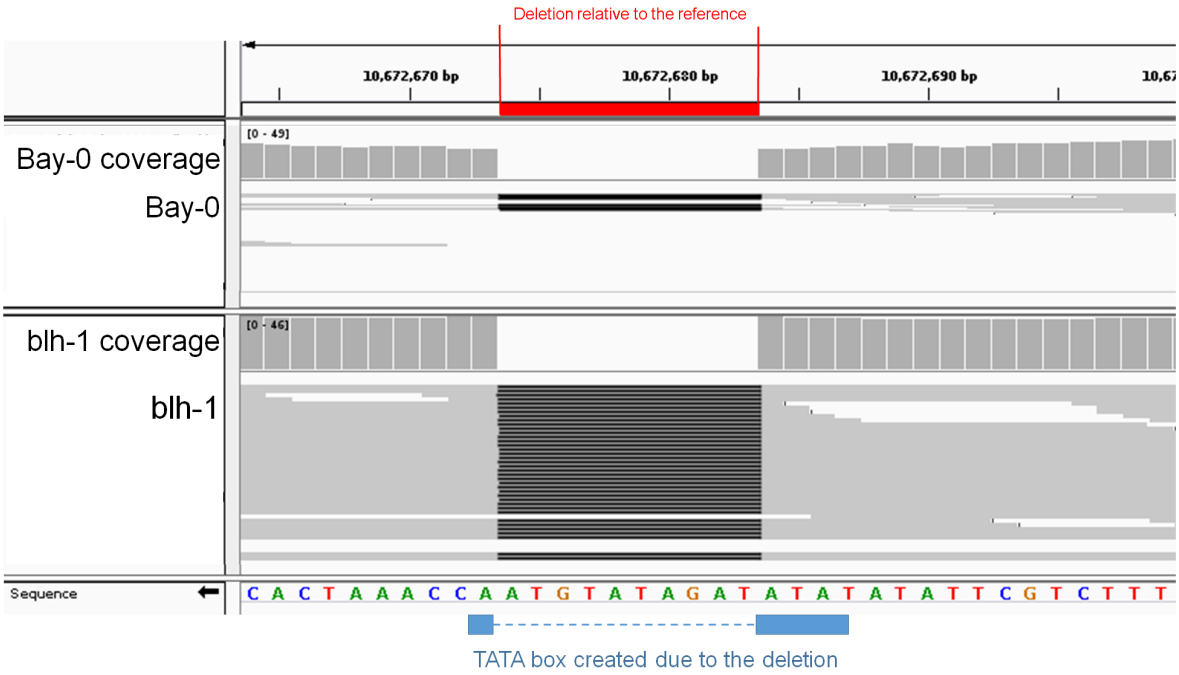

Capsella rubella
Camelina sativa
Boechera stricta
Arabidopsis lyrata
Arabidopsis thaliana
Leavenworthia alabamica
Brassica napus
Brassica rapa
Sisymbrium irio
Schrenkiella parvula
Eutrema salsugineum
Aethionema arabicum
Tarenaya hassleriana
Carica papaya
Theobroma cacao
Gossypium raimondii
Other dicots
Grasses
Other monocots
Amborella trichopoda
Picea abis
Ferns
Sellaginella moellendorfii
Physcomitrella patens
Brassicaceae
Cleomaceae
Core Brassicales
Brassicales
Dicots
Monocots
Basal Angiosperms
Angiosperms
Gymnosperms
Pterophytes
Lycophytes
Bryophytes

