# Calcium-dependent protein kinases: Understanding functional diversification and specificity of plant signalling hubs using the most conserved members

**Gardette Raquel Valmonte**

A thesis submitted to
Auckland University of Technology
in fulfilment of the requirements for the degree of
Doctor of Philosophy (PhD)

2016
School of Science
Supervisors:
Colleen Higgins, PhD (Lond.)
Robin MacDiarmid, PhD (Auck.)

# Abstract

Agricultural productivity and food security are declining globally, because of factors such as climate change, crop disease, natural calamities, population growth and pollution. Research on how to improve plant stress tolerance, disease resistance and crop productivity is paramount in preparing for future agro-environmental difficulties brought about by a changing world.

Calcium-dependent protein kinases (CPKs) are plant proteins that directly bind calcium ions before phosphorylating substrates involved in osmosis, hormone response, stress and pathogen signalling pathways. CPKs are considered as 'hubs' in plant signalling; members of this large multigene family may function redundantly or complementarily to multiple stresses and stimuli. This research project aimed to answer three questions about the functional diversification and specificity of CPKs. Firstly, how did CPKs diversify and what is the most conserved CPK group in plants? Secondly, what is the role of the most conserved CPKs in plant stress and pathogen responses? Lastly, what influences CPK functional specificity?

A comprehensive genome-wide phylogenetic analysis of CPKs from algae to higher plants showed that CPKs diversified in parallel with the transition of plants into terrestrial life, possibly providing support to plants in response to the stress of this transition; and that the most conserved members of this gene family in plants are those that belong to Group IIb. In Arabidopsis, CPKs that belong to this group are AtCPK3, 17 and 34.

AtCPK3 and its orthologues (Group IIb.1) in rice and kiwifruit change in transcript accumulation in response to most abiotic stresses and pathogens such as *Botrytis cinerea, Pseudomonas syringae,* and various plant viruses, as inferred from meta-analysis of publicly available transcript data and as validated from biological experiments carried out in this project. Knocking out or overexpressing AtCPK3 in Arabidopsis and AcCPK16 in kiwifruit appeared to change the way the limited number of experimental plants respond to stress and pathogens. In Arabidopsis, overexpressors were slightly more tolerant to drought, bacterial, fungal and viral infections, whereas knockouts had little difference or were slightly more susceptible to WT. In kiwifruit, overexpressors were slightly more tolerant to drought and more susceptible to fungal infections, whereas knockouts had little difference or were slightly more susceptible to WT.

AtCPK17 and 34 and their orthologues in rice and kiwifruit (Group IIb.2) were only expressed in floral tissue and mainly function in pollen development. Gene structure and predicted protein structure analysis of Group IIb CPKs in Arabidopsis and rice identified promoter regions

and several protein motifs correlated to CPK function. Seed and pollen germination assays showed some degree of similarity in responses among AtCPK3 and AtCPK34 single overexpressors, suggesting that tissue localisation influences CPK gene function. Gene structure, several protein motifs and tissue localisation, all contribute to CPK functional specificity, which may explain why CPK functions are usually redundant and overlapping making them useful as plant signalling hubs.

This project provides new insights and hypotheses with regards the evolution of CPKs and recommends further research with regards the use of group IIb.1 CPKs for novel molecular and diagnostic approaches in managing plant abiotic and biotic stress across a broad range of plant species.

# Table of Contents

# Attestation of Authorship

I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person (except where explicitly defined in the acknowledgements), nor material which to a substantial extent has been submitted for the award of any other degree or diploma of a university or other institution of higher learning.

Gardette Raquel Valmonte

# List of Figures

14

# List of Tables

# List of Abbreviations

| | |
|---|---|
| aa | Amino acid |
| ABA | Abscisic acid |
| ABF4 | ABA-responsive element binding factor |
| AJ | Autoinhibitory junction |
| At | *Arabidopsis thaliana* |
| AUT | Auckland University of Technology |
| Bc | *Botrytis cinerea* |
| BEAST | Bayesian evolutionary analysis sampling trees |
| BLAST | Basic local alignment search tool |
| CAD | Calmodulin-like domain, also called Ca2+-activation domain |
| CAF | CCR4-associated factor |
| CaM | Calmodulin |
| CaMK | Ca2+/CAM-dependent protein kinase |
| CaML | Calmodulin-like proteins |
| CaMV | Cauliflower mosaic virus |
| CBL | Calcineurin B-like |
| CCaMK | Ca2+/CAM-dependent protein kinase |
| cDNA | Copy DNA |
| CFU | Colony-forming units |
| CIPK | CBL-interacting protein kinase |
| Cp | *Cryptosporidium parvum* |
| CPK | Calcium-dependent protein kinases |
| Cre | *Chlamydomonas reinhardtii* |
| CT | C-terminus |
| CTAB | Cetyltrimethylammonium bromide |
| CymMV | Cymbidium mosaic virus |
| DNA | Deoxyribonucleic acid |
| DREB | Drought response element binding |
| EBI | European bioinformatics institute |
| EF-1α | Elongation factor- 1 α |
| ELISA | Enzyme-linked immunosorbent assays |
| ERF | Ethylene response factor |
| EST | Expressed sequence tags |
| ETI | Effector-triggered immunity |
| ETS | Effector-triggered susceptibility |
| F-BOX | F-Box family protein |
| GA3 | Gibberellic acid |
| gDNA | Genomic DNA |
| GEO | Gene expression omnibus |
| GFP | Green fluorescent protein |
| Gm | *Glycine max* |
| GSR | General stress response |
| HR | Hypersensitive response |
| ICE-CBF/DREB1A | Inducer of CBF Expressions- Dehydration responsive element-binding factor 1a |
| ICMP | International Collection of Microorganisms from Plants |
| I-TASSER | Iterative Threading Assembly Refinement |
| JA | Jasmonic acid |
| Ka | Nonsynonymous nucleotide substitution |
| KEGG | Kyoto Encyclopedia of Genes and Genomes |

| | |
|---|---|
| KO | Knock-out |
| Ks | Synonymous nucleotide substitution |
| LB | Luria-Bertani |
| LEA | Late-embryogenesis abundant |
| MAMP | Microbial- or pathogen-associated molecular patterns |
| MAPK | Mitogen activated protein kinase |
| MeJA | Methyl jasmonate |
| MEME | Maximisation for Motif Elicitation |
| ML | Maximum likelihood |
| MRCA | Most common recent ancestors |
| mRNA | Messenger RNA |
| MS | Mass spectrometry |
| MS | Murashige and Skoog |
| MUSCLE | Multiple Sequence Comparison by Log-Expectation |
| MYA | Million years ago |
| NASC | Nottingham Arabidopsis Stock Centre |
| NB-LRR | Nucleotide-binding leucine-rich repeat |
| NCBI | National Center for Biotechnology Information |
| NGS | Next generation sequencing |
| NJ | Neighbour-joining |
| NMR | Nuclear magnetic resonance |
| NO | Nitric oxide |
| N-VD | N-terminal variable domain |
| Os | *Oryza sativa* |
| OX | Overexpressor |
| PAMP | Pathogen-associated molecular patterns |
| PC1 | Physical containment level 1 |
| PC2 | Physical containment level 2 |
| PCR | Polymerase chain reaction |
| PDA | Potato dextrose agar |
| PDF | Plant defensin gene |
| PDF2 | Protodermal factor 2 |
| Pf | *Plasmodium falciparum* |
| PFR | The New Zealand Institute for Plant and Food Research (PFR) |
| PK | Protein kinase |
| POPTR | *Populus trichocarpa* |
| Pp | *Physcomitrella patens* |
| PPV | Plum pox virus |
| PR | Pathogenesis-related |
| PRR | Pattern recognition receptors |
| Pss | *Pseudomonas syringae pv. syringae* |
| PTI | Pathogen-triggered immunity |
| Pto DC3000 or Pst DC3000 | *Pseudomonas syringae pv. Tomato* DC3000 |
| qPCR | Quantitative polymerase chain reaction |
| R genes | Resistance genes |
| RIN | RNA integrity |
| RNA | Ribonucleic acid |
| RNAi | RNA interference |
| ROS | Reactive oxygen species |
| RT | Reverse transcriptase |
| RT-PCR | Reverse-transcriptase polymerase chain reaction |
| RT-qPCR | Reverse-transcriptase quantitative polymerase chain reaction |
| RWR | Rapid wound responsive |
| SA | Salicylic acid |

| | |
|---|---|
| SAND | SAND family protein |
| SAR | Systemic acquired resistance |
| Sb | *Sorghum bicolour* |
| SD | Standard deviation |
| SDS-PAGE | Sodium dodecyl sulfate polyacrylamide gel electrophoresis |
| SE | Standard error |
| SnRK | SNF1-related kinase |
| SOS | Salt overly sensitive |
| Ta | *Triticum aestivum* |
| TAIR | The Arabidopsis Information Resource |
| tBLASTn | Translated nucleotide BLAST |
| TBS-T | Tris-buffered saline-tween |
| TDI | T-DNA insertion |
| T-DNA | Transfer DNA |
| TF | Transcription factor |
| Tg | *Toxoplasma gondii* |
| TMV | Tobacco mosaic virus |
| TSWV | Tomato spotted wilt virus |
| TuMV | Turnip mosaic virus |
| TYMV | Turnip yellow mosaic virus |
| UPGMA | Unweighted pair group method with arithmetic |
| Vv | *Vitis vinifera* |
| WAG | Whelan and Goldman |
| WGD | Whole genome duplication |
| WT | Wild-type |
| Zm | *Zea mays* |

# Acknowledgements

The past years of working on this project as I pursue a PhD degree has been filled with fun and laughter as well as challenges and tears. I had a wonderful opportunity to work alongside awesome teams and groups of people who selflessly helped, encouraged, and inspired me throughout this journey. I have to say that it was teamwork that brought me towards the completion of this project. No words will be enough to express my gratitude to all the people who helped me and have been part of my life during my PhD years. I dedicate this piece of work to all of you.

Firstly, I would like to thank my amazing supervisors, Dr Colleen Higgins and Dr Robin MacDiarmid. Thank you so much for happily accepting me to do this project, for guiding me through and through, and for pushing me towards the end especially during the times that I felt like giving up. Thank you for picking me up when I made mistakes, for honing my skills and intellect and especially for lifting up my spirit in good days and in bad days. Thank you so much for all your hard work reading and editing my thesis, for organising things for me and for giving me great insights. I will always treasure your advice and all the things you have imparted to me. I have learned a lot from you --- things that are so valuable not only in my career but also in my life. I will cherish them forever. Thank you for opening doors for me; for your unwavering support, patience and love. I am indebted to you so much that I cannot express how thankful I am to have you as my supervisors, my friends and my 'second' parents.

My deepest gratitude and love to all the teams I have been part of: the Molecular Genetics Research Team at AUT, Joe Chang, Subuhi Khan, Smriti Nair, Kate Olliver, Chris Puliuvea, Priya Ajithkumar and everyone; the Virus and like Organisms Team and Bioprotection Technologies group at PFR, Dr Kieren Arthur, Dr Tracey Immanuel, Dr Sonia Lilly, Sam Edwards, Dr Karmun Choi, Arnaud Blouin, Dr Dan Cohen and everyone; the Plant Pathology group at the University of Auckland and AUT School of Science teaching and laboratory team. Thank you so much for your all your help, support, suggestions, and friendship. You all made a great impact to my project, career and life; and for that I will always be thankful.

Much gratitude goes to Kieren and Tracey for their support and advice throughout this project. Huge thanks to Sonia for sharing with me the cDNA material from her virus inoculation

# Chapter One

# General Introduction

## 1.1 Rationale and Significance of the study

Agricultural productivity and food security is currently declining globally as a result of a range of environmental factors and human activity, such as climate change, natural calamities, population growth and pollution (Dar and Laxmipathi Gowda 2013; Shao et al. 2007). Our world is approaching critical stages of health quality, eco-environment stability, and food supply (Shao et al. 2007). It has been projected that within fifty to one hundred years, average temperatures will rise dramatically, which further threatens crop yield (Dar and Laxmipathi Gowda 2013). Research on how to improve crop productivity, stress tolerance and disease resistance is therefore highly important in addressing the current problems of food security and preparing for future agro-environmental difficulties brought about by a changing world. Molecular diagnosis of plant disease and molecular approaches to managing plant abiotic and biotic stress are useful aspects of such promising fields of research.

How plants respond to stress such as environmental pressures, pathogens, insects, symbionts and other stimuli is currently of great interest in the plant sciences and agricultural research due to the impacts these stresses have on crop productivity and ecosystem sustainability (Gust et al. 2010; Qin et al. 2011; Rodriguez et al. 2005). Stress is defined as an external factor that exerts a disadvantageous influence on the plant. Stress can be abiotic or biotic; abiotic stresses include drought, extreme temperatures, high salinity and nutrient starvation; while biotic stresses include herbivory, bacterial, viral and fungal infections. Due to a sessile lifestyle, plants have limited capacity to avoid these unfavourable conditions (Galindo et al. 2007). Plant genomes have adapted to overcome environmental challenges and many plant genes are dedicated to stress protective mechanisms (Gechev and Hille 2012). Much research is being done to look for and characterise genes responsible for environmental stress tolerance and disease resistance in plants with the aim of understanding plant stress responses and discovering novel approaches to improve plant stress tolerance.

Plants have sophisticated molecular chemical strategies to defend themselves against abiotic and biotic stress (Galindo et al. 2007). Environmental stress and pathogen responses in plants involve an intricate relay of intercellular and intracellular signals. As an adaptation, plants have specialised signal transduction pathways in response to adverse conditions, infection and insect attack (Iriti and Faoro 2007; Jones and Dangl 2006; Kaur and Gupta 2005; Qin et al. 2011). These pathways lead to cellular responses, which include changes in cell cycle and division, cell membranes, cell wall architecture, and metabolism (Galindo et al. 2007). This then culminates in physiological changes such as hormone secretion, stomatal closure, root tip movement and localised cell death (Boudsocq and Sheen 2013).

Calcium ($Ca^{2+}$) signalling is one of the sophisticated signalling networks that play fundamental roles in growth, development, and stress response, not only in plants but in all eukaryotes. Cytosolic $Ca^{2+}$ levels elevate in complex spatio-temporal patterns (called '$Ca^{2+}$ signatures') in response to various developmental and stress stimuli, leading to specific cellular responses including stomatal movement, increased water retention, microbial detection, and tip structure movement (DeFalco et al. 2010; Guex et al. 2009; Hashimoto and Kudla 2011). In plants, $Ca^{2+}$ signatures are detected by a vast array of $Ca^{2+}$-sensors and responder proteins, such as calmodulins (CaMs), calmodulin-like proteins (CaMLs), $Ca^{2+}$/CAM-dependent protein kinases (CaMKs), $Ca^{2+}$ and $Ca^{2+}$/CAM-dependent protein kinases (CCaMKs), Calcineurin B-like proteins (CBLs), CBL-interacting protein kinases (CIPKs), and calcium-dependent protein kinases (CPKs). These proteins undergo conformational changes upon binding with $Ca^{2+}$ and as a consequence transmit specific signals to their substrates through phosphorylation (DeFalco et al. 2010; Guex et al. 2009; Kiefer et al. 2009). Among proteins involved in calcium signalling, CPKs are unique because they have both a $Ca^{2+}$ sensor domain that directly binds $Ca^{2+}$ ions, and a responder (protein kinase) domain that phosphorylates specific protein targets. In contrast, other calcium signalling proteins only have a single function; either a sensor or a responder. These latter proteins therefore require the extra step of finding their corresponding sensor/responder protein in order to direct signalling; while in CPKs the calcium sensing and kinase functions occur in tandem (DeFalco et al. 2010).

Despite the increasing evidence on the importance of CPKs in various plant stress responses, a comprehensive genome-wide analysis of CPKs from algae to higher plants that focuses on CPK evolution in land plants and its applications to agriculture and ecology has not

yet been undertaken (Figure 1.1). CPKs are present in protists (Billker et al. 2009), oomycetes (as detected through Basic Local Alignment Search Tool (BLAST)), Broad Institute of Harvard and MIT 2010), green algae (Baillie et al. 2000; McCurdy and Harmon 1992) and plants (Anil et al. 2003; Asano et al. 2010; Kiselev et al. 2010; Li et al. 2008b), but are not found in animals or fungi. At the time when this thesis project was started, comparative genome-wide phylogenetic analyses of CPKs and their closely-related gene families had so far been described only in protists (Billker et al. 2009; Nagamune and Sibley 2006) and a small number of model plants: *Arabidopsis thaliana* (thale cress or Arabidopsis) (Cheng et al. 2002; Hrabak et al. 2003)*, Oryza sativa* (rice) (Asano et al. 2011), and *Triticum aestivum* (wheat) (Li et al. 2008a). CPKs are a large multigene family divided into four major evolutionary groups; with 34 isoforms in the model dicotyledon plant, Arabidopsis, and 31 isoforms in the model monocotyledon plant, rice (Asano et al. 2005; Cheng et al. 2002; Hrabak et al. 2003). A phylogenetic analysis of CPKs from protists through to plants was reported over a decade ago (Zhang and Choi 2001) but this was limited by the CPK sequences available at that time, with only 22 CPKs from Arabidopsis, one each from maize and liverwort, and four from protists. Wide-range analyses of CPKs and their relatives (belonging to the same superfamily, called the CPK-SnRK superfamily) in protists, algae and plants have provided good representation of their phylogeny, but were also limited in the number of CPK isoforms to represent the genome of each species included in the analysis (Harmon et al. 2000; Hrabak et al. 2003). With the recent developments in sequencing and annotation of several plant genomes, CPKs from a number of complete and nearly completed genomes can be mined to perform a broadly sampled phylogenetic analysis. At the beginning of this project, there were no reports of a genome-wide analysis of CPKs from lower to higher plants. Therefore, one of the aims of this research was to fill this gap. This also led to the identification of potential CPK orthologous groups, which is important in predicting CPK function.

The function and specificity of genes can be inferred by analysing different factors, such as protein structures and gene regulatory regions. The three dimensional (3D) protein structures of some protist CPKs have been described recently (Wernimont et al. 2011; Wernimont et al. 2010); however, the full length structure of a plant CPK protein has not yet been elucidated (Figure 1.1). Structural studies have been performed in *Arabidopsis* and *Glycine max* (soybean); however, their 3D structures did not encompass the whole protein (Christodoulou et al. 2004; Weljie et al. 2000; Weljie et al. 2004; Weljie et al. 2003; Weljie and Vogel 2004). The structure of

a plant CPK that includes the protein kinase domain could provide more information for predicting its function, substrate specificity and/or explaining its molecular activity in response to certain conditions. Furthermore, there have been no comprehensive studies reporting on the gene regulatory regions of CPKs. Currently, there are computer programs available to predict tertiary and/or quaternary protein structures and gene promoter regions making it possible to make inferences and hypotheses about molecular responses (e.g. transcriptional activation) and specificity (e.g. binding proteins or phosphorylation substrates) (Sadowski and Jones 2009; Shin et al. 2007; Whisstock and Lesk 2003). Thus, further aims of this project were to utilise these bioinformatics tools in predicting or understanding CPK functional diversification and specificity through analysis of the predicted 3D structure and regulatory sequences.