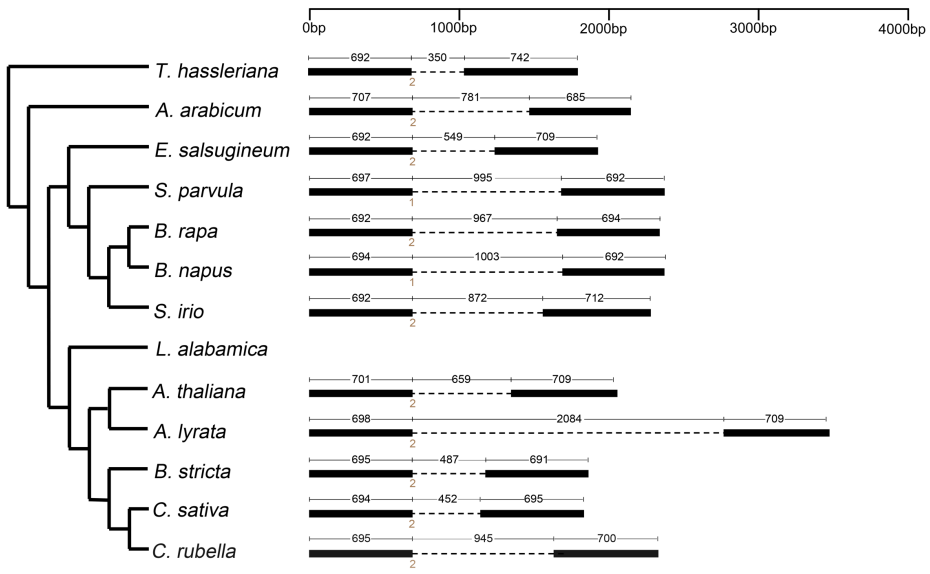

CIPK 1/7, 21 clade
CIPK 24, 8 clade
CIPK 3, 26 clade
CIPK 9, 23 clade
Intron harbouring
G. raimondii|Gorai.009G124100
T. cacao|Name: 5385586, CGD0024963
G. raimondii|Gorai.010G118500
G. raimondii|Gorai.001G037600
G. raimondii|Gorai.009G062200
T. cacao|Name: 5469000, CGD0031897
[CIPK16/5/25] homologue NCBs
C. papaya|16407386 peptide
C. papaya Location: 382244-383491 (length: 1248), Chromosome: contig DS981647
B. rapa|Bra032042
B. napus|BnaA04g28130D
S. irio|ID=SI scaffold680 9 prot/length=467
S. parvula|Tp4g04100 1
A. alpina|KFK32596.1
E. salsugineum|Thhalv10000155m.g
A, lyrata|481363
A. thaliana|CIPK16
B. stricta|Bostr.26326s0163
C. rubella|Carubv10023156m.g
C. sativa|Csa09g090280.1
A. arabicum|ID=AA scaffold5755 2 prot/length=463
T. hassleriana|Name: Th22509
CIPK 16 clade
T. hassleriana|Name: Th16487
E. salsugineum|Thhalv10004116m.g
S. parvula|Tp2g22410 1
A. arabicum|ID=AA scaffold5353 16 prot/length=465
S. irio|ID=SI scaffold631 32 prot/length=456
B. napus|BnaA06g39390D
B. napus|BnaC07g29570D
L. alabamica|ID=LA scaffold1022 13 prot/length=418
A. lyrata|894415
A. thaliana|CIPK25
B. stricta|Bostr.5763s0010
C. rubella|Carubv10002785m.g
C. sativa|Csa08g022720.1
A. alpina|gi|KFK27772.1
A. arabicum|ID=AA scaffold5674 135 prot/length=423
L. alabamica|Location: 122783-124108, Chromosome: Scaffold 31, Strand: 1
S. parvula|Tp6g32560
B. rapa|Bra009003
B. napus|BnaA10g21470D
B. napus|AGA37208.1
B. napus|BnaC09g45580D
B. oleraceae|Bol043617
A. alpina|KFK25373.1
C. sativa|Csa13g013650.1
A. lyrata|487924
A. thaliana|CIPK5
C. rubella|Carubv10003564m.g
B. stricta|Bostr.13175s0012
CIPK 5/25 clade
Intronless
CIPK 6 clade
CIPK 20 clade
CIPK 15 clade
CIPK 10/2 clade
CIPK 7/4 clade
CIPK 22, 14, 11 clade
CIPK 13/18, 19/12 clade
Root
1

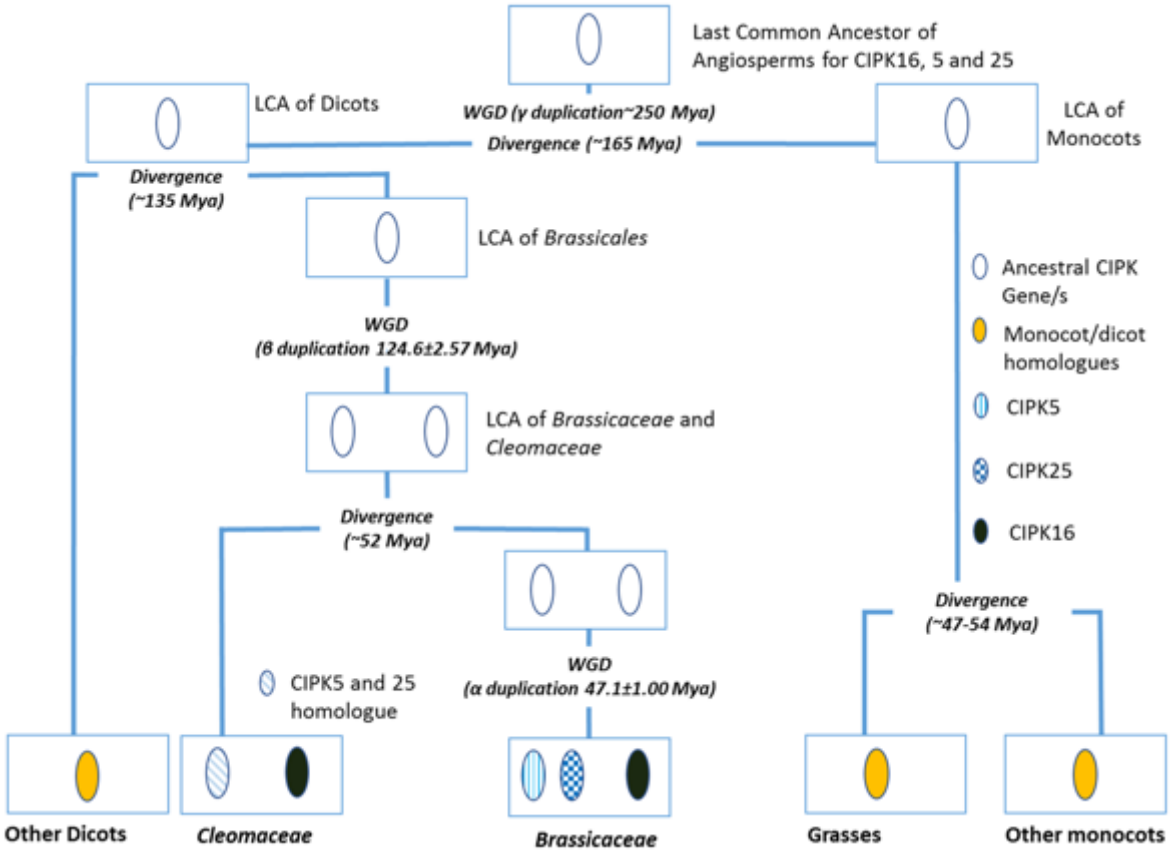

T. hassleriana|Name: Th16487
74
72
CIPK 5/25 clade
T. hassleriana|Name: Th22509
100
CIPK 16 clade
97
[CIPK16/5/25] homologue monocots
90
84
[CIPK16/5/25] homologue dicots
A. trichopoda|Location: 3502995-3504275, Chromosome: AmTr v2.0 scaffold00070, Strand: 1
Root
0.5