Research on CPKs could be of great benefit to agriculture, as these are highly important genes in plant stress signalling. CPKs are possible gene targets for novel approaches to improved plant stress tolerance, particularly those that respond to various types of stresses, and are conserved in a wide range of plant species. Understanding how CPK functions have diversified through evolution and what determines their functional specificity are important in how these genes can be applied to molecular approaches in managing plant stress and infections. A summary of the key evolutionary, functional and structural studies on CPKs including this research project is shown in Figure 1.1.

**Evolutionary and functional studies**

| ✓ | Unrooted, non-genome-wide phylogenetic analyses of the CPK superfamily from various algae, plants and protists | Harmon, et al. (2000 and 2001) |

| ✓ | Molecular evolution of CPKs in very few plants and protists (not genome-wide) | Zhang and Choi (2001) |

| ✓ | Genome-wide identification, phylogenetic analysis and functional review of CPKs from the model dicot, Arabidopsis | Cheng, et al. (2002) |

| ✓ | Genome-wide identification and phylogenetic analysis of CPKs and closely-related gene families from the model monocot, rice | Asano, et al. (2005) |

| ✓ | Comparative genomic and phylogenetic analyses of calcium-regulated proteins in apicomplexan protists | Nagamune and Sibley (2006) |

| ✓ | Evolutionary and functional study of the CPK gene family in another monocot, wheat | Li, et al. (2008) |

| ✓ | Phylogenetic analysis of CPKs in apicomplexans (genome-wide) and few plants (non-genome-wide) | Billker, et al. (2009) |

| ✳ | Broadly-sampled, genome-wide phylogenetic analysis of CPKs from lower to higher plants, using protist CPKs as outgroups | This project |

| ✳ | Identification of the most conserved CPK in plants | This project |

**Structural studies**

| ✓ | Comparative modeling studies of the calmodulin-like domain of soybean CPKs | Weljie, et al. (2000) |

| ✓ | Conformational changes in the Ca2+-regulatory region of soybean CPKα studied | Weljie, et al. (2003) |

| ✓ | Evidence for differing roles for each lobe of the calmodulin-like domain in an Arabidopsis CPK | Christodoulou et al. (2004) |

| ✓ | Solution structure and backbone dynamics of the N-terminal region of the calcium regulatory domain from soybean CPKα | Weljie, et al. (2004) |

| ✓ | Structure of the regulatory apparatus of an Arabidopsis CPK | Chandran, et al. (2006) |

| ✓ | Structures of apicomplexan calcium-dependent protein kinases reveal mechanism of activation by calcium | Wernimont, et al. (2010 and 2011) |

| ✳ | 3D structure prediction of the most conserved CPK in Arabidopsis (first *In silico* 3D structure of a full plant CPK sequence) | This project |

| ✳ | Functional characterisation of the most conserved CPK in Arabidopsis, rice and kiwifruit | This project |

**Figure 1.1. Flow chart of key evolutionary, functional and structural studies on CPKs that led to the goals of this research project.** Ticked boxes show previous studies reported by different authors/research groups. Green circles with asterisk show analyses that had not yet been reported and were the aims of this PhD study.

**1.2 Statement of the problem and aims of the study**

In order to advance our knowledge of the evolution, structure and function of CPKs, the aims of this research project were to: 1) conduct a broadly-sampled, genome-wide evolutionary analysis of CPKs using sequences from databases of complete (or nearly completed) plant genomes; 2) identify the most conserved CPK in plants; 3) predict their 3D structure; and 4) understand its function in the model plant *Arabidopsis* and in two economically important crop plants that represent monocotyledons and dicotyledons (rice and kiwifruit). The most conserved member(s) of this gene family would serve as a good representative for structural and functional studies of CPKs because the sequence would have been maintained with little change through evolution and, thus, may have a greater degree of similarity with other CPK homologues within and between different plant species. The most highly conserved CPK (s) may therefore have an essential function that is conserved in a wide range of species. Moreover, analysing the structure and function of the most conserved CPK can provide insights into the fundamental function of CPKs in plant signal transduction, particularly in stress response and developmental pathways. This information may also lead to potential targets for molecular markers in selectively breeding crops with increased stress tolerance. Such markers are of great importance to commercial and natural ecosystems, as they provide a tool to measure how well plants are coping with their environment. These tools will become increasingly important as the effects of climate change become more apparent.

This PhD project had three primary aims, each with a number of specific objectives:

**Aim 1: To determine how CPKs diversified from lower to higher plants and identify the most conserved CPK group in plants**

Objective 1.1    Carry out data mining of all available CPK sequences from the genome of representative lower and higher plant species

Objective 1.2    Carry out a phylogenetic analysis of the CPKs identified in Objective 1.1 and identify the most conserved members

Objective 1.3    Determine the correlation between CPK sequence evolution and CPK functional diversification from lower to higher plants

**Aim 2: To determine the function of the most conserved CPK in response to abiotic and biotic stresses**

> Objective 2.1    Determine the transcript accumulation of the most conserved CPK in Arabidopsis plants in response to abiotic stresses including drought and high salinity and biotic stresses including specific bacterial, viral and fungal infections
>
> Objective 2.2    Determine the physical status and stress responsiveness of Arabidopsis plants when the expression of the most conserved CPK gene is ablated (knocked-out) or increased (overexpressed)
>
> Objective 2.3    Establish whether this function is conserved amongst important crops belonging to plant families different to Arabidopsis: rice *(O. sativa)* and kiwifruit (*Actinidia chinensis*)

**Aim 3: To determine if protein structure, gene structure and/or tissue localisation correlate with functional specificity of CPKs in the most conserved group**

> Objective 3.1    To determine motifs in the CPK protein sequences that correlate with evolutionary grouping and specific biological function
>
> Objective 3.2    To analyse predicted CPK tertiary structure particularly within the motifs that correlate with specific biological function
>
> Objective 3.3    To determine CPK gene structure and regulatory regions that correlate with specific biological function
>
> Objective 3.4    To determine if the CPK genes' tissue localisation correlate with functional specificity

**1.3 Synopsis of the Study Design**

The following sections provide a summary of methods carried out to address the aims and objectives, and the chapters where these are presented:

***Aim 1: To determine how CPKs diversified from lower to higher plants and identify the most conserved CPK group in plants***

Experiments, results and inferences regarding Aim 1 comprise **Chapter 3** of the thesis: "*How did CPKs diversify and what is the most conserved CPK group in plants?*"

***Objective 1.1  Carry out data mining of all available CPK sequences from the genome of representative lower and higher plant species***

In order to analyse CPK sequence evolution and diversity among lower and higher plants, all known CPKs from fifteen selected algae and plant genomes were identified. Representative CPK sequences from the model plant Arabidopsis were used as query terms to perform protein BLAST (Altschul et al. 1990) and translated nucleotide BLAST (tBLASTn) searches to identify all (or most) CPK homologues in the genome of selected plant species. Selected plant genomes included green algae (*Volvox carteri* and *Chlamydomonas reinhardtii*), bryophyte moss (*Physcomitrella patens*), pteridophyte fern (*Selaginella moellendorfii*), gymnosperm (*Picea sitchensis*), monocotyledons (*O. sativa, T. aestivum, Sorghum bicolour* and *Zea mays*) and eudicotyledons (*Vitis vinifera, G. max, Populus trichocarpa, Carica papaya*, and *Solanum tuberosum*). Identified sequences were downloaded from the plant genome databases GreenPhylDB (Conte et al. 2008) and Phytozome (Sheen 1996). Each sequence was examined to validate if it was a true CPK and if all the introns were properly removed from the sequence.

***Objective 1.2  Carry out a phylogenetic analysis of the CPKs identified in 1.1 and identify the most conserved members***

A phylogenetic analysis was carried out to explore the evolutionary history of CPKs and to identify the most conserved CPK in plants. Multiple sequence alignment and phylogenetic analysis were carried out using the software GeneiousPro 5.6 (Kearse et al. 2012). A total of 352 plant CPK protein sequences were aligned using the ClustalW program (Larkin et al. 2007) in GeneiousPro 5.6. Five protist CPKs consisting of TgCPK1 and TgCPK3 from *Toxoplasma gondii*, PfCPK3 from *Plasmodium falciparum* and CpCPK1 and CpCPK3 from *Cryptosporidium parvum*

were included in the alignment and were used as outgroups. Distance (Jukes-Cantor model) and likelihood (Whelan and Goldman [WAG] model) trees were constructed using the neighbour-joining (NJ) and maximum likelihood (PHYML) methods, respectively, with 1000 bootstrap replicates. The phylogenetic tree represents the evolutionary history of CPKs based on existing nucleotide and protein sequence data, from lower to higher plants. The most conserved CPK group, with the shortest branch lengths on average, were then identified based on this tree.

### Objective 1.3   Determine the correlation between CPK sequence evolution and CPK functional diversification from lower to higher plants

To explore the functional importance of CPK gene expansion and diversification events in plant evolution, an extensive literature review and expression profile examination was undertaken. Literature that reported CPK function in development, stress responses and other stimuli were collected and the corresponding CPK sequence in each report downloaded. An NJ tree that includes all the sequences of CPKs with reported function was constructed to illustrate any correlation between sequence relationships of the CPKs with similar function.

### Aim 2:    To determine the function of the most conserved CPK in response to abiotic and biotic stresses

Experiments, results and inferences regarding Aim 2 comprise **Chapter 4** of the thesis: "*What is the role of the most conserved CPKs in plant stress and pathogen responses?*"

### Objective 2.1   Determine the transcript accumulation of the most conserved CPK in Arabidopsis plants in response to abiotic stresses such as drought and high salinity and biotic stresses such as bacterial, viral and fungal infections

*In silico* and *in planta* approaches were carried out to determine the transcript accumulation of Arabidopsis CPKs (AtCPKs) that belong to the most conserved group. Since AtCPK 17 and 34 are not found in plant tissues other than the floral tissue (specifically in pollen), only AtCPK3 was analysed using the *in planta* approaches. For the *in silico* approach, transcript accumulation levels of Arabidopsis CPKs were analysed using Affymetrix 22K microarray data available in The Arabidopsis Information Resource (TAIR) (Swarbreck et al. 2008) and in the online platform Genevestigator V3 (https://www.genevestigator.com/gv/index.jsp) (Hruz et al.

2008). Data from the literature reporting microarray and/or quantitative Polymerase Chain Reaction (PCR) result were also noted, involving CPK responses to various stimuli including hormones, developmental signals, abiotic and biotic stresses.

For the *in planta* approach, the transcript accumulation of the most conserved CPK in Arabidopsis in response to desiccation, high salinity and pathogen treatments were measured using reverse transcriptase-quantitative polymerase chain reaction (RT-qPCR). Triplicate samples for each plant species were exposed to either an environmental stress or control treatment. Leaf tissue samples (and root tissue for some treatments) were collected at certain time points: for salinity, at 0, 15, 30 min, 1 h, 4 h, 12 h, 24 h, and 48 h; for drought, at 0, 7 and 14 days, and for pathogen infections, at 0, 2, 3, 7, 14, and 21 days post inoculation (dpi). Phenotypes such as symptom severity, development (plant height), number of leaves, leaf rosette diameter and dry weight were measured. RT-qPCR experiments were performed conforming to the standards set by the Minimum Information for Publication of Quantitative Real-Time PCR Experiments (MIQE) guidelines (Bustin et al., 2010; Bustin et al., 2009).

### Objective 2.2   Determine the physical status and stress responsiveness of Arabidopsis plants when the expression of the most conserved CPK gene is ablated (knocked-out) or increased (overexpressed)

To determine the role of the most conserved CPK in plant stress responses, reverse genetics and overexpression approaches were employed. Arabidopsis lines with the AtCPK3 and AtCPK34 genes knocked out by T-DNA insertion were obtained from the Nottingham Arabidopsis Stock Centre (NASC). Arabidopsis plants that constitutively overexpress the native form of the most conserved CPK were also created through *Agrobacterium tumefaciens*- mediated plant transformation. The stress responses investigated in Objective 2.1 were also tested using these loss-of-function mutants and overexpressing plants.

### Objective 2.3   Establish whether this function is conserved amongst important crops belonging to different families: rice (O. sativa) and kiwifruit (A. chinensis)

To find out if the functions observed in Arabidopsis (a rosid dicotyledon) is conserved in monocotyledons and other dicotyledon families, functional analyses were performed in rice (a

monocotyledon belonging to the grass family) and kiwifruit (a dicotyledon belonging to the asterid clade). The CPKs in rice that were most similar to the most conserved Arabidopsis CPK were identified from the phylogenetic tree constructed in Objective 1.2. In a separate collaborative study, the full array of CPK gene family members in kiwifruit were identified from the nearly assembled *Actinidia* genome available at The New Zealand Institute for Plant and Food Research (PFR). Generic primers for CPKs were designed to physically isolate and sequence CPKs from kiwifruit. To find out whether the stress-responsiveness observed in Arabidopsis is conserved in rice and kiwifruit, the transcript accumulation of the most conserved CPK orthologues in these crops in response to desiccation, high salinity and pathogen treatments was measured using RT-qPCR. In kiwifruit, the most conserved CPK gene was rendered non-functional using RNA interference approaches (RNAi). These knock out plants were then challenged with abiotic and biotic stresses and the responses were compared with wild-type (WT) plants.

***Aim 3: To determine if protein structure, gene structure and/or tissue localisation correlate with Group IIB CPK functional specificity***

Experiments, results and inferences regarding Aim 3 comprise **Chapter 5** of the thesis: *"What influences CPK functional specificity?"*

> ***Objective 3.1   To determine motifs in the CPK protein sequences that correlate with evolutionary grouping and specific biological function***

To determine group-specific motifs, a motif analysis of the 34 AtCPKs was carried out using the publicly available software, Multiple Expectation Maximisation for Motif Elicitation (MEME) v4.7.0 (http://meme.sdsc.edu/meme/cgi-bin/meme.cgi). Peptide sequence patterns that are unique to each of the four evolutionary CPK groups in Arabidopsis were identified. To determine function-specific motifs, CPK sequences with reported function were analysed by visual observation to identify response-specific amino acid (aa) sequence patterns. The alignments, consensus sequence and sequence logos were constructed using GeneiousPro 5.6. In particular, developmental and stress response-specific motifs were compared between members of the most conserved CPK group.

***Objective 3.2 To analyse predicted CPK tertiary structure and substrate specificities particularly within the motifs that correlate with specific biological function***

To determine if protein tertiary structures contribute to CPK function specificity, 3D structures of AtCPKs that belong to the most conserved group (AtCPK3, 17 and 34) were predicted. 3D protein structure predictions were done through publicly available software, Swiss-Model (http://swissmodel.expasy.org/) (Bordoli et al. 2009) and I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/)(Sheen 1996). 3D structures were constructed using a model-based approach, using as models the published 3D structures of CPKs coming from the apicomplexan parasites *Toxoplasma gondii* and *Cryptosporidium parvum*, namely TgCPK1, CpCPK1 and TgCPK3 (Wernimont et al, 2010). The available soybean and Arabidopsis CPK structures (incomplete structures) were also attempted to use as models (Chandran et al. 2006). The regions that contain the group-specific and function specific motifs identified in Objective 3.1 were analysed and compared.

***Objective 3.3 To determine CPK gene structure and regulatory regions that correlate with specific biological function***

To explore gene structure evolution and potentially determine highly conserved gene regions within and between groups, an intron-exon analysis was performed. The intron/exon organisation 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 sequence of these CPKs were obtained from Phytozome (http://www.phytozome.net/) (Sheen, 1996).

To determine gene regulatory regions that may influence function, promoter analysis of the most conserved CPKs in Arabidopsis and rice was carried out. Promoter analysis based on transcription factor binding sites was done using the software MatInspector (Quandt et al. 1995). The identified sites were compared between developmental and stress-response CPKs in the most conserved group.

***Objective 3.4   To determine if the CPK genes' tissue localisation correlate with functional specificity.***

To determine any potential correlation between tissue localisation and gene function, the reported functions of AtCPKs in the most conserved group was compared. AtCPK3, 17 and 34, which all belong to this group, have different biological functions: AtCPK3 primarily responds to abiotic stress and some biotic stresses, while AtCPK17 and 34 were developmental regulators and were pollen-specific. Seed germination assays and pollen germination assays involving knockouts and overexpressors of the genes AtCPK3 and 34 were carried out.

# Chapter Two

# Literature Review

This review starts with a brief introduction of the molecular tools and techniques that are used in understanding function of genes and/or proteins. To further justify the rationale of this project, this review then describes what is currently known about plant stress responses, $Ca^{2+}$-dependent protein kinases and their involvement in plant responses to abiotic stresses and pathogen infection. Finally, discussed is the possible benefits to agriculture and the plant sciences of understanding the function of the most conserved members of the plant CPK gene family.

## 2.1 Understanding Gene Function

To understand the function of genes and proteins, molecular biologists utilise several tools that mainly fall under two categories: biological approaches and bioinformatic approaches (Figure 2.1). The central dogma of molecular biology states that genes (DNA) control cellular processes by encoding proteins, using mRNA as an intermediate. It is however currently well-established that the environment has a great impact on this process; various stimuli can influence each step of the expression and activity of a certain gene (Brooker et al. 2007). Functional studies use the known sequence of a gene or protein and utilise molecular biology techniques to analyse gene transcription, translation and molecular activities in various environmental conditions.
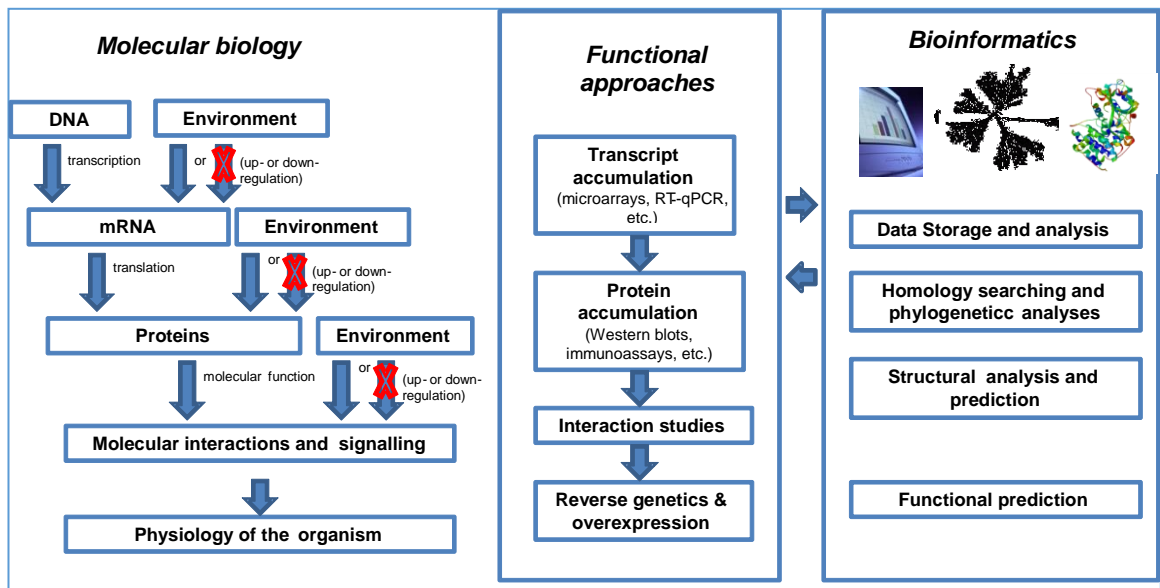
**Figure 2.1. An outline of the central dogma of molecular biology functional and bioinformatic approaches in understanding gene function.** Blue arrows with an overlapping red 'X' mark indicates down-regulation.

### 2.1.1 Molecular approaches to understanding gene function

There are various molecular tools and approaches to analyse the function of genes. Each technique is used for a specific purpose; however, the data that result from one technique must be validated using another technique in order to obtain strong experimental evidence to establish the molecular function of a certain gene.

### 2.1.1.1 Quantitative analyses of gene transcript and protein

Transcript and protein accumulation are commonly used as indicators of upregulation and downregulation of genes/proteins in a certain type of tissue, developmental stage or stimulus response. This is because the expression of a gene and/or activity of a protein may be turned on or off in response to single or multiple stimuli. Transcript accumulation is detected using hybridisation-based and PCR-based molecular techniques, such as northern hybridisation (RNA gel blots), *in situ* hybridisation (use of labelled complementary RNA strand/probe to find the localisation of a specific sequence in a tissue), reverse transcriptase PCR (RT-PCR) (use of the enzyme reverse transcriptase to convert mRNA to its DNA complement, followed by amplification using PCR), quantitative RT-PCR (RT-qPCR) (use of fluorescent dyes to accurately measure the relative number of mRNA copies between two biological samples), and microarray analysis (as for RT-qPCR but uses thousands of shorter nucleotide sequences in a 'gene chip'). Protein accumulation of a specific gene product is detected using immunoassays (for which a specific

antibody is required) such as enzyme-linked immunosorbent assays (ELISA) and western blot analyses and other molecular techniques such as two-dimensional polyacrylamide gel electrophoresis (2D-PAGE), protein arrays and mass spectrometry (MS) (Graves and Haystead 2002).

### 2.1.1.2 Analysing gene function using transgenic plants

The biological function of a gene and its products, including its impact on the physiology or morphology of the organism, can be determined using transgenic approaches. These approaches are very powerful in elucidating gene function as the roles are verified *in vivo* (or *in planta* in the case of plants). Transgenesis may involve gene overexpression or reverse genetics. Gene overexpression involves the introduction of a particular gene and/or promoter to increase the level of gene expression while reverse genetics involves insertional mutagenesis or gene silencing to knock-out or knock-down the expression of a gene. Both approaches determine whether a biological function is enhanced or diminished by the change in a particular gene expression level.

Production of transgenic plants has four essential components: suitable vectors, selectable genetic markers, plant transformation and plant regeneration (Kung and Wu 1993). A particular gene of interest is inserted into gene transfer vectors, which are usually plasmids that contain selectable markers and constitutive promoters for both the gene of interest and the marker. The recombinant vectors containing the gene of interest are propagated, commonly using bacteria (e.g. specialised strains of *Escherichia coli*). The recombinant DNA can be transferred into plants using various technologies, most commonly by infecting plant tissue with recombinant *Agrobacterium tumefaciens* (Agrobacterium) or by introducing DNA directly into plant cells by particle bombardment-mediated transformation (Finer and Taniya 2008).

Agrobacterium is the causative agent of crown-gall disease in plants, which has a unique ability to transfer part of its DNA into the plant genome (Tzfira et al. 2004). The transferred DNA is called the T-DNA, carried on an extrachromosomal plasmid called the Ti (tumour-inducing) plasmid (Radhamony et al. 2005). In Agrobacterium-mediated transformation, binary vectors contain the left (L) and right (R) borders of the T-DNA, in which the gene of interest is inserted. Agrobacterium strains have been developed to be avirulent. The Ti plasmid no longer causes tumour formation, but still transfers the T-DNA region. Agrobacterium vectors are called binary

vectors, because they are the second of two plasmids that are required for the overall process. Once a plant tissue is inoculated with Agrobacterium, chemical signals are released by the plant; and if the tissue is wounded it produces acetosyringone which activates the bacterial virulence (vir) genes that then initiate the T-DNA transfer process (Finer and Taniya, 2008). The amount of acetosyringone produced varies between plants, so current Agrobacterium inoculation techniques include the addition of synthetic acetosyringone to enhance transformation efficiency. Agrobacterium transformation is usually done by direct injection or agroinfiltration into leaves of a suitable target plant (e.g. *Nicotiana benthamiana*) or by floral dipping of developing buds (usually done on Arabidopsis). In agroinfiltration, an Agrobacterium suspension is forced into the internal leaf airspace by tightly holding a syringe (without needle) to the leaf and pushing the plunger. The effects of gene transfer can be quickly determined using this procedure; however, the gene introduced is not passed on to the next plant generation (Finer and Taniya, 2008). For vertical transmission of the new gene, whole plants must be transformed either through regeneration of transformed cells through transformation of developing seeds by floral dipping.

A gene of interest can be overexpressed by inserting the gene's open reading frame, or a cDNA of the gene, into the vector along with promoters that direct expression of the gene constitutively or through inducers. On the other hand, a gene of interest can be knocked out (total absence of expression) or knocked down (reduced expression level) through different reverse genetics approaches such as insertional mutagenesis, RNA-mediated interference, virus-induced gene silencing, fast-neutron mutagenesis and chemical mutagenesis (Gilchrist and Haughn, 2010). Each of these approaches has its own advantages and disadvantages. Of these approaches, RNAi is the simplest and has the advantage of being sequence-specific and of not involving another microbe once the transgenic plant has been generated. RNAi involves delivering a gene that generates a double stranded RNA to initiate targeting and degradation of homologous transcripts (Kotak et al. 2007).

### 2.1.2 Bioinformatic approaches to understanding gene function

#### 2.1.2.1 Homology searching and phylogenetic analysis

Using bioinformatic tools, homologues may be identified for sequences with unknown or unresolved function. Homologous sequences often have similar or related functions and thus can provide clues regarding the unknown function of a gene or protein of interest. Homology implies

common ancestry that may be recent or not. Homologous genes can arise through speciation events, in which case they are called 'orthologues' (equivalent genes in another species); or through duplication events, in which case they are called 'paralogues'. Homologous sequences can be aligned and phylogenetic trees can be built to visualise and analyse their evolutionary relationship (Pevsner 2009). Similarity searching is most commonly done using BLAST, available in the National Center for Biotechnology Information (NCBI) web site (Suntio and MÄKinen 2012). Multiple sequence alignments provide information about characters (nucleotide bases or amino acids) and regions (also called motifs or domains) that show similarities and differences. These also provide measures or scores of similarity, which are then used in constructing phylogenetic trees which shows possible homology. Multiple sequence alignments can be done using various programmes such as ClustalX (Larkin et al. 2007) and Multiple Sequence Comparison by Log-Expectation MUSCLE (Edgar 2004).

Phylogenetics investigates the evolutionary history of a species, related taxonomic groups, genes or gene families using information available from fossils, morphology and/or genetic data (Campbell and Reece 2008). This involves the construction of a phylogenetic tree that represents a hypothesis about the evolution of the organism(s) or genes being studied (Campbell and Reece 2008). Trees are constructed using shared and derived characters as a measure to infer relationships. For DNA, RNA or protein, this is based on the multiple alignment of sequences (Pevsner 2009). There are a number of methods to build phylogenetic trees, falling into two categories: distance-matrix methods which involve clustering or algorithmic methods; and character-based methods that engage tree searching (Baldauf 2003). Examples of the former are NJ method and unweighted pair group method with arithmetic mean (UPGMA), while examples of the latter method are maximum parsimony (cladistics), maximum likelihood and Bayesian state (statistical phylogenetics) (Baldauf 2003; Bininda-Edmonds 2009). The robustness of phylogenetic trees are estimated by statistics; the most commonly used is the bootstrapping method, which determines the number of times (in percentage) a specific phylogenetic relationship is preferred (Baldauf 2003; Pattengale et al. 2010).

GeneiousPro, developed by Biomatters Ltd. is an almost complete package of bioinformatics tools that allows users to upload sequences, search for homologous sequences, edit, align and construct a tree, all in one continuous programme (Kearse, 2012).

### 2.1.2.2 Expression databases: Transcriptomes and Proteomes

Many databases are available to the public that contain information about the expression and function of genes, complete genomic and functional data of a certain organism, gene ontology and gene interaction networks, among other information. The most common gene expression information that is available in public databases are transcript accumulation data based on hybridisation and/or microarray experiments, such as those in Gene Expression Omnibus (GEO) (Barrett and Edgar 2006), ArrayExpress (Parkinson et al. 2009), PlexDB (Wise et al. 2007) and species-specific sites such as TAIR (Rhee et al. 2003). Recently, transcriptome data generated by next generation sequencing (NGS) are becoming more available.

Research aiming to identify and functionally characterise specific genes or gene families commonly start with a bioinformatics approach to identify related members from the genome or transciptome, and predict the function through sequence homology, meta-analysis of expression data, or 3D structure determination. This is usually followed by functional experiments to test biologically the predicted functions and/or interactions *in planta*.

### 2.1.2.3 Protein motif analysis and tertiary structure prediction

The tertiary, or 3D structure of proteins has an important role in determining their specificity and function. Tertiary structures of proteins are determined using analytically complex techniques such as X-ray crystallography, nuclear magnetic resonance (NMR), circular dichroism and cryo-electron microscopy. These methods measure the density distribution of electrons in the protein thus deducing the relative positions, or 'coordinates' of each atom in the molecule (Sadowski and Jones 2009). While these methods determine the actual structure of an extracted protein, the techniques are expensive and laborious as they mostly require pure protein and in some cases require protein crystallisation.

To advance protein structure research, software programs have been developed which allow prediction of protein 3D structure without having to do any of the methods mentioned above. Currently, there are various bioinformatics tools that can predict the 3D structure of an unknown protein. These tools fall into two main approaches: 1) *de novo* approach, which is based on the inherent properties of amino acids but is statistically unreliable in general; and 2) model-based approach, which utilises the alignment between the target protein and a closely related protein with known 3D structures resolved from standard protein structure determination techniques such

as NMR (Guex et al. 2009; Kiefer et al. 2009). 3D structure and related information of various proteins are publicly available in databases such as the Protein Data Bank (Berman et al. 2003). Examples of most commonly used 3D structure prediction software are Modeller, Swiss Model and I-TASSER (Roy et al. 2010).

**2.2 How plants respond to stress**

Stress is defined as a change in environmental and/or physiological conditions that goes beyond the organism's optimal state such that it impairs homeostasis (Kilian et al. 2012). Stress can be caused by abiotic factors such as adverse environments and temperature, or by biotic factors such as microbial pathogens, parasites, and ecological competitors. As sessile organisms, plants are constantly and unavoidably exposed to various forms of abiotic and biotic stresses. As a result, extant plants have been shaped and are continually being shaped through evolution to develop cellular and molecular networks specialised for rapid and efficient responses against stresses (Chisholm et al. 2006; Des Marais and Juenger 2010). Stress perception and response in plants occur through a complex cascade of signals at the molecular, cellular, and physiological levels, that are unique for every type of stress, but at the same time are interconnected (Fujita et al. 2006; Lee and Luan 2012; Mantri et al. 2010; Zhao et al. 2008). Likewise, different plants may possess distinct abilities to respond to stress due to their own distinct molecular, cellular and physiological make up. Signalling of different stress stimuli may involve similar types of receptors, messengers and overlapping pathways differing only in molecular interaction kinetics. The abundance, activity, and specificity of cellular proteins and nucleic acids involved in all of these processes as well as the interaction and crosstalk between them are vital factors that affect the plant's ability to restore homeostatic conditions and/or develop disease resistance (Cohn et al. 2001; Das and Pandey 2010; Zou et al. 2010).

Primary stress signals such as excess ions and pathogen-associated molecular patterns (PAMPs) are detected by cellular receptors that trigger the release of secondary stress signals such as intracellular secondary messengers, phytohormones, and reactive oxygen species (ROS) (Kaur and Gupta 2005; Xiong and Zhu 2002) (Figure 2.2).

**Primary Stress Signals**

**ABIOTIC STRESS:**
Water deficit, heat, cold, mechanical stress

**BIOTIC STRESS:**
MAMPs/PAMPs

**Detection by cellular receptors**

**ABIOTIC STRESS:**
Transport molecules, channels, etc.

**BIOTIC STRESS:**
Pattern recognition receptors (PRRs)

**Release of secondary stress signals:**

**Secondary messengers:**
$Ca^{2+}$
Phospholipids
Nitric oxide (NO)
Inositol triphosphate (I3)

**Phytohormones**
such as ABA, JA and SA

**Reactive oxygen species (ROS)**

**Increase in superoxide dismutase**

**Increase in cell wall cross-linking to inhibit pathogen movement**

**ABA**

**JA**

**SA**

**Induce stomatal closure**

**Systemic acquired resistance (SAR)**

**Water retention, reduced entry of foreign materials**

**Release of protease inhibitors and defence compounds**

**Signalling cascades:**
Phosphorylation cascades (such as MAPK and $Ca^{2+}$ kinase relays)
Redox reactions
Regulation of stress-inducible genes and transcription factors
(osmoprotectants, anti-freeze proteins, detoxifiers, etc.)
Synthesis of antimicrobial products or pathogenesis-related (PR) proteins

**ABIOTIC STRESS:**
LEA pathways, SOS signalling

**BIOTIC STRESS:**
Pathogen-triggered immunity (PTI)

**Pathogen effectors**

**Short and long-term responses to maintain homeostasis**

**Effector-triggered susceptibility (ETS)**

**Effector-triggered immunity (ETI)**
Hypersensitive responses, PR and SA accumulation

**Systemic acquired resistance (SAR)**

**Figure 2.2. An overview of plant abiotic and biotic stress responses.** Information in this figure was derived and modified from Jones and Dangl (2006), Kaur, Schumaker and Zhu (2002), and Rodriguez, Canales and Borras-Hidalgo (2005).

Intracellular secondary messengers are molecules that pass on signals from a receptor to a target molecule, as part of a signalling cascade, called signal transduction. Signals are passed through different molecular interactions such as protein conformational change, phosphorylation (addition of phosphate groups), and redox (reduction-oxidation) reactions. Examples of intracellular secondary messengers are ions such as $Ca^{2+}$, phospholipids, and gases like nitric oxide (NO). Hormones were initially coined as the 'primary messengers' in signal transduction, particularly in developmental processes; however, in stress responses these are considered as secondary stress signals because their concentration increases in reaction to stress. Jasmonates (jasmonic acid (JA) and derivatives), salicylates (salicylic acid (SA) and derivatives) and abscisic acid (ABA) and derivatives are the three major phytohormone classes that, apart from having growth and developmental roles, regulate plant stress responses (Erb and Glauser 2010). Jasmonates are involved in wound and insect responses by triggering protease inhibitors and defence compounds (Birkett et al. 2000; Stratmann 2003) and in pathogen defence by participating in local and systemic resistance (Halim et al. 2006). Salicylates play important roles in photosynthesis, ion transport, leaf anatomy, and the development of systemic acquired resistance (SAR) (Ashraf et al. 2010; Erb and Glauser 2010; Halim et al. 2006). ABA induces stomatal closure which reduces water loss and mediates the expression of various defence proteins through inositol triphosphate (I3) and $Ca^{2+}$ signalling pathways (Maksimov 2009; Wasilewska et al. 2008). ROS are molecules such as superoxide and hydrogen peroxide that can stimulate an increase in cell wall cross-linking and strength to inhibit the movement of pathogens to other parts of the plant (Xiong and Zhu 2002). ROS act as secondary stress signals that activate scavenging enzymes such as superoxide dismutase. The increase in activity of these enzymes supports ABA signalling and abiotic stress tolerance (Li et al. 2012; Ozfidan et al. 2012). Both primary and secondary stress signals lead to further signalling pathways and regulation of a large array of stress inducible genes, including those that produce osmoprotectants, anti-freeze proteins, chaperones, detoxification enzymes, and defensins. Important signalling systems that interplay in plant stress response include mitogen activated protein kinase (MAPK) pathways and $Ca^{2+}$ dependent signalling (Wurzinger et al. 2011).

### 2.2.1 Plant responses to abiotic stresses

Osmotic and ionic pressure caused by drought, salt, flooding, extreme temperatures and drastic climate change are the main abiotic stresses that limit plant survival and productivity (Kaur and Gupta 2005; Qin et al. 2011; Rodriguez et al. 2005; Xiong and Zhu 2002).

### 2.2.1.1 Osmotic pressures: Drought and salinity

Severe imbalance in water and ionic distribution damages the cell, inhibits growth and photosynthetic activities and may eventually lead to death (Rodriguez et al. 2005). Plants exhibit a number of long and short–term responses to compensate for or prevent water loss (reviewed in detail by Chaves et al. 2003; Xiong and Zhu 2002). Water deficit is essentially managed immediately by ion and water transport adjustments such as stomatal closure, decreased carbon assimilation, hydraulic changes in the xylem and osmotic adjustment in the roots; and in long-term by morphological and developmental changes such as turgor maintenance, increased absorption area, leaf and shoot growth inhibition, sustained root growth and life cycle changes (Chaves et al. 2003; Xiong and Zhu 2002). These responses are mediated by phytohormones and several signalling systems including ROS and MAPK pathways that produce antioxidant compounds and osmolites (Ozfidan et al. 2012), late-embryogenesis abundant (LEA) protein pathways that protect cell structure and repair cell damage, and salt overly sensitive (SOS) signalling that targets ion transporters to restore ionic homeostasis (Rodriguez et al. 2005).

### 2.2.1.2 Extreme temperatures

High temperature stress occurs when plants are exposed to temperatures beyond a threshold for a period of time, which may cause impaired growth and development (Hasanuzzaman et al. 2013). On the other hand, low temperature stress can either be freezing stress, producing injury due to ice crystal formation in plant tissues; or chilling stress, causing injury without ice crystal formation (Hasanuzzaman et al. 2013; Miura and Furumoto 2013). Both high and low temperature stress lead to generation of toxic compounds in the plant cells, including ROS (Hasanuzzaman et al. 2013). Heat stress response in plants is characterised by the accumulation of heat shock proteins regulated by heat stress transcription factors (Kotak et al. 2007). Heat shock proteins serve as molecular chaperones during heat stress; however, the exact mechanism of how these proteins contribute to heat tolerance is not very well known (Kotak et al. 2007). Heat stress transcription factors regulate the expression of heat shock proteins and other heat stress-induced genes. This leads to multiple signalling pathways, including ABA, SA, ethylene, calcium, and ROS signalling resulting in thermotolerance (Kotak et al. 2007). In cold stress, reduced membrane fluidity causes $Ca^{2+}$ levels increase in the cytoplasm. The $Ca^{2+}$ signatures are decoded by calcium sensors and responder proteins and activate signalling pathways such as the Inducer of CBF Expression-Dehydration responsive element-binding factor 1a (ICE-CBF/DREB1A) pathway and pathways involving cold-responsive (COR) genes

(Heidarvand and Maali Amiri 2010; Miura and Furumoto 2013). MAPK and LEA pathways are also involved in cold stress signalling, as well as phytohormone responses such as ABA, auxin, gibberellic acid, SA, and ethylene responses (Miura and Furumoto 2013).