10
20
30
A.thaliana_CIPK16 DFGLSALMMPEGLGGRRGSSDDLLHTRCGTPAYVAPE
A.lyrata_CIPK16 DFGLSALMMPEGLGGRRGSSDDLLHTRCGTPAYVAPE
C.rubella_CIPK16 DFGLSALMMPEGLGGRRGSSDDLLHTRCGTPAYVAPE
B.stricta_CIPK16 DFGLSALMMPEGLGGRRGSSDDLLHTRCGTPAYVAPE
B.oleraceae_CIPK16 DFGLSALMMPEGLGGRRGSSDDLLHTRCGTPAYVAPE
C.sativa_CIPK16 DFGLSALMMPEGLGGRRSSSDDLLHTRCGTPAYVAPE
A.alpina_CIPK16 DFGLSALMMPEGLGGRRGSSDDLLHTRCGTPAYVAPE
B.napus_CIPK16 DFGLSALMMPEGLGGRRGSSDDLLHTRCGTPAYVAPE
B.rapa_CIPK16 DFGLSALMMPEGLGGRRGSSDDLLHTRCGTPAYVAPE
S.irio_CIPK16 DFGLSALMMPEGLGGRRGSSDDLLHTRCGTPAYVAPE
E.salsugineum_CIPK16 DFGLSALMMPELLGGRRGSSDDLLHTRCGTPAYVAPE
S.parvula_CIPK16 DFGLSALMMPEELGGRRGSSDDLLHTRCGTPAYVAPE
A.arabicum_CIPK16 DFGLSAMMMPEKLGGRRSSSDDLLHTRCGTPAYVAPE
T.hassleriana_CIPK16 DFGLSAMLLPEGE-SHGGGEEALLRTRCGTPAYVAPE
A.thaliana_CIPK5 DFGLSAL--PEQI--LQ---DGLLHTQCGTPAYVAPE
A.thaliana_CIPK25 DFGLSAL--PEQI--LQ---DGLLHTQCGTPAYVAPE
A.thaliana_CIPK8 DFGLSAL--PE----QG---VTILKTTCGTPNYVAPE
A.thaliana_CIPK24 DFGLSAL--PQ----EG---VELLRTTCGTPNYVAPE
A.thaliana_CIPK9 DFGLSAF--SRQV--RE---DGLLHTACGTPNYVAPE
A.thaliana_CIPK23 DFGLSAL--PQQV--RE---DGLLHTTCGTPNYVAPE
A.thaliana_CIPK26 DFGLSAL--SRQV--RG---DGLLHTACGTPNYAAPE
A.thaliana_CIPK3 DFGLSAL--SQQV--RD---DGLLHTSCGTPNYVAPE
A.thaliana_CIPK17 DFGLSAL--SQHY--RE---DGLLHTTCGSPNYVAPE
A.thaliana_CIPK1 DFGLSAL--PQHF--RD---DGLLHTTCGSPNYVAPE
A.thaliana_CIPK21 DFGLSAV--PK-----S---GDMLSTACGSPCYIAPE
A.thaliana_CIPK7 DFGLSAL--PEHL---Q---NGLLHTACGTPAYTAPE
A.thaliana_CIPK4 DFGLSAL--PEHR--SN---NGLLHTACGTPAYTAPE
A.thaliana_CIPK18 DFGLSAV--AEQL--RQ---DGLCHTFCGTPAYIAPE
A.thaliana_CIPK12 DFGLSAV--SDQI--RQ---DGLFHTFCGTPAYVAPE
A.thaliana_CIPK19 DFGLSAV--SDQI--RQ---DGLFHTFCGTPAYVAPE
A.thaliana_CIPK13 DFGLSVV--SEQL--KQ---EGICQTFCGTPAYLAPE
A.thaliana_CIPK15 DFGLSAL--SDSR--RQ---DGLLHTTCGTPAYVAPE
A.thaliana_CIPK2 DFGLSAL--ADCK--RQ---DGLLHTTCGTPAYVAPE
A.thaliana_CIPK10 DFGLSAL--ADCK--RQ---DGLLHTTCGTPAYVAPE
A.thaliana_CIPK6 DFGLSAF--TEHL--KQ---DGLLHTTCGTPAYVAPE
A.thaliana_CIPK20 DFGLSAL--RESK--QQ---DGLLHTTCGTPAYVAPE
A.thaliana_CIPK11 DFGLSAL--TDQI-----RPDGLLHTLCGTPAYVAPE
A.thaliana_CIPK14 DFGLSAL--TDQI-----RPDGLLHTLCGTPAYVAPE
A.thaliana_CIPK22 DFGLSAM--KEQI-----HPDGMLHTLCGTPAYVAPE
Consensus for ALI
LMMPEGLGGRRGSS
Conservation
Consensus
DFGLSALMMPEQLGGRRGSSDGLLHTRCGTPAYVAPE