### 2.2.1.3 Mechanical stress

Mechanical stress or wounding in plants involves the expression of rapid wound responsive genes (RWR) (Walley et al. 2007). These genes lead to different pathways such as the general stress response (GSR) which generates ROS, and various phytohormone pathways (Walley et al. 2007). A number of identified RWR genes are also abiotic stress inducible genes, such as the ethylene response factor 18 (ERF-18) and CCR4-associated factor (CAF-1) (Walley et al. 2007). Mechanically-induced stress happens in nature when plant parts are moved or rubbed by different agents such as wind, rain, and animals (Biddington 1986). Mechanical stress affects the growth of plants, which commonly results in shorter plants characterised by reduction in stem height, leaf length and petiole length (Biddington 1986).

### 2.2.2 Plant responses to biotic stresses

Plant biotic stresses include pests such as insects, and diseases such as infection by certain viruses, bacteria, fungi, and protists. Pathogenic infections often have unique symptoms; they can be local or systemic, and in most cases affect the plant's morphology, photosynthetic activities, development, reproduction and survival as the pathogens proliferate. Plants, like animals, have an immune system capable of recognising and distinguishing between self and non-self entities. However, plants do not have circulating immune cells and antibodies (Jones and Dangl 2006). Plants rely only on innate immunity and do not possess the adaptive immunity exhibited by vertebrates (Iriti and Faoro 2007). The plant innate immune system is characterised by two branches (Jones and Dangl 2006) (Figure 2.2). The first branch involves the detection of microbial- or pathogen-associated molecular patterns (MAMPs or PAMPs) by host transmembrane pattern recognition receptors (PRRs). This results in signalling pathways directed to fight against the pathogen, such as the release of ROS and NO, intracellular pH changes, and synthesis of antimicrobial products or pathogenesis-related (PR) proteins (Cohn et al. 2001). This initial recognition-response phase is known as pathogen-triggered immunity (PTI) and is then counteracted by pathogens that release 'effector' molecules that improve the pathogen's virulence. Effectors obstruct PTI and may render the host susceptible, a state known as effector-triggered susceptibility (ETS). Plants have developed a second branch of innate

immunity, involving effector-triggered immunity (ETI) where the host specifically recognises the effector molecules of a pathogen and once again trigger immune responses against the pathogen. The specific recognition of diverse types of effectors is usually performed by nucleotide-binding leucine-rich repeat (NB-LRR) protein products which are encoded by most disease resistance (R) genes. *R* gene products specifically recognise corresponding effectors called avirulence (*avr*) proteins from the pathogen. ETI responses are similar to PTI but are faster and stronger. As these processes can be toxic (i.e. accumulation of ROS, NO, prolonged pH changes) to the host cell itself, this may lead to localised cell death at the site of infection, thus are referred to as hypersensitive responses (HRs). Moreover, HR gives rise to local and systemic accumulation of a wide assortment of PR proteins. With the elevation of endogenous SA hormone levels, PR protein abundance leads to SAR that confers protection against subsequent attack by a broad spectrum of pathogens (Durrant and Dong 2004). Through natural selection, pathogens avoid ETI by diversifying effectors, which also results in new *R* gene specificities as plants evolve. Polymorphisms (nucleotide sequence differences between organisms of the same population) in the *R* gene loci of different plants bring about variation in plants' susceptibility to different pathogens (Miura and Furumoto 2013).

$Ca^{2+}$ is important in secondary stress signalling. $Ca^{2+}$ is a well-studied secondary messenger that plays important roles both in abiotic and biotic stress signalling in plants. $Ca^{2+}$ signals regulate a myriad of cellular functions, including stomatal movement, tip structure development, cytoskeletal movements, pathogenic or symbiotic interactions, among others. Upon exposure to a specific stimulus (stress or developmental), plants undergo transient and repetitive $Ca^{2+}$ concentration flux in the cytoplasm (Dodd et al. 2010). This happens as $Ca^{2+}$ transporters become activated and promote $Ca^{2+}$ movement into the cell (influx) as a result of release from internal stores such as vacuoles, chloroplasts and mitochondria (DeFalco et al. 2010; Hashimoto and Kudla 2011; Rodriguez et al. 2005). $Ca^{2+}$ oscillation patterns are unique in each type of stimulus and occur at specific frequencies, amplitude and location in the cell (referred to as '$Ca^{2+}$ signatures') (DeFalco et al. 2010; Laude and Simpson 2009). These signatures are decoded by $Ca^{2+}$ binding proteins, which regulate downstream effects that are also specific to a certain type of stimulus. The mechanism and kinetics of how specific $Ca^{2+}$ transporters and $Ca^{2+}$-binding proteins become activated to a particular stimulus and lead to downstream effects is still unclear (Boudsocq et al. 2010; Hashimoto and Kudla 2011). A great number of studies in the past decade have focused on $Ca^{2+}$-sensing proteins and how the

signatures are decoded into certain signalling pathways (Boudsocq et al. 2010; Hashimoto and Kudla 2011; Reddy et al. 2011; Reddy and Reddy 2004).

Ca$^{2+}$-binding proteins have been reported to be involved in signalling pathways that are related to growth, development, abiotic stress responses and pathogen defence, such as MAMP, MAPK, ROS, SOS, drought response element binding (DREB) proteins, ABA, and JA signalling, as well as in the regulation of stress-responsive genes like plant defensin gene (PDF1.2) and alcohol dehydrogenase (ADH1)(DeFalco et al. 2010; Dodd et al. 2010). Ca$^{2+}$-binding proteins in plants, briefly described in Chapter 1, include CaMs, CMLs, CBLs, and CPKs. Upon binding with Ca$^{2+}$, these proteins undergo conformation changes and either directly or indirectly result in phosphorylation reactions via their kinase activity. CaMs, CMLs and CBLs are classified as sensor-relays because they only sense Ca$^{2+}$ signatures and bind to responder molecules such as CCaMKs and CIPKs which phosphorylate specific targets. CPKs on the other hand are referred to as Ca$^{2+}$ sensor-responders, having both Ca$^{2+}$-binding and kinase domains. All of the Ca$^{2+}$-binding proteins have a characteristic EF-hand motif, a type of motif that canonically has four Ca$^{2+}$ binding loops (with a characteristic shape called EF-hand), but can have between one to five EF-hands. These are approximately 12 aa long, starting with an aspartate and ending with glutamate (Grabarek 2006; Kawasaki et al. 1998; Zhou et al. 2006).

## 2.3 What are CPKs?

### 2.3.1 Nomenclature and canonical structure of CPKs

CPKs are ubiquitous in plants and directly bind Ca$^{2+}$ ions before phosphorylating substrates involved in metabolism, osmosis, hormone response, and stress signalling pathways (Harmon et al. 2001; Klimecka and Muszyńska 2007). Most authors refer to CPKs as "calcium-dependent protein kinases" (Hamel et al. 2014; Hrabak et al. 2003; Munemasa et al. 2011) while some authors use "calmodulin-like domain protein kinases" (Zhang and Choi 2001). Based on the proposed nomenclature 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 uppercase) and species (in lowercase), followed by the abbreviation 'CPK' and a number (e.g. AtCPK1 for Arabidopsis). Most authors have followed this nomenclature style; however, some authors have used "CDPK" in naming sequences (e.g. TgCDPK1) (Billker et al. 2009; Jaworski et al. 2010; Kugelstadt et al. 2007; Wernimont et al. 2011; Wernimont et al. 2010), while some have used the abbreviation 'CPK' instead of 'CDPK' when referring to CPKs in general (Arimura

and Maffei 2010; Kanchiswamy et al. 2010; Murillo et al. 2001). This thesis follows the proposed gene nomenclature style of Hrabak et al. (2003) that has been adapted by most plant researchers. However, throughout the text the abbreviation 'CPK' is used in order to maintain uniformity and avoid confusion.

CPKs consist of four domains (Harmon et al. 2001): a variable N-terminal domain (N-VD), a catalytic protein kinase domain (PK), an autoinhibitory junction domain (AJ) and a calmodulin-like domain, also called $Ca^{2+}$-activation domain (CAD) (Figure 2.3). Some authors consider another domain, the C-terminus (CT), which is as highly variable as the N-terminus, but generally shorter in length (Klimecka and Muszyńska 2007). CPKs have different isoforms and each isoform has different substrate specificities, $Ca^{2+}$ sensitivity, cellular localisation and function. Binding of $Ca^{2+}$ ions occurs in the CAD, containing one to five loops called EF-hand loops (usually four). Each loop is 12 aa long and is flanked by two α-helices, thus having a helix-loop-helix type of arrangement (Cheng et al. 2002).



**Figure 2.3. Characteristic primary structure of CPKs.** N-VD= N-terminal variable domain, PK = catalytic protein kinase domain, AJ= autoinhibitory junction domain, CAD= $Ca^{2+}$ activation domain commonly containing four EF hands (black boxes) and CT= C-terminal variable domain.

By inference from its similarity with CaMK (Harmon et al. 2000) and based on structural studies of the soybean CPK AJ-CAD region (Weljie et al. 2004; Weljie and Vogel 2004), it has been proposed that $Ca^{2+}$ binds to the EF hands within CAD and causes a conformational change in CAD so that it binds to the AJ domain (Figure 2.4a and b). This in turn releases the AJ domain from the substrate-binding site of the PK domain where it is normally bound (Figure 2.4b). The host cellular substrate can then bind to the PK domain, rendering the CPK active. However, recent structural studies on protist CPKs indicate that it is the CAD that actually blocks the active site of PK and changes conformation and position when bound with $Ca^{2+}$ (Lim et al. 2012). This mode of activation, however, needs to be confirmed in plants by resolving the complete (full sequence) 3D structure of a full-length plant CPK.
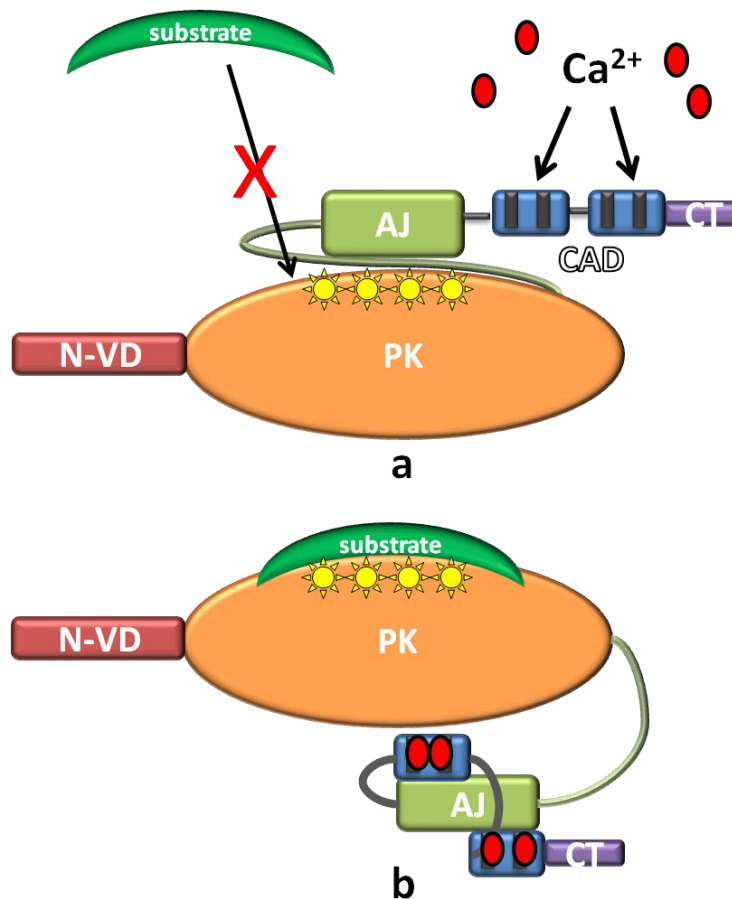
**Figure 2.4. Mechanism of activation of CPKs.** (a) Inactive CPKs have their substrate-binding sites concealed by the AJ-CAD domains. (b) Ca2+- activated CPKs undergo conformational changes that shifts the AJ-CAD domains to one side of the PK domain, rendering the active sites available for substrate-binding. N-VD= N-terminal variable domain, PK = catalytic protein kinase domain, AJ= autoinhibitory junction domain, CAD= Calcium-activation domain commonly containing four EF hands (black boxes) and CT= C-terminal variable domain, Green half-moon= substrate target, yellow sun-shaped figures= ATP binding sites and active sites.

CPKs were first described by Hetherington and Trewavas (1982) from garden pea extract (*Pisum sativum*), and were initially purified and characterised by Harmon et al. (1987) from soybean (*G. max*) (as cited in Cheng et al. 2002). CPKs are found in protists, algae, oomycetes and plants, but not in animals or fungi (Harmon et al. 2001; Harmon et al. 2000; Hashimoto and Kudla 2011; Suntio and MÄKinen 2012). Instead, fungi and animals are abundant with CaM and CaMKs (Figure 2.5), which conversely are rare in protists, algae and plants (Billker et al. 2009). Based on sequence similarity and intron-exon structure, it is thought that the CPK gene family (Figure 2.5) arose through the fusion of genes encoding a CaMK and a calmodulin (Harmon et al. 2000; Harper et al. 2004; Zhang and Choi 2001). As mentioned in section 1.1, this characteristic structure is unique to CPKs among the other members of its superfamily, the CPK-SnRK superfamily of protein kinases (Hrabak et al. 2003) and among other $Ca^{2+}$-binding proteins (DeFalco et al. 2010; Hashimoto and Kudla 2011) (Figure 2.5). In contrast with CPKs, the

calcium-binding proteins CaM, CMLs and CBLs are separate from their respective kinases, and only bind to the AJ domains of their kinases once activated by $Ca^{2+}$. This binding then brings conformational change to the AJ domain, allowing the protein kinase to phosphorylate their specific substrates.
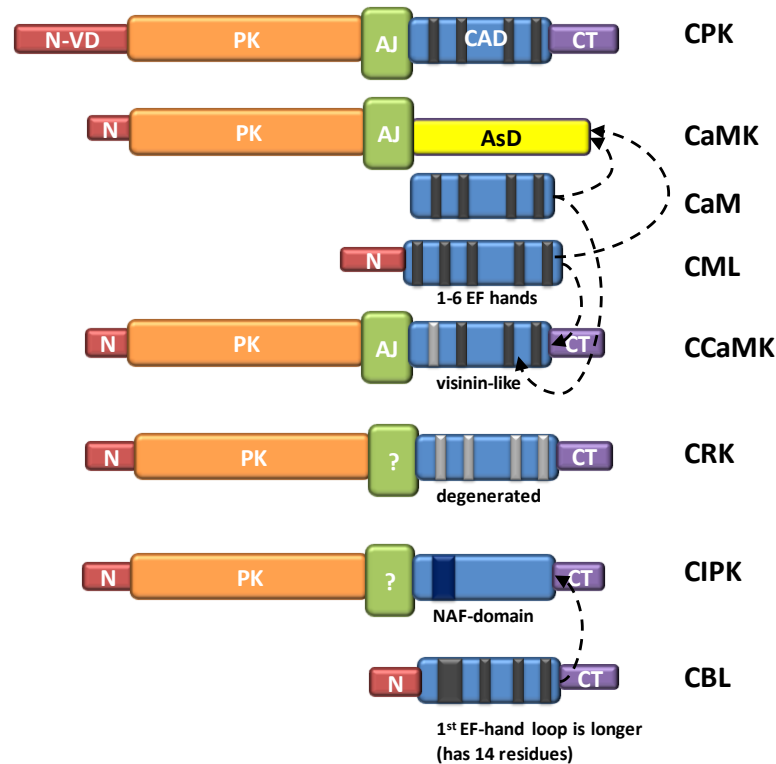


**Figure 2.5. The CPK-SnRK superfamily and Ca2+-binding proteins in plants.** The CPK -related protein kinases are CaMK, CCaMK, CRK and CIPK while the other Ca2+-binding proteins are CaM, CML and CBL. N-VD= N-terminal variable domain, N= N-terminus, PK = catalytic protein kinase domain, AJ= autoinhibitory junction domain, CAD= Calcium-activation domain commonly containing four EF hands (black boxes), CT= C-terminal variable domain, AsD= association domain. Black broken arrows indicate kinases and their corresponding Ca2+ sensor.

Tertiary structures of CPK domains have been reported in recent years, but the structures of a complete CPK protein has only been resolved for the apicomplexan protists *Toxoplasma gondii* and *Cryptosporidium parvum*, namely TgCPK1, TgCPK3, and CpCPK1 (Wernimont et al. 2011; Wernimont et al. 2010). Structural studies have been performed for Arabidopsis (Chandran et al. 2006) and soybean (Weljie et al. 2004; Weljie and Vogel 2004) CPKs; however these 3D structures only included the C and AJ domain. The 3D structure for a complete plant CPK sequence has not yet been elucidated.

The 3D structure of one of the protist CPKs, TgCPK1, is shown in Figure 2.6. Briefly, the protein kinase is bi-lobed (Figure 2.6a), with the first lobe (PK lobe 1) having many beta-

pleated sheets shown in yellow arrows, and the second lobe (PK lobe 2) having a series of α

helices (pink ringlet structures) with several turns (blue lines). There is an opening between the

two lobes, whose shape may define the substrate specificity of the kinase. As the figure shows

an activated CPK, the CAD and AJ regions do not cover the substrate-binding regions but are

instead twisted towards the other side of the kinase. Figure 2.6b focuses on the CAD, showing

the EF hand loops in more detail, where $Ca^{2+}$ molecules bind.
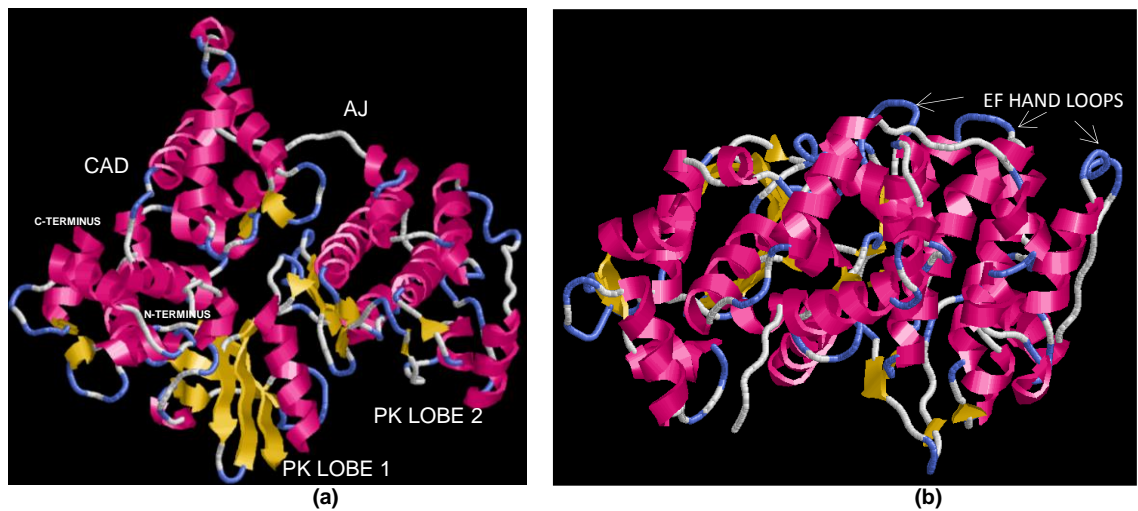


**Figure 2.6. The 3D structure of TgCPK1 in the presence of Ca2+.** (a) TgCPK1 top view, N- and C-terminus on the lower-left hand; (b) TgCPK1 side view, N- and C- terminus on the lower-left hand. Figures downloaded from Protein Data Bank, http://www.rcsb.org/pdb/explore/explore.do?structureId=3HX4.

### 2.3.2 The CPK gene family in plants

CPKs have been described as a large multigene family, with 34 members in Arabidopsis (Cheng et al. 2002; Suntio and MÄKinen 2012), 31 members in rice (Asano et al. 2005), an estimate of 26 members in wheat (Li et al. 2008b), and 30 members in poplar (cited in Li et al. 2008b). CPKs from Arabidopsis and rice were identified from genome sequence and mRNA expression analyses while CPKs from wheat were obtained only from tentative consensus sequences and expression analyses. No specific paper has identified or characterised CPKs from poplar, although there are sequences predicted to be CPKs based on the completed genome sequence. Similarly, there are several CPK genes present in protist gene models: 12 members have been identified in *Toxoplasma gondii;* seven in each of the genomes of *Plasmodium falciparum*, *Cryptosporidium parvum,* and the ciliate *Tetrahymena thermophila* (Billker et al. 2009). CPK sequences have also been identified and analysed in a few species of green algae, moss, and liverworts, and several crop species and model plants, although these reports have focused only on one or two CPKs from each organism (Giammaria et al. 2011; Jain et al. 2011; McCurdy and Harmon 1992; Mitra and Johri 2000; Murillo et al. 2001; Nishiyama et al. 1999; Sugiyama et al. 2000; Yuasa and Muto 1992).

The 34 Arabidopsis CPKs (AtCPKs) are divided into four major evolutionary groups, based on unrooted phylogenetic analyses of protein and nucleotide sequence (Figure 2.7a) (Cheng et al. 2002; Hrabak et al. 2003). The most divergent AtCPKs are those belonging to group IV, while the most complex group is group II, having 13 members. Since the rice and wheat CPKs (OsCPKs and TaCPKs, respectively) were published, groups II and III have both been further subdivided into two subgroups (a and b), as shown in the phylogenetic tree in Figure 2.7b) (Asano et al. 2011; Li et al. 2008b).

CPKs are highly homologous to each other, with protein sequence similarities ranging from 56% to 96% in Arabidopsis and 32% to 99% in rice. Some CPKs are very closely related, having very high sequence identities with each other. In Arabidopsis, there are eight closely-related pairs or sets: AtCPK4 and 11 (95% similarity), AtCPK17 and 34 (93% similarity), AtCPK7 and 8 (90% similarity), AtCPK10 and 30 (86% similarity), AtCPK9 and 33 (85% similarity), AtCPK1 and 2 (81% similarity), AtCPK21 and 23 (81% similarity), and lastly AtCPK5, 6 and 26 (85%–88% similarity) (Cheng et al. 2002). In rice, there are eleven closely-related pairs or sets: OsCPK1 and 15 (86% similarity), OsCPK2 and 14 (87% similarity), OsCPK3 and 16 (92% similarity), OsCPK4 and 18 (82% similarity), OsCPK5 and 13 (81% similarity), OsCPK7 and 23 (71%

similarity), OsCPK8 and 20 (75% similarity), OsCPK11 and 17 (79% similarity), OsCPK21 and 22 (71% similarity), OsCPK24 and 28 (86% similarity) and OsCPK25 and 26 (99.6% similarity) (Asano et al. 2005). Some of these pairs are encoded by close gene loci within the same chromosome, such as AtCPK17 and 34, AtCPK7 and 8, and AtCPK21 and 23, suggesting that these genes were derived from tandem duplications. On the other hand, some are encoded on different chromosomes but in recently duplicated genome segments (Paterson et al. 2004, as cited by Asano et al. 2005), such as OsCPK25/26, OsCPK1/15, OsCPK2/14, OsCPK3/16, OsCPK5/13,OsCPK11/17, OsCPK21/22.

### 2.3.3 Reported functions of CPKs

CPKs have been shown to respond to different stimuli and participate in various processes such as development, abiotic stress responses, pathogen defence, as well as cellular transport, movement, and division. Most of this evidence comes from studies of Arabidopsis, rice, and wheat CPKs. CPKs have also been identified and functionally characterised in some crops such as *V. vinifera* (grape), (Yu et al. 2006), *Z. mays* (maize) (Estruch et al. 1994; Murillo et al. 2001; Szczegielniak et al. 2005; Takezawa et al. 1996), *S. tuberosum* (potato) (Gargantini et al. 2009; Giammaria et al. 2011; Kobayashi et al. 2007; Raices et al. 2001), *Solanum lycopersicum* (tomato) (Chang et al. 2011; Chang et al. 2009; Rutschmann et al. 2002), and *Nicotiana tabacum* (common tobacco) (Tai et al. 2009). An overview of the reported response and biological functions of CPKs is shown in Table 2.1. The involvement of CPKs in plant development and stress responses will be discussed in the following sections.
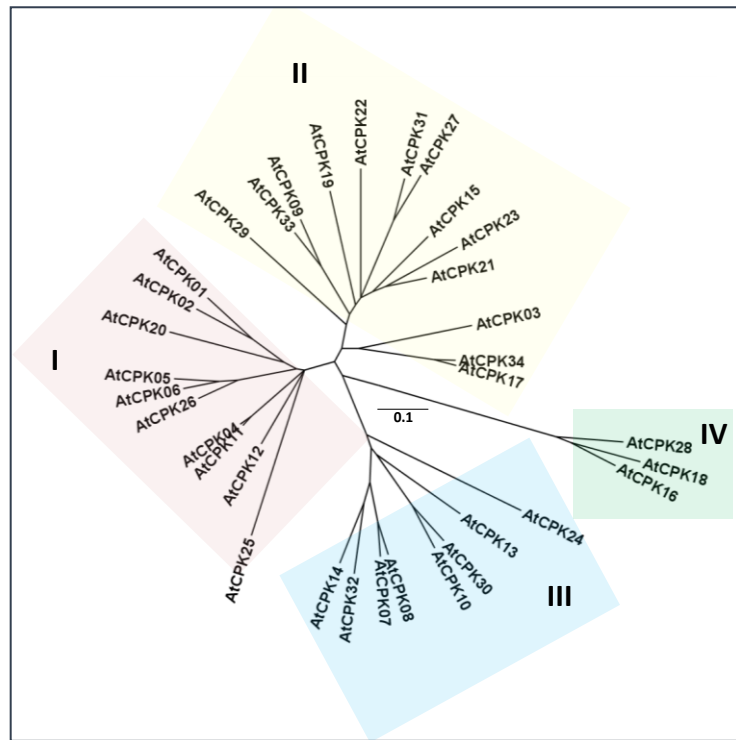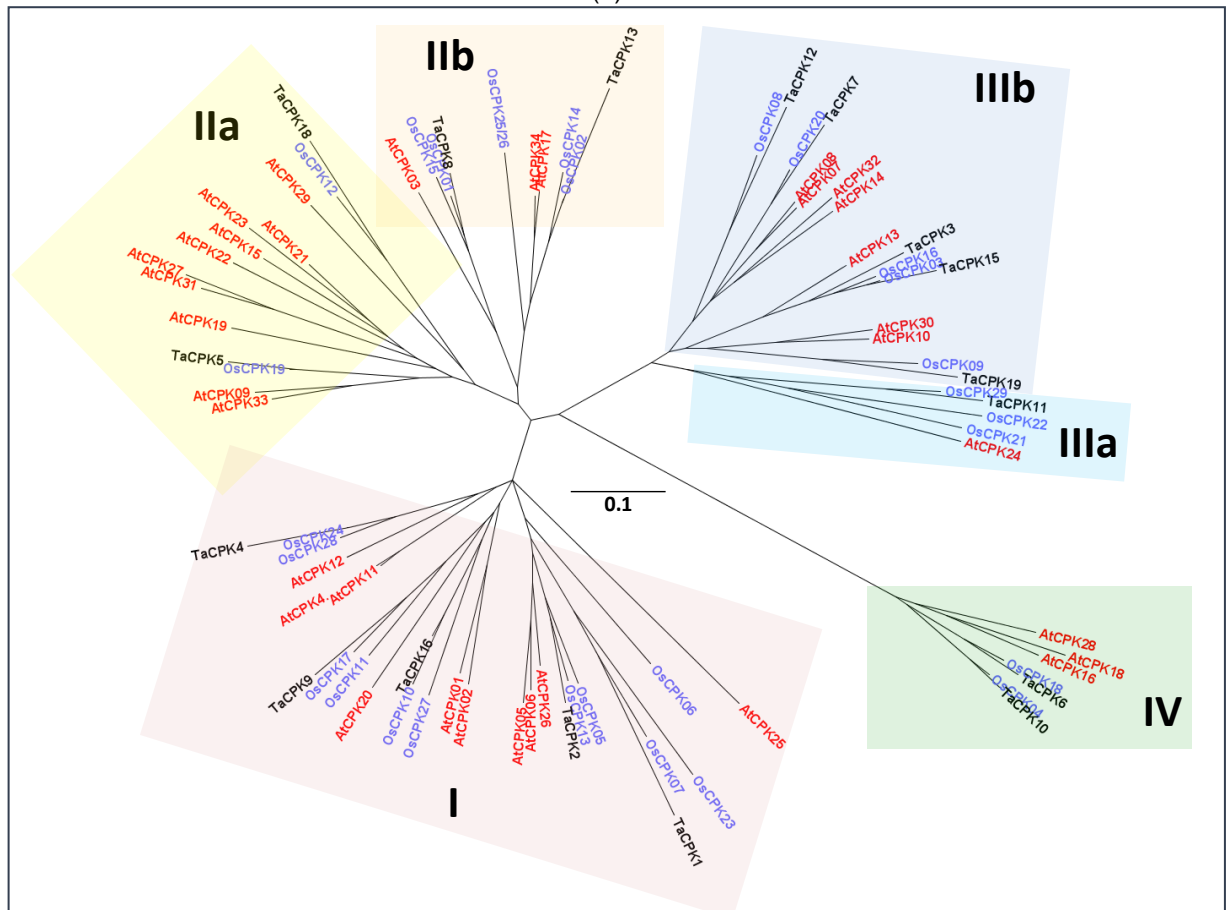
**Figure 2.7. Evolutionary grouping of CPKs based on previous reports**. Phylogenetic tree of (a) Arabidopsis CPKs as published by Cheng et al. (2002) and (b) Arabidopsis, rice and wheat as published by Li et al. (2008). In (b) Arabidopsis CPKs are in red font while rice CPKs are in blue font. Groups II and III have each been divided into two subgroups.

**Table 2.1. A summary of functional information regarding the involvement of CPKs to development and stress signaling.** Tick marks indicate response to a specific hormone, or developmental function. Arabidopsis CPKs are in red font while rice CPKs are in blue font. Other plant CPKs are in black font. CPKs listed here may be up- or down-regulated. More detailed information is presented in Appendices 1 and 2. At, Arabidopsis; Os, rice; Ta, wheat; Vv, grape; St, potato; Le, tomato.

| | ABIOTIC STRESS | | | | | BIOTIC STRESS | | | | OTHER RESPONSES/DEV'T FUNCTION | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Drought | High Salt | Cold | Heat | Wound | Fungi & Elicitors | Bacteria & flg22 | Virus | Insects | Stomatal closure | Hormones | ABA | GA(3) | SA | JA/MeJA | Development | Pollen-specific | Pollen tube elongation | Flower-specific | seed/early plant devt |
| **Group I** | AtCPK04<br>AtCPK06<br>AtCPK11<br>OsCPK05<br>OsCPK07<br>OsCPK10<br>OsCPK13<br>TaCPK01<br>TaCPK09 | AtCPK04<br>AtCPK06<br>AtCPK11<br>OsCPK05<br>OsCPK07<br>OsCPK13<br>TaCPK04<br>TaCPK09 | OsCPK05<br>OsCPK07<br>OsCPK13<br>TaCPK04<br>TaCPK01 | | ZmCPK11 | AtCPK01<br>ZmCPK10<br>TaCPK02<br>TaCPK04<br>TaCPK011<br>TaCPK01 | AtCPK01<br>AtCPK04<br>AtCPK05<br>AtCPK06<br>AtCPK11<br>AtCPK26<br>AtCPK04<br>AtCPK05<br>AtCPK06 | | | AtCPK06<br>VvCPK01 | AtCPK01<br>AtCPK04<br>AtCPK06<br>AtCPK11<br>OsCPK13<br>TaCPK02<br>TaCPK04<br>TaCPK09<br>VvCPK01 | AtCPK04 ✓<br>AtCPK06 ✓<br>AtCPK11 ✓<br>OsCPK13 ✓<br>TaCPK04 ✓<br>TaCPK09 ✓<br>VvCPK01 ✓ | OsCPK13 ✓<br>TaCPK02 ✓<br>TaCPK04 ✓ | AtCPK01 ✓ | AtCPK06 ✓ | OsCPK27<br>AtCPK11<br>TaCPK01<br>OsCPK11<br>ZmCPK11<br>OsCPK24<br>OsCPK23<br>VvCPK1 | OsCPK27 ✓<br>AtCPK11 ✓ | | OsCPK27 ✓<br>OsCPK11 ✓<br>OsCPK24 ✓ | TaCPK01 ✓<br>ZmCPK11 ✓<br>OsCPK23 ✓<br>VvCPK1 ✓ |
| **Group IIa** | AtCPK21<br>AtCPK23<br>OsCPK12<br>TaCPK18 | AtCPK21<br>AtCPK23<br>OsCPK12<br>TaCPK18 | TaCPK05 | | LeCPK1 | OsCPK12<br>LeCPK1<br>AtCPK9<br>AtCPK15 | AtCPK9<br>AtCPK15 | | | AtCPK23 | AtCPK33<br>TaCPK05<br>StCPK1<br>OsCPK12 | AtCPK33 ✓<br>TaCPK05 ✓<br>StCPK1 ✓<br>OsCPK12 ✓ | TaCPK05 ✓<br>StCPK1 ✓ | | StCPK1 ✓ | AtCPK9<br>LeCPK1 | | | AtCPK9 ✓<br>LeCPK1 ✓ | |
| **Group IIb** | AtCPK03<br>OsCPK01<br><br>OsCPK15 | AtCPK03<br>OsCPK01<br><br>OsCPK15 | AtCPK03 | AtCPK03 | | AtCPK3 | AtCPK03 | AtCPK03 | AtCPK03 | AtCPK03 | | | | | | AtCPK17<br>AtCPK34<br>OsCPK02<br>OsCPK14<br>OsCPK25&26 | AtCPK17 ✓<br>AtCPK34 ✓<br><br><br>OsCPK25&26 ✓ | AtCPK17 ✓<br>AtCPK34 ✓ | OsCPK02 ✓<br>OsCPK14 ✓<br>OsCPK25&26 ✓ | OsCPK02 ✓ |
| **Group IIIa** | OsCPK21 | OsCPK21 | OsCPK21 | | | | | | | OsCPK21 | OsCPK21 | OsCPK21 ✓ | | | | AtCPK24<br>OsCPK29<br>OsCPK21&22 | AtCPK24 ✓<br>OsCPK29 ✓<br>OsCPK21&22 ✓ | | | |
| **Group IIb** | AtCPK10<br>OsCPK16 | AtCPK32<br>TaCPK19 | TaCPK03<br>TaCPK07<br>TaCPK12<br>TaCPK15 | | | TaCPK03<br>TaCPK07<br>TaCPK12<br>TaCPK15<br>TaCPK19<br>AtCPK7,8,10,32<br>OsCPK9 | AtCPK10<br>AtCPK13<br>AtCPK30<br>AtCPK32<br>CaCPK2 | AtCPK14<br>AtCPK32 | AtCPK13 | AtCPK10 | AtCPK10<br>AtCPK30<br>TaCPK03<br>TaCPK07<br>TaCPK12<br><br>TaCPK15 | AtCPK10 ✓<br>AtCPK30 ✓<br>TaCPK03 ✓<br>TaCPK07 ✓ | AtCPK30 ✓<br>TaCPK03 ✓<br><br>TaCPK12 ✓<br>TaCPK15 ✓ | | | AtCPK14<br>AtCPK32<br>AtCPK19 | AtCPK14 ✓<br>AtCPK32 ✓ | AtCPK14 ✓<br>AtCPK32 ✓ | AtCPK19 ✓ | |
| **Group IV** | TaCPK06 | OsCPK18<br>TaCPK10 | OsCPK04<br>TaCPK06 | | LeCPK2 | OsCPK04<br>OsCPK18<br>TaCPK10 | AtCPK28 | AtCPK18<br>AtCPK28 | | | TaCPK06<br>LeCPK2<br>TaCPK10 | TaCPK06 ✓<br>TaCPK10 ✓ | | LeCPK2 ✓ | LeCPK2 ✓ | AtCPK16<br>LeCPK2 | AtCPK16 ✓ | | LeCPK2 ✓ | |

57

### 2.3.3.1 Developmental functions of CPKs

CPKs appear to have important roles in plant development and reproduction. This is mostly based on high transcript accumulation in seeds, seed development and reproductive organs such as pollen/stamen, ovary/ovules, flowers and fruit. A number of CPKs have relatively abundant transcripts in seed and panicle (developing flower) tissues, such as OsCPK11, 23, and 24 (Ray et al. 2007), ZmCPK11 (Szczegielniak et al. 2005) and TaCPK1 (Li et al. 2008b) from evolutionary group I and OsCPKs 2, 14, 25 and 26 (Ye et al. 2009) from evolutionary group IIb. Other CPKs are specifically abundant (in the stamen or pollen, for example OsCPK11 and 27) (Ray et al. 2007) from group I, AtCPK17 and 34 (Zhou et al. 2009) from group IIb and AtCPK24 from group IIIa. OsCPKs 21, 22 and 29 (also group IIIa) have high accumulation of transcript in the panicle and stamen (Ray et al. 2007), while AtCPK14 and 32 (group IIIB) are transcripts pollen-specific (Zhou et al. 2009). A tomato CPK (LeCPK2) has very high transcript accumulation in flower tissues (Chang et al. 2009). In addition, a CPK from grape berry (*V. vinifera* x *Vitis labrusca*), VvCPK1 (also named as ACPK1), which is closely related to OsCPK24 and ZmCPK11, is preferentially expressed in the mesocarp (fleshy layer of fruit) and seeds, and exhibits an increase in enzymatic activity during fruit development as induced by the ABA hormone (Yu et al. 2006; Yu et al. 2007).

There are very few studies that used reverse genetics and overexpression approaches to demonstrate the role of CPKs in development. These studies have mostly focused on pollen and seed development. Group IIb CPKs AtCPK17 and 34 (93% identical in aa sequence) are both important in pollen growth and development (Zhou et al. 2009). Transient overexpression of CPKs fused with green fluorescent protein (GFP), or CPK-GFP-fusion proteins, in tobacco pollen has shown that pollen tube growth depolarisation (reduced elongation/ increased width) can be induced by the overexpression of AtCPK34 but is not affected by the overexpression of AtCPK17 (Zhou et al. 2009). Myers et al. (2009) reported that Arabidopsis *cpk17* and *cpk34* double mutants (T-DNA insertion knockout mutants) show 350-fold reduction in pollen transmission efficiency and three-fold reduction in pollen tube growth rate. This information supports the importance of AtCPK 17 and 34 in phosphorylating substrates that are involved in cellular elongation and movement, particularly in pollen tube formation.