10
20
30
40
50
60
A.thaliana_CIPK16
A.lyrata_CIPK16
C.rubella_CIPK16
B.stricta_CIPK16
B.oleraceae_CIPK16
C.sativa_CIPK16
A.alpina_CIPK16
B.napus_CIPK16
B.rapa_CIPK16
S.irio_CIPK16
E.salsugineum_CIPK16
S.parvula_CIPK16
A.arabicum_CIPK16
T.hassleriana_CIPK16
A.thaliana_CIPK5
A.thaliana_CIPK25
A.thaliana_CIPK8
A.thaliana_CIPK24
A.thaliana_CIPK9
A.thaliana_CIPK23
A.thaliana_CIPK26
A.thaliana_CIPK3
A.thaliana_CIPK17
A.thaliana_CIPK1
A.thaliana_CIPK21
A.thaliana_CIPK7
A.thaliana_CIPK4
A.thaliana_CIPK18
A.thaliana_CIPK12
A.thaliana_CIPK19
A.thaliana_CIPK13
A.thaliana_CIPK15
A.thaliana_CIPK2
A.thaliana_CIPK10
A.thaliana_CIPK6
A.thaliana_CIPK20
A.thaliana_CIPK11
A.thaliana_CIPK14
A.thaliana_CIPK22
Consensus for JDNLS
PPEPPA----------KKKKKKKKKDN
Conservation
Quality
Consensus
RKNFTPPVFAFSIDETIPS+V++DPPEPPA+ESEES+EAS+VSES+++KKKKKKKD+NEEED


Deletion relative to the reference
10,672,670 bp
10,672,680 bp
10,672,690 bp
Bay-0 coverage
[0 - 49]
Bay-0
blh-1 coverage
[0 - 46]
blh-1
Sequence
C A C T A A A C C A A T G T A T A G A T A T A T A T A T T C G T C T T T
TATA box created due to the deletion


0bp
1000bp
2000bp
3000bp
4000bp
T. hassleriana
692
350
742
2
A. arabicum
707
781
685
2
E. salsugineum
692
549
709
2
S. parvula
697
995
692
1
B. rapa
692
967
694
2
B. napus
694
1003
692
1
S. irio
692
872
712
2
L. alabamica
A. thaliana
701
659
709
2
A. lyrata
698
2084
709
2
B. stricta
695
487
691
2
C. sativa
694
452
695
2
C. rubella
695
945
700
2


Last Common Ancestor of Angiosperms for CIPK16, 5 and 25
LCA of Dicots
WGD (γ duplication~250 Mya)
Divergence (~165 Mya)
LCA of Monocots
Divergence (~135 Mya)
LCA of Brassicales
Ancestral CIPK Gene/s
Monocot/dicot homologues
CIPK5
CIPK25
CIPK16
WGD (β duplication 124.6±2.57 Mya)
LCA of Brassicaceae and Cleomaceae
Divergence (~52 Mya)
Divergence (~47-54 Mya)
CIPK5 and 25 homologue
WGD (α duplication 47.1±1.00 Mya)
Other Dicots
Cleomaceae
Brassicaceae
Grasses
Other monocots