OsCPK2 is closely related to AtCPK17 and 34. It has been reported that its protein abundance is very low in leaves exposed to light but very high in the dark (Morello et al. 2000).

OsCPK2 appears to be important in light-responsive signalling involved in seed development, because plants that overexpress OsCPK2 appear normal in morphology but have disrupted seed formation and light exposure represses its overexpression (Morello et al. 2000).

AtCPK24, 14, and 32, which are pollen-specific CPKs, also appear to be regulators of pollen tube development, as reported by Zhou et al. (2009) through their transient overexpression experiments of GFP-tagged AtCPKs. AtCPK24 in tobacco pollen slightly inhibited pollen tube elongation, while AtCPK14 and 32 significantly induced depolarisation of pollen tube growth (reduced elongation); with AtCPK32 having the most severe outcome (Zhou et al. 2009).

### *2.3.3.2 Role of CPKs in abiotic stress response*

**Drought and Salinity**

As described in section 2.2.1.1, drought and high salt concentration in the soil pose osmotic pressures to plants, limiting their productivity and survival (Kaur and Gupta, 2005; Qin, et al., 2011; Rodriguez, et al., 2005; Xiong and Zhu, 2002). Water loss is most commonly compensated for by increased water transport, closing of stomata (to prevent exit of water from leaf surfaces), and osmotic adjustments in the stems and roots (Chaves, et al., 2003; Xiong and Zhu, 2002). These responses are mediated by phytohormones, primarily by ABA; as well as by other signalling systems such as ROS and MAPK pathways (Ozfidan, et al., 2012),

Many CPKs respond to drought and salt stress (see Table 2.1, first and second columns), as indicated by an increase or decrease in transcript accumulation or protein levels upon exposure to these stresses. CPKs that are upregulated by drought and/or salt stress include AtCPK4, 6, and 11,  OsCPK5, 7, and 10 and TaCPK1 and 9 from group I; AtCPK23 and OsCPK12 from group IIa; OsCPK15 from group IIb;  OsCPK21 from group IIIa; AtCPK10, AtCPK32 and TaCPK19 from group IIIb; and TaCPK6 and 10 from group IV (Li et al. 2008a; Ray et al. 2007; Swarbreck et al. 2008). On the other hand, AtCPK3 and a few rice CPKs, OsCPK13 (group I), OsCPK1 (group IIb), OsCPK16 (group IIIb) and OsCPK18 (group IV) are downregulated by drought and/or salt stress (Ray et al. 2007; Swarbreck et al. 2008). Interestingly, some TaCPKs have contrasting responses to drought and salt. For example, the transcript levels of TaCPK9 (group I) increased in response to drought but decreased in response to salt, while TaCPK18 (group IIa) shows the opposite response pattern (Li et al. 2008a). The transcript response patterns of wheat CPKs were based on semi-quantitative RT-PCR alone and would require further analysis to elucidate the type of response of these CPKs upon drought and salt stress.

A number of reverse genetics and overexpression approaches have been reported looking at the function of CPKs in response to drought and salt. Plants that overexpress group I CPKs OsCPK7, AtCPK4 and 11 have enhanced tolerance to salt and drought (Saijo et al. 2000; Saijo et al. 2001; Zhu et al. 2007). Similarly, AtCPK6 (also from group I) overexpressing plants exhibit greater capability to retain water and are more tolerant to salt and drought stress (Xu et al. 2010). More than 60% of AtCPK6 overexpressing plants were shown to survive two weeks of salt and drought stress, in contrast to wild type plants that mostly died (Xu et al. 2010). OsCPK12 (group IIa) overexpressing plants exhibit increased tolerance to salt stress, less $H_2O_2$ accumulation in leaves, ABA-induced seed growth inhibition and higher expression levels of ROS scavenging enzyme-encoding genes (Asano et al. 2012). AtCPK3 (group IIb) overexpressing plants have increased germination under salt stress and AtCPK3 overexpressing protoplasts have increased AtCPK3 kinase activity in high salt concentration (Mehlmer et al. 2010). AtCPK32 (group IIIb) overexpressing plants showed enhanced ABA and salt sensitivities during germination and promoted the expression of ABF4-regulated genes and ABA responsive genes rd29A, rab18, and rd29B (Choi et al. 2005). AtCPK10 (group IIIb) overexpression mutants have enhanced tolerance to drought (Zou et al. 2010). AtCPK10, 30 and 32 have also been shown to interact with ABF4 (ABA-responsive element binding factor), supporting the role of these CPKs in ABA signalling (Choi et al. 2005). In general, these findings show that plants overexpressing certain CPKs are more tolerant to drought and salt stresses.

In contrast, when these CPKs are knocked-out, plants appear to be more susceptible to drought and salt. AtCPK3 T-DNA insertion (TDI) mutant plants are salt-sensitive and have decreased germination rate under salt stress (Mehlmer et al. 2010). OsCPK12 loss-of-function mutants (retrotransposon and RNAi-silenced) were more sensitive to salinity (Asano et al. 2012). TDI mutants where the AtCPK4 and 11 genes have been disrupted exhibit ABA insensitivity in seed development and stomatal movement, salt insensitivity in seed germination, and decreased tolerance of seedlings to salt stress (Zhu et al. 2007). The double mutants *cpk4-1/cpk11-1* and *cpk4-1/cpk11-2* displayed more severe ABA insensitivity (Zhu et al. 2007). Guard cells of AtCPK3 and 6 TDI knockout mutant plants are impaired in the ABA and $Ca^{2+}$ activation of slow-type anion channels and $Ca^{2+}$ permeable channels, which results in reduced stomatal closure when challenged with osmotic stress (Mori et al. 2006). AtCPK10 TDI knockout mutants are more sensitive to drought (20-day withholding irrigation on 1-week old seedlings) than WT and are insensitive to ABA induction of stomatal closure (Zou et al. 2010). Interestingly, TDI knockout

mutants of two group IIa CPKs, AtCPK21 and 23 (which have 81% identical aa sites) have been reported to be more tolerant to drought and salt stress. In contrast, overexpression of these two genes in plants brings about susceptibility to these stresses and increased stomatal apertures (AtCPK23) or show accumulation of stress marker genes DREB1a, COR15A, Rd29A upon mannitol treatment (AtCPK21) (Franz et al. 2011; Ma and Wu 2007). This information suggests that AtCPK21 and 23 may have important roles in negative feedback pathways in response to drought and salt stress tolerance. Therefore, certain AtCPKs appear to participate in drought and salt stress responses in plants; some are important in developing stress tolerance, while some are involved in the negative feedback pathways.

**Extreme temperatures**

Cold and heat stress also apply osmotic pressure to plants, particularly as extreme temperature affects water conservation and cell water potential. Both have similar physiological effects to drought, as cold hinders water movement between cells and tissues, while high temperature and humidity increases transpiration and evaporation.

Some CPKs have been reported to respond to cold and heat. In response to cold, the transcript accumulation of OsCPK7 and13 (group I), TaCPK5 (group IIa), OsCPK21 (group IIIa), TaCPK7, 12 and 15 (group IIIb) and OsCPK4 (group IV) have been shown to increase while TaCPK1 and 4 (group I), TaCPK3 (group IIIb) and TaCPK6 (group IV) were shown to decrease (Li et al. 2008a; Ray et al. 2007). AtCPK3 transcript accumulation is not affected by cold or heat (Swarbreck et al. 2008), although HA-epitope tagged CPK3 in protoplasts showed increased protein kinase activity 15 min after treatment with cold (4°C) and heat (37°C) (Mehlmer et al. 2010). In addition, AtCPK3 phosphorylates a heat shock factor HsfB2a, which then promotes transcriptional activation of a plant defensin gene PDF1.2. OsCPK13 transcript and protein are highly abundant in cold-tolerant rice varieties and the gene confers cold tolerance when overexpressed (Abbasi et al. 2004; Komatsu et al. 2007; Ray et al. 2007). Interestingly, however, its transcript and protein accumulation are suppressed by ABA, drought and salt stress (Abbasi et al. 2004). Similarly, plants that overexpress OsCPK7 are more tolerant to cold (Saijo et al. 2000; Saijo et al. 2001; Zhu et al. 2007). There are very few studies that have focused on the involvement of CPKs in heat or cold stress responses. Further research investigating the homologues of the CPKs mentioned here in various crop plants may be beneficial to agriculture.

**Wounding**

There are very few studies that have investigated the role of CPKs in response to wounding, apart from transcript profiling research. AtCPK3 (group IIb) transcript accumulation increases upon wounding (3-12 hr postwounding in roots and 30 mins to 24 hr in shoots), while tomato CPKs LeCPK1 (group IIa) and LeCPK2 (group IV) appear to participate in rapid and systemic wound response and desiccation (Chang et al. 2009; Chico et al. 2002). Likewise, ZmCPK11 (group I) transcript accumulation increases in leaves at 1, 3, 6 and 24 hours postwounding (Szczegielniak et al. 2005). Neighbouring leaves also have elevated ZmCPK11, indicating a systemic response (Szczegielniak et al. 2005).

### 2.3.3.3 Response to biotic stress

The detection of MAMPs or PAMPs by plant transmembrane PRRs results in signalling pathways directed to fight against the pathogen, such as the release of ROS and NO, intracellular pH changes, and synthesis of antimicrobial products or pathogenesis-related (PR) proteins (Cohn et al. 2001) (Figure 2.2). The detection of MAMPs leads to specific $Ca^{2+}$ oscillations, which are then decoded by $Ca^{2+}$-sensors and/or responders that activate plant defence signalling pathways (as described in Chapter 2.2.2). How CPKs are specifically involved in these responses is currently unclear, although a number of CPKs have been shown to be up- or down-regulated in response to various infections and/or pathogen elicitors (Table 2.1). Moreover, it has been reported that CPKs may have positive roles in plant immune responses, particularly in MAMP signalling and MAPK activation cascades (Boudsocq et al. 2010). These responses are discussed briefly in the following sections.

**Bacteria and flg22**

In response to bacterial infection by *Pseudomonas syringae*, some Arabidopsis CPKs show increases in transcript accumulation based on microarray data (Table 2.1). These include AtCPKs 4, 5 and 6 (group I), AtCPK9 and 15 (group IIa), AtCPK3 (group IIb), AtCPK7, 10 and 32 (group IIIb) and AtCPK28 (group IV). Mesophyll protoplasts with constitutively active group I CPKs AtCPK4,11, 5, 6, and 26 (AJ and CAD domains deleted) demonstrated five to twenty-fold increase in promoter activity of the NDR1/HIN1-LIKE 10 gene (NHL10), a gene that responds significantly to bacterial flagellin (flg22) (Boudsocq et al. 2010). Constitutively active group IIIB CPKs AtCPK10, 30, 13 and 32 can also increase the promoter activity of this gene more than five-fold (Boudsocq et al. 2010). In the same study, constitutively active AtCPK3 (group IIb) also

induced NHL10 promoter activity more than ten-fold; however, the authors considered this activity to be caused by the high endogenous concentrations of AtCPK3 and not due to flg22 response. Further experimental evidence is required to support this, because based on immunocomplex kinase assays in protoplasts, AtCPK3 kinase activity increased within 15 min of treatment with flagellin (Mehlmer et al. 2010).  Double (*cpk5/cpk6*), triple (*cpk5/cpk6/cpk11)* and quadruple TDI knockout mutants (*cpk4/cpk5/cpk6/cpk11)* of AtCPKs 4, 5,6 and 11 have shown reduction in oxidative burst induced by flg22 and transcript levels of flg22-induced genes PHI-1, NHL10, PER62 and PER4, as well as increased susceptibility to *Pseudomonas syringae* pv. *tomato* (tested in double and triple mutants only) (Boudsocq et al. 2010). TDI knock-out mutants of AtCPK1 are more susceptible to infections by *P. syringae* (Coca and San Segundo 2010). These CPKs therefore appear to be important regulators of plant defence against bacterial infection.

**Fungi and elicitors**

The transcript accumulation of Arabidopsis, rice and wheat CPKs change in response to fungal infections. In Arabidopsis, AtCPK1, 4, 5 and 6 (group I), AtCPK9 and 15 (group IIa), AtCPK3 (group IIb), and AtCPK 7, 8 10 and 32 (group IIIb) showed increased transcript accumulation from 2 to 7 dpi with the fungus *Erysiphe* (Swarbreck et al. 2008). In rice, OsCPK4 and 18 (group IV) appear to be involved in the early response to a symbiont, *Glomus intraradices*, as they are transcriptionally activated by inoculation (presymbiotic phase) and upregulated by secreted molecules from this fungus (Campos-Soriano et al. 2011). OsCPK9, which is closely related to AtCPK10 and 30 (group IIIb) showed increased transcript accumulation 12-24 hr after infection with rice blast fungus (Asano et al. 2005). In wheat, the transcript levels of TaCPK2, 4 and 11 (group I), TaCPK3, 7, 12, 15, and 19 (group IIIB) and TaCPK10 (group IV) increased in response to *Blumeria graminis tritici* (powdery mildew) infection (Li et al. 2008a). In contrast, TaCPK1 showed a decrease in transcript accumulation.

Knocking-out certain CPK genes causes increased susceptibility to fungal infections, while overexpression of some CPK genes confers tolerance and/or resistance to fungi. For example, TDI knock-out mutants of AtCPK1 were more susceptible to infections by fungi such as *Fusarium oxysporum* and *Botrytis cinerea* (Coca and San Segundo 2010). Conversely, plants that overexpress AtCPK1 were less susceptible to both pathogens compared to WT plants and exhibit accumulation of SA together with constitutive expression of SA-regulated pathogen

defence genes. Interestingly, AtCPK2, which is closely related to AtCPK1 (one of the paralogue pairs, having 81% identical aa sites) did not respond to fungal elicitors, although this was only based on transcript accumulation data and no further protein or mutation experiments were done to support this (Coca and San Segundo 2010). AtCPK1 is highly homologous to a maize CPK, ZmCPK10, which was reported to respond to *Fusarium moniliforme* infection (based on transcript accumulation), and was present in cell types where the *PRms* (pathogenesis-related protein in maize) mRNA was also present, suggesting a role for ZmCPK10 in the regulation of pathogen responsive genes (Murillo et al. 2001). However, OsCPK12 overexpressing plants showed increased susceptibility to a blast fungus *Magnaporthe grisea* (Asano et al. 2012). This was explained by the authors as a result of the involvement of OsCPK12 in the higher expression levels of ROS scavenging enzymes, causing a reduction of ROS species, thus hindering localised cell death to prevent pathogen spread.

**Insects**

Only AtCPK3 and 13 have been reported to participate in plant responses to insect attack. During insect herbivore attack, Arabidopsis CPK3 and CPK13 mutants had lower transcript levels of PDF 1.2, compared to WT plants, suggesting negative feedback regulation of wounding and insect response (Kanchiswamy et al. 2010). This response was investigated in TDI knockout mutant lines of AtCPK18, 7, 8, 10, 5, 2, 20, 11, 33, 19, 21, and 22, but these mutants did not show significant change in transcript levels of PDF1.2 compared to WT. Moreover, *in vitro* kinase assays showed that AtCPK3 phosphorylates ATL2, a member of RING-H2 zinc finger family that functions as E3 ubiquitin ligases and is a potent regulator of PDF1.2 transcription (Kanchiswamy, et al., 2010). Despite the potential role of CPKs in plant responses to insect attacks as shown in these findings, very few studies have been done in relation to this. Exploring the function of the homologues of AtCPK3 and 13 in other plants in relation to insect attack may benefit agricultural science.

**Viruses**

No virus response experiments have yet been reported focusing on CPKs. Information regarding CPK involvement in virus infections only comes from general transcript profiling studies investigating the expression of various genes in response to a certain virus. General transcriptome analysis of *Plum pox virus* (PPV)–infected Arabidopsis plants showed that

AtCPK32 and 14 transcript levels increase by 3.32 and 5.29-fold, respectively, at 17 dpi (Babu et al. 2008). AtCPK28 and 18 appeared to respond to viral infection, as indicated by changes in their transcript levels in response to PPV; AtCPK28 was upregulated by 2.63-fold while AtCPK18 was downregulated 2.7-fold 17 dpi (Babu et al. 2008).

### 2.3.4 CPKs from algae and non-vascular plants

CPK sequences have been identified and analysed in a small number of green alga such as *Ulva compressa* (Contreras-Porcia et al. 2010), *Ventricaria ventricosa* (Sugiyama et al. 2000), *Dunaliella tertiolecta* (Yuasa and Muto 1992) and *Chara coralline* (McCurdy and Harmon 1992), as well as in the moss *Funaria hygrometrica* and liverwort *Marchantia polymorpha* (Nishiyama et al. 1999). In the marine alga *U. compressa* cultivated with 10 µM copper, CPK transcripts increased more than two-fold from 3 to 5 days suggesting a possible role for CPKs in copper acclimatisation or tolerance (Contreras-Porcia et al. 2011). A 52-KDa CPK from the unicellular green alga *V. ventricosa* has been suggested to be involved in $Ca^{2+}$-mediated wound response by simultaneously participating in the organisation and contraction of F-actin (Sugiyama et al. 2000). A CPK was isolated and characterised from *D. tertiolecta,* a salt-tolerant green alga (Yuasa and Muto 1992). This CPK phosphorylated casein, myosin light chain and histone IIIS, appeared to be bound to microsomes, and phosphorylated microsomal proteins. A CPK in the green alga *Chara* was also found to be involved in the regulation of cytoplasmic streaming, being localised in the subcortical actin bundles, organelle surfaces and components of streaming endoplasm (McCurdy and Harmon 1992). In the moss *F. hygrometica,* a 518 aa CPK has been identified and exhibited an increase in transcript accumulation within 24-48 hr of nitrogen, phosphorus or sulphur starvation (Mitra and Johri 2000). Only the CPKs identified from *M. polymorpha* have so far been included in CPK comparative and evolutionary analyses (Zhang and Choi 2001). These CPKs have been shown to be encoded by one gene that generates two mature mRNAs. One of the splicing variants is preferentially accumulated in the liverwort's male sexual organ (Nishiyama et al. 1999). CPKs from algae and non-vascular plants appear to function in cytoskeletal organisation, osmotic pressure responses and development.

### 2.3.5 A summary of CPK activity

In summary, CPKs are important in signalling cascades in response to stresses with osmotic effects (drought, salt, extreme temperatures), mechanical stress and pathogen attack, and in processes that require cellular or cytosolic movement. Each stimulus has a specific array of CPKs that are up- or down-regulated. There is a high degree of conservation among CPK sequences, which correlates with the functional redundancy and overlap between closely related CPK sequences, as well as CPKs belonging to different evolutionary groups. Each evolutionary group has abiotic, biotic and development-responsive members, which suggests that the specificity of CPKs to a stimulus and to a type of response might possibly be influenced by very few differences in the amino acid sequence. On the other hand, some CPK isoforms, particularly those that arose through recent duplication events, may have similar or contrasting (despite having a high degree of similarity) expression levels in response to a given stimulus or process. While most CPKs are involved in ABA-mediated stress responses, some CPKs are also involved in MeJA signalling, SA responses and MAPK cascades. Most of the information regarding CPK function is based on transcript accumulation measured using various techniques, and there is a need to support or validate these data through protein activity and interaction-based and mutation-based experiments. Moreover, most of the fully supported information is about abiotic stress and developmental responses; very few studies have reported the involvement of CPKs in pathogen defence and insect attack responses. Certain CPKs have been shown to be important in flg22 signalling (Boudsocq et al. 2010) and confer tolerance to bacterial infection when overexpressed. Little is known about CPK responses to viral infections. Characterisation of CPKs that can confer protection or resistance against infection will be of enormous value to agriculture, providing potential opportunities to develop novel strategies to protect against disease.

## 2.4 How will an understanding of the most conserved members of the CPK gene family benefit agriculture and the plant sciences?

Based on the information presented above, it appears that CPKs are important in $Ca^{2+}$-mediated signal transduction pathways involved in cellular/cytoskeletal movement, in membrane transport systems to maintain cellular integrity and water potential, and in regulation of stress response genes. CPKs play crucial roles in processes such as cellular elongation (pollen tube growth and root elongation), stomatal closure, cellular division in meristematic zones, cytoplasmic streaming, MAMP detection and phytohormone response, which all involve cellular/cytoskeletal

movement and transport of molecules between the plasma membrane. As most of the CPKs are localised to the plasma membrane, but are not transmembrane proteins, it is possible that they mediate the relationship between transmembrane transport or receptor proteins and cytosolic proteins. They also phosphorylate transcription factors that regulate stress-response genes. CPKs therefore have diverse roles in cellular signalling pathways that are vital to plant development and survival.

There are three main reasons why the identification and functional characterisation of the most conserved CPK(s) may be beneficial to plant sciences and agriculture. Firstly, it has been maintained through evolution to a greater degree than the other CPK isoforms, indicating that its sequence elements have been unchanged to ensure function has been maintained. Studying the most conserved CPK(s) may lead to the discovery of elements that are important in plant responses to different kinds of abiotic and biotic stresses. Secondly, the complete 3D structure of any plant CPK has not yet been reported. As previously noted, the 3D structure of proteins provides information regarding their molecular function and may explain how these proteins are activated and/or how they interact with their substrates or other proteins. The 3D structures of certain protist CPKs have been published recently and can be used as a model to predict the 3D structure of plant CPKs; however, the protist sequences do not have a high degree of amino acid sequence similarity with most plant CPKs. There is a greater chance that there are more sequence similarities between protist CPKs and the most conserved plant CPK(s) than between protist and other plant CPK isoforms. Lastly, the most conserved member(s) of this multigene family has greater degree of similarity with other CPK isoforms within and between different plant species from lower to higher plants. As there are many CPK isoforms, the most conserved CPK may be a good target for identifying or developing stress tolerant or resistant plants, that is applicable to a broad range of plant species.

# Chapter Three

# How did CPKs diversify and what is the most conserved CPK group in plants?

## 3.1 Introduction

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 available CPK sequences 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 (Billker et al. 2009; Nagamune and Sibley 2006) and a small number of plants, namely, Arabidopsis (Cheng et al. 2002; Hrabak et al. 2003), rice (Asano et al. 2011), and *T. 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 (Boudsocq et al. 2012; Harmon et al. 2000; Hrabak et al. 2003).

CPKs studied to date have different tissue and cellular localisations, 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 completion of several plant genomes, this chapter 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. As this chapter includes an analysis of reported CPK functions, some of the information presented in Chapter 2 is revisited in this chapter in light of the evolutionary groupings of CPKs.

This chapter aims to answer the main question: How did CPKs diversify and what is the most conserved CPK group in plants? To address this question, this chapter has the following specific objectives: (1) to carry out data mining of all available CPK sequences from the genome of representative lower and higher plant species; (2) to carry out a phylogenetic analysis of the CPKs identified and identify the most conserved members; and (3) to determine the correlation between CPK sequence evolution and CPK functional diversification from lower to higher plants.

The majority of the work carried out in this chapter has been published in Plant Physiology in 2014, entitled "Calcium-Dependent Protein Kinases in Plants: Evolution, Expression and Function" (Valmonte et al. 2014).

## 3.2 Materials and Methods

### 3.2.1 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 (AtCPKs). AtCPKs 1, 21, 8, and 16 have the highest percent pair-wise 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 andv4 u10.2)*, P. patens* (v1.1)*, S.moellendorfii* (v1.0)*, P. sitchensis* (expressed sequence tags (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) and *Z. mays* (v4 and v5a.59), *A. thaliana* (The Arabidopsis Information Resource (TAIR) r10)*, V. vinifera* (8x and 12x 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: (1) a cut-off BLAST score of at least 250 and an E-value of $e^{-2}$ or less; (2) presence of the five CPK domains: N-VD, PK, AJ, CAD and CT domain; and (3) having one to five 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 Appendix 1. When alternative splicing variants were present, only one protein sequence was chosen (one with the longer sequence) to be included in the analysis.

### 3.2.2 Notes regarding Nomenclature

As mentioned in section 2.3.1, this thesis followed the gene nomenclature used by Hrabak et al. (2003). In cases where the CPK gene has been previously named by other authors, for consistency the abbreviation "CPK" was used for all sequences but the assigned numbers were retained (e.g. OsCDPKs to OsCPKs). 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.

### 3.2.3 Multiple sequence alignment and phylogenetic analysis

Multiple sequence alignments and phylogenetic analyses were carried-out using GeneiousPro 5.6 (Kearse et al. 2012). 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, AJ and CAD (Figure 2.3, section 2.3.1). For the phylogenetic analysis of all identified CPKs, distance (Jukes-Cantor model) and likelihood (Whelan and Goldman (WAG) model) trees were constructed using the neighbour-joining (NJ; Geneious Tree Builder) and maximum likelihood (ML; PhyML) methods, respectively, with 1000 bootstrap replicates. Trees were viewed and coloured using FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

To compare trees constructed using full CPK sequences and the trees constructed using trimmed sequence, multiple sequence and tree construction was carried out using full CPK sequences. As the ClustalW program was not appropriate for untrimmed sequences, another method called MUSCLE was used to align the sequences. Only an NJ tree was constructed as the program did not function well in all of the attempts made to construct an ML tree using full CPK sequences.

Bayesian Markov chain Monte Carlo (MCMC) analysis in the software Bayesian Evolutionary Analysis Sampling Trees (BEAST) (Drummond and Rambaut 2007) was used to construct a tree with relaxed molecular clock. Calibration points for the most common recent ancestors (MRCAs) were set at 1700 million years ago (MYA) for apicomplexan (Hedges 2002; Hedges et al. 2004) and 1400 MYA for green algae (Yoon et al. 2004). The Tree Prior used was the Yule model.

The evolutionary relationship between CPKs with reported biological or molecular function was also determined using phylogenetic analysis. CPK sequences for which published sequence information was available were aligned with ClustalW using GeneiousPro 5.6. An NJ tree was constructed using Geneious Tree Builder, with default parameters.

### 3.2.4 $K_a/K_s$ ratio analysis

The ratio of nonsynonymous ($K_a$) to synonymous ($K_s$) nucleotide substitution rates ($K_a/K_s$ ratio) were calculated and the $K_a/K_s$ tree was generated using the calculation tool from Universitetet I Bergen (http://services.cbu.uib.no/tools/kaks). A $K_a/K_s$ calculation was carried out using the conserved domains of CPKs (PK, AJ and CAD). Calculation could not be undertaken within the N-VD and CT domains as these regions are highly variable.

### 3.2.5 Gene structure analysis

The intron/exon organisation 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/).

### 3.2.6 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). Second, experimental data on specific CPK responses to biotic and abiotic responses, hormones, developmental signals and other genes were collated from the literature.

### 3.3 Results

#### 3.3.1 How did CPK sequences diversify?

##### *3.3.1.1 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. As described in section 3.2.1, this included two green alga, *V. carteri* and *C. reinhardtii* (Fukuzawa et al. 2008); a bryophyte, *P. patens* (Rensing et al. 2008); a pteridophyte, *S. moellendorfii* (Banks et al. 2011); a gymnosperm, *P. sitchensis* (Ralph et al. 2008); four monocots, *O. sativa* (Asano et al. 2005), *T. aestivum* (Li et al. 2008b), *S. bicolour* (sorghum) (Paterson et al. 2009) and *Z. mays* (maize) (Schnable et al. 2009); and six eudicots, *A. thaliana* (Cheng et al. 2002), *V. vinifera* (grape) (Jaillon et al. 2007)*, G. max* (soybean) (Schmutz et al. 2010)*, P. trichocarpa* (poplar) (Tuskan et al. 2006; Zuo et al. 2013), *C. papaya* (papaya) (Ming et al. 2012), and *S. tuberosum* (potato) (Xu et al. 2011) (Figure 3.1 and Appendix 1). 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* (Du et al. 2013; Engstrom 2011; Goodstein et al. 2012; Rouard et al. 2011; Zhang et al. 2012). On the other hand, *V. 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 (refer to section 3.2.1 for 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 aa; except for two putative CPKs from *C. reinhardtii* which were 1042 and 1801 aa 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 domains, and occasionally due to the number of EF-hands in the CAD. 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 localisation and functional specificity.

The total numbers of CPK genes within the genomes examined were consistent with the pattern of genome duplication and polyploidisation events that have occurred through plant evolution (summarised in Figure 3.1). The divergence between red algae and Viridiplantae (green algae and land plants) occurred about 1200–1600 MYA, while the split between green algae and land plants happened 700–1000 MYA as estimated in several studies (Hedges et al. 2004; Parfrey et al. 2010; Yoon et al. 2004; Zimmer et al. 2007). 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: (1) the ancestral seed plant (ζ) and (2) ancestral angiosperm (ε) WGD events (Jiao et al. 2011a); (3) the ancestral eudicot triplication event (γ) (Bowers et al. 2003; Jaillon et al. 2007; Jiao et al. 2011a; Tang et al. 2008); and (4) the ancestral monocot (σ) WGDs. Consistent with these events, green algae had the least number of CPKs (eight??? in *C. reinhardtii* and ten in *V. carteri*), whereas angiosperms generally had more CPKs than other land plants, except for *V. vinifera, C. papaya* and *T. aestivum* (Figure 3.1). It must be noted, however, that *C. papaya* and *T. aestivum* genomes had not yet been completely annotated. *V. vinifera*, on the other hand, has the least number of CPK genes among fully or nearly completely sequenced eudicot genomes (17 CPK genes), most likely 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 Figure 3.1). 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). *G. max* had the most CPKs among eudicots, possibly due to several rounds of polyploidisation within this species (Gill et al. 2009; Schmutz et al. 2010). Independent genome duplications in Fabaceae, Solanaceae, Brassicaceae (α and β events) and *Populus* (Gill et al. 2009; Jiao et al. 2011a; Magallon and Castillo 2009; Soltis et al. 2009; Tuskan et al. 2006) also corresponded to the increase in CPK genes present in the representative species (47, 27, 34 and 30 CPK genes, respectively).
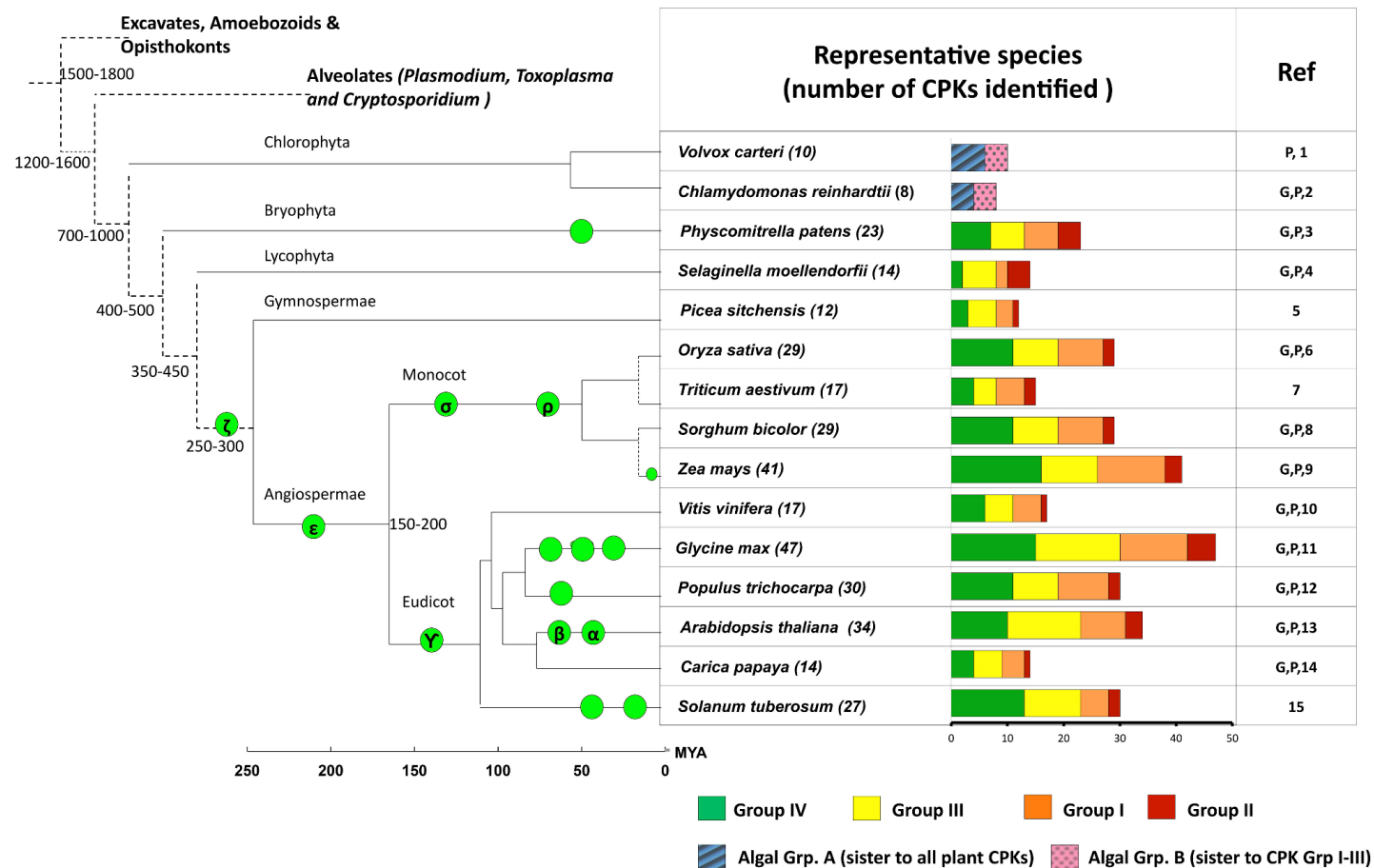
**Figure 3.1 Distribution of CPK evolutionary groups among the representative genomes.** Circles indicate whole genome duplication (WGD) events that occurred within lineages, with some events designated with Greek symbols by previous literature. Branches in broken lines denote that the diversification ages are not drawn to scale. References for sequence mining: (G) – GreenPhylDB Database (http://greenphyl.cirad.fr/v2/cgi-bin/index.cgi), (P) – Phytozome Database (www.phytozome.net) Genome versions: (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 TCs from TIGR r2 and Li et al. (2008), (8) S. bicolor v1.4, (9) Z. mays v4 and v5a.59, (10) V. vinifera 8x and 12x 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.

### 3.3.1.2 CPK evolution from algae to angiosperms

Phylogenetic analyses were undertaken using the amino acid sequences of only the conserved region, consisting of PK, AJ and CAD domains (previously illustrated in Figure 2.3, section 2.3.1). The N-VD and CT were excluded from multiple sequence alignments (Appendix 2) 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 NJ and ML trees  appeared similar to that of Li et al. (2008b) and Boudsocq and Sheen (Boudsocq et al. 2012). A condensed view of the ML tree generated from the alignment of the conserved region is shown in Figure 3.2, while the detailed topology for each evolutionary group is shown in Appendices 3 to 6. The NJ tree constructed is shown in Appendix 7. To determine whether the N-VD and CT of the CPKs affect the evolutionary groupings, full-length CPK sequences were also aligned, using another method called MUSCLE (Edgar 2004) (Appendix 8). 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 group are shown in (Appendix 9 to 13). Despite similarity in topology, the trees constructed using full CPK sequences generally had lower bootstrap values and have shown numerous cases of polytomy. For this reason, the succeeding analyses focused on the trees constructed using trimmed CPK sequences.

As illustrated in Figure 3.2, 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; and Group IV, 0.858 aa 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 (Figure 3.2, 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*), 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 CAD region (Appendix 2 and 8). 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.

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 (Figure 3.2, 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 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 (for example, 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 (Figure 3.2).

Monocot and eudicot CPKs form several clusters within the evolutionary groups. Group IIIa and IV both had only one monocot-eudicot cluster (Figure 3.2, cluster IV and IIIa), while Groups I, IIa, IIb and IIIb had two to three monocot-eudicot clusters (cluster 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 aa sequence (80 to 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 with the increase in CPK gene numbers among angiosperms.
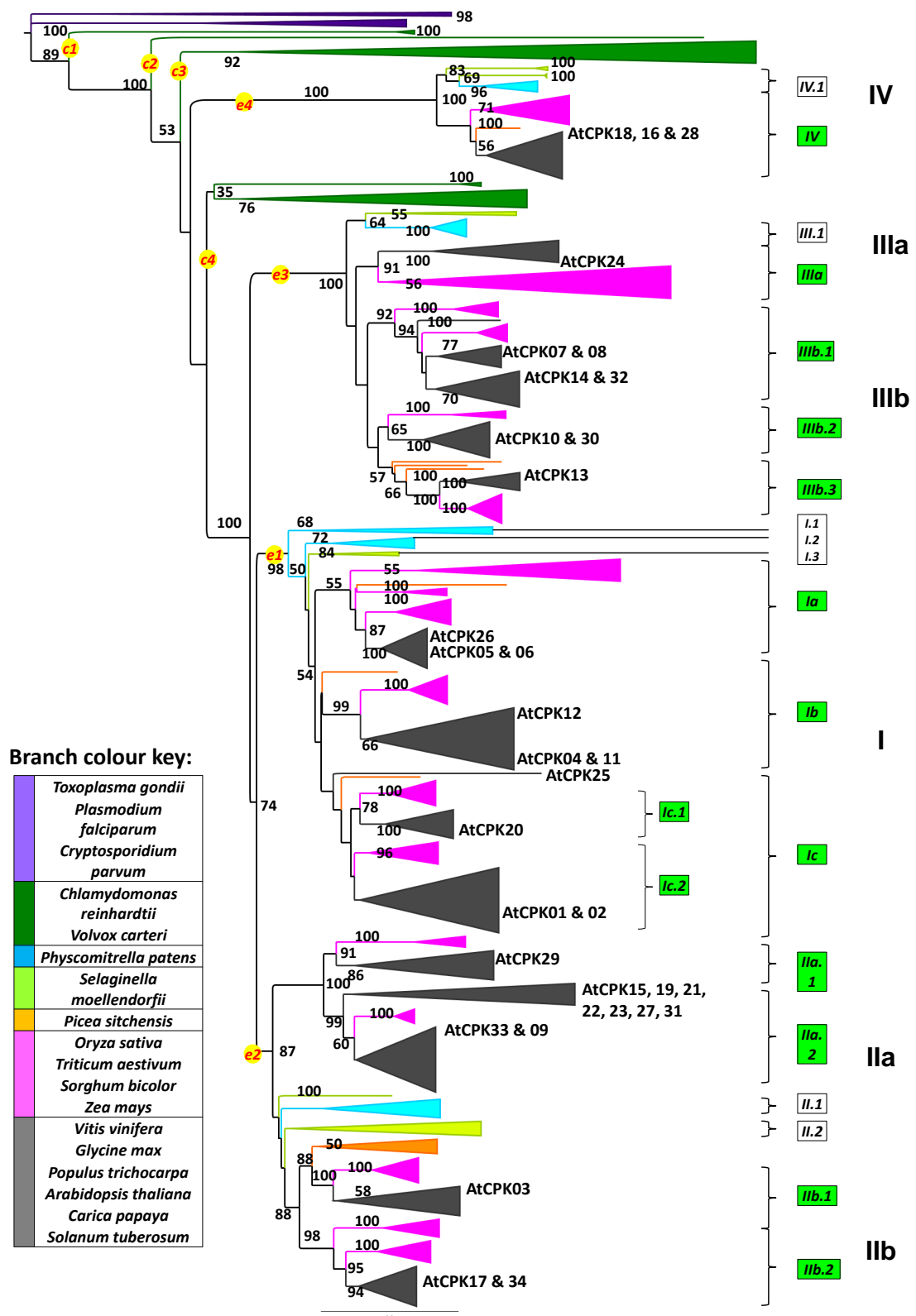
**Figure 3.2 Maximum Likelihood tree of CPKs from algae to higher plants. Phylogenetic tree of CPKs from algae to higher plants.** The tree was constructed by Maximum Likelihood method, with 1000 bootstrap replicates using GeneiousPro 5.6 software (Drummond et al. 2010). A total of 352 plant (including green algae) and 5 apicomplexan (as outgroup) CPK protein sequences were included in this analysis. Branch colours match the species colours in the left box. Branching points as described in-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. NJ tree show similar general topology (Appendix 7).

To estimate the timing of CPK diversification, a tree with a relaxed molecular clock was constructed by performing an MCMC analysis used by the program BEAST (Drummond and Rambaut 2007) (Figure 3.3). Similar to the NJ and ML trees (Figure 3.2 and Appendices 3-6, 9-13), 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 (Figure 3.2), 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.

**Figure 3.3 Bayesian evolutionary tree of CPKs with relaxed molecular clock.** The tree was constructed using BEAST software (Drummond and Rambaut, 2007), with calibration points at 1700 million years ago (MYA; Hedges, 2002) for apicomplexan and 1400 MYA for green algae MRCAs (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 colours match the species colours in Figure 3.2.

### 3.3.1.3 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 the intron/exon patterns were similar between CPKs belonging to the same evolutionary group and taxon, but different between taxa (Figure 3.4). 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 (Figure 3.2, Appendix 7 and 9), which shows that group IV CPKs form a separate clade of earlier lineage.
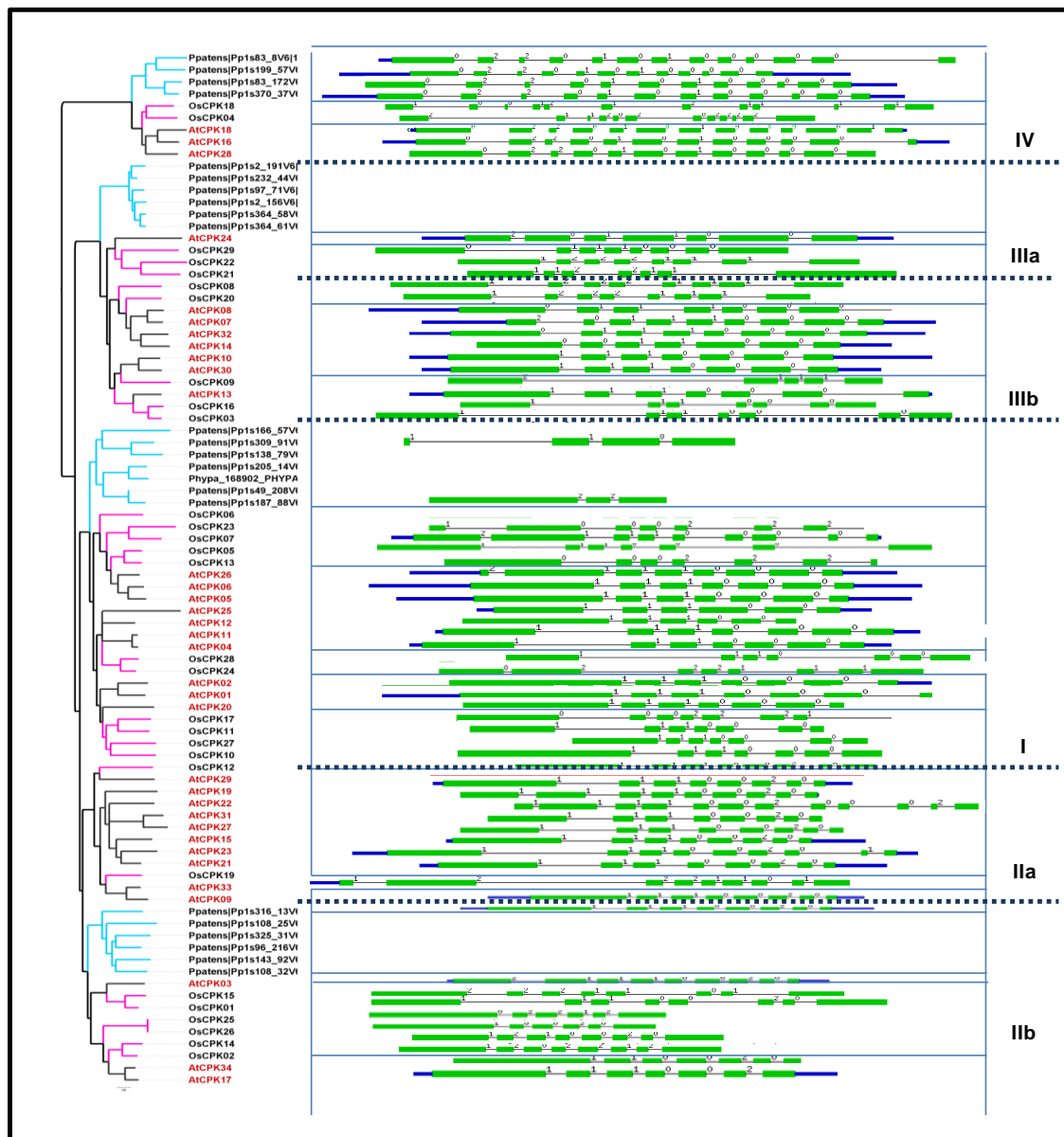
**Figure 3.4 Gene structure analysis of representative CPKs.** Intron/exon analysis was performed with CPK genes from representative moss (*P. patens*), dicot (*A. thaliana*) and monocot (*O. sativa*). Intron/exon patterns are similar between CPKs belonging to the same evolutionary group and taxon, however patterns between taxa are different. Intron/exon patterns were visualised using the online tool Gene Structure Display Server (http://gsds.cbi.pku.edu.cn). No intron/exon information could be gathered for some of the *P. patens* CPK sequences.

### 3.3.2 Functional diversification of plant CPKs

#### *3.3.2.1 Function and phylogenetic grouping did not correlate except in some subgroups*

The functional importance of CPK gene expansion and diversification events along 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 characterisations undertaken on individual CPKs from different species such as *V. vinifera*, (Yu et al. 2006), *Z. mays* (Estruch et al. 1994; Murillo et al. 2001; Szczegielniak et al. 2005; Takezawa et al. 1996), *S. tuberosum* (Gargantini et al. 2009; Giammaria et al. 2011; Kobayashi et al. 2007; Raices et al. 2001), *S. lycopersicum* (Chang et al. 2011; Chang et al. 2009; Rutschmann et al. 2002), 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 Figure 3.5 (detailed information shown in Appendices 14-16). 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 Figure 3.6. It should be noted that CPKs in this tree are not necessarily included in the broadly sampled phylogenetic tree (Figure 3.2) as some of the CPKs with functional information come from genomes with incomplete or no data available. Likewise, some CPKs that have been identified in genomes do not have functional data.

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 (Figure 3.5, Appendix 14-15, 17). Most abiotic stress studies have focused on drought and salt conditions, with very few studies having focused on extreme temperatures and wounding (Figure 3.5b). The majority of biotic stress data has come from microarray experiments in Arabidopsis and semi-quantitative reverse-transcriptase 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 (Figure 3.5c).

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 (Figure 3.6). Each group (major or subgroup) had CPK genes involved in developmental and stress responses (Figure 3.5). 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 stamen), which indicate developmental function(s) (Figures 3.5 and 3.6, Appendix 17). 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 abiotic and biotic 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 structure of AtCPK12 have a few differences with those of AtCPK4 and 11 (Figure 3.4). Another example within Group IV, AtCPK18, showed decrease in transcript accumulation upon viral infection (Babu et al. 2008), while its closest homologue, 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 (Figure 3.4). An interesting relationship was also observed between likely paralogues OsCPK1 and 15 (Group IIb.1), wherein both CPKs responded to drought and salt, but one was upregulated (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 (Figure 3.6). 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 response: Group Ia (monocots only), IIIb.2, IIIb.3, and small eudicot clades in Ia and Ic.2 (Figure 3.6). Most of the other CPK evolutionary subgroups (Ib, Ic.1, IIa, IIb.1, IIIb.1, and IV) responded to different developmental and abiotic and/or biotic stress signals with no apparent pattern. The orthologous CPK groups mentioned are described briefly in the following sections.
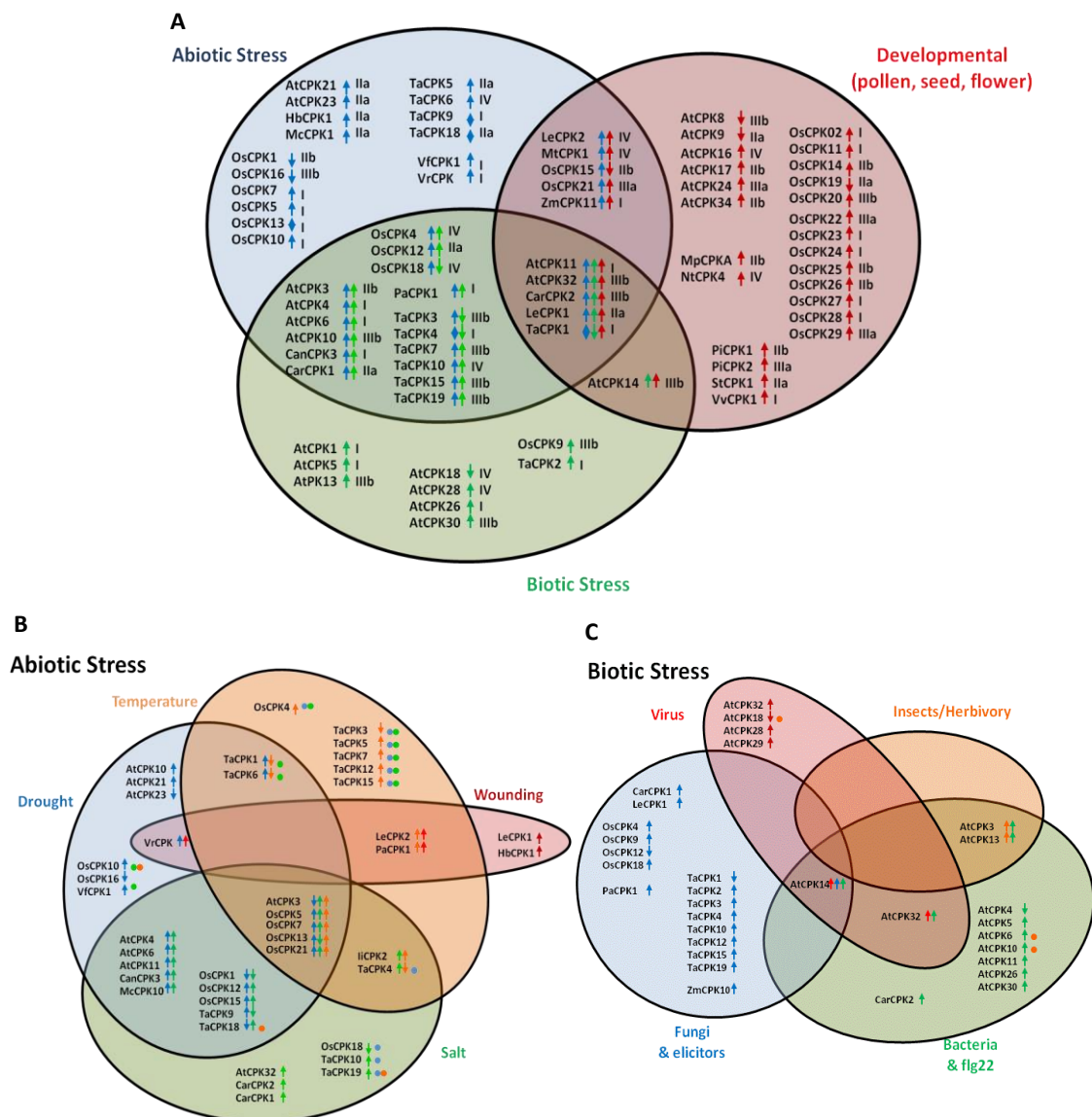
**Figure 3.5 Overview of plant CPK functional information based on the literature** (a) Developmental, abiotic and biotic stress responses. (b) Abiotic stress response. (c) Biotic stress responses. Information includes transcript and protein accumulation, enzyme activity, gene knockout and overexpression experiments. Upward arrows sign indicate up-regulation, while downward arrows 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. Colour of arrows and dots correspond to the font colour of stress. Refer to Appendices 15 and 16 for detailed information and citations.

| 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 Glyma2g48160 (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 |

**Figure 3.6 Plant CPKs with reported function and identified orthologous groups.** The NJ tree was constructed using CPKs for which function has been reported, with 100 bootstrap replicates using GeneiousPro 5.6 software (Drummond et al., 2010). The table highlights potential orthologues within clades that seem to have a functional pattern. Clades are highlighted based on type of response. Colour of boxes correspond to the type of response. Refer to Appendices 16 and 17 for detailed information and citations.

### 3.3.2.2 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 (Figure 3.6). 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), OsCPK 2, 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 (Appendices 14-15,17). Arabidopsis *cpk17* and *cpk34* double mutants had 350-fold reduced pollen transmission efficiency and three-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 depolarisation of pollen tube growth (Zhou et al. 2009). Based on their position in the phylogenetic tree (Appendix 4), OsCPK2 and 14 (85% pairwise aa identity with each other) and OsCPK25 and 26 (99.6% pairwise aa identity) are potential orthologues of AtCPK 17 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 signalling involved in seed development (Morello et al. 2000). Another Group IIb.2 CPK, PiCPK1, from petunia, appeared to have similar functions as 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 (two and three-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 (Figure 3.6) have no specific information available but may also be important in development, although no specific information is yet available. 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.

### 3.3.2.3 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 upregulated 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 orthologue 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 paralogue of TaCPK1 and a potential orthologue 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 (Figure 3.6) may verify this observation, such as the three genes in maize and the four genes in sorghum listed in Figure 3.6. One conifer CPK, Picea_TC127192 belongs this group (Appendix 3). 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.

### 3.3.2.4 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 paralogue of AtCPK10 (86% aa similarity), showed no change in transcript accumulation in the same study. In Group IIIb.3, positive responses to *Blumeria 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 to fungal infection. It may be valuable to study the responses of CPKs within Group IIIb.2 and IIIb.3 in different seed plants (Figure 3.6) against certain fungi that have wide host ranges.

### 3.3.2.5 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 (Boudsocq et al. 2010; Coca and San Segundo 2010; Tsai et al. 2007). However, the closest paralogue of AtCPK1, AtCPK2 (81% similarity in aa) did not appear to respond to bacteria nor 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.

### 3.3.3 What is the most conserved CPK group?

### 3.3.3.1 $K_a/K_s$ analysis in Arabidopsis CPKs

As suggested in section 3.3.1, Group IIb CPKs appears to be the most conserved CPKs as they have the shortest main branch and the lowest average branch length (0.547). To further verify this hypothesis, a $K_a/K_s$ ratio analysis was performed among Arabidopsis CPKs. The $K_a/K_s$ ratio provides an indication of selective pressure for a protein coding gene to be maintained. A $K_a/K_s$ ratio greater than one indicates that there is positive selection, or that there is a driving change to the gene. A ratio of one indicates neutral selection; while a ratio less than one indicates purifying or stabilising selection, or that there is a pressure for the gene to be maintained

(Hasanuzzaman et al. 2013). The $K_a/K_s$ tree generated using the calculation tool from Universitetet I Bergen (http://services.cbu.uib.no/tools/kaks) is shown in Figure 3.7 (detailed figure and analysis in Appendix 18 and 19). The general topology of the $K_a/K_s$ tree minimally resembled that of the AtCPK trees constructed in this thesis and by Sheen et al (2002). While most of the evolutionary groupings were retained, certain branches were different and some genes grouped differently. For example, AtCPK24 (IIIa) grouped together with AtCPK1 and 2 (I) and AtCPK29 grouped together with AtCPK17, 34 and 3 (IIb). Moreover, in this tree it appeared that group II CPKs showed an earlier separation from the other CPK groups (node 32), followed by some members of group I (node 19). Group IIb separated in to one clade (node 3) but showed paraphyly as it included AtCPK29, whereas some members of groups I and IIIa and all members of group IV belonged to one clade, showing polyphyly among group I CPKs and polyphyly among group III CPKs. The possible reason behind this difference is the method used by the $K_a/K_s$ calculation tool, which used binary rooted phylogenetic tree construction method in Newick format, whereas the phylogenetic trees constructed in this thesis were calculated using either NJ or ML methods.

Despite the difference in topology, the $K_a/K_s$ ratios support high conservation among the CPK genes as inferred from the branch lengths in the ML tree. All the $K_a/K_s$ ratios computed for all branches were less than one, ranging from 0.0455 to 0.6682, which supports the inference that in the CPK gene family, conservation has been maintained. From the main node (node 33), the $K_a/K_s$ ratio of the branch towards group II CPKs was lower (0.3011) than that of the other groups (0.3075). The $K_a/K_s$ ratios of the branches towards IIb and some members of IIa (0.1912 and 0.1852, respectively) were comparatively lower than that of the other groups (0.3841, 0.2257, and 0.2304 for groups IIIb, IV/IIIa and I, respectively). The branch towards group IIb CPKs (plus AtCPK29) showed a 0.2063 $K_a/K_s$ ratio, which is lower than the average of all the branches towards group IIa CPKs (0.2918 averaged from 0.2211, 0.1998 and 0.4546). Moreover, the branch that leads to a monophyletic clade for group IIIb, IV, and I were higher than that of group IIb (0.3841, 0.6682, and 0.2304, respectively). These consistently low $K_a/K_s$ ratio values indicate that throughout the evolution of AtCPKs, there was relatively more pressure for group IIb CPKs to be maintained.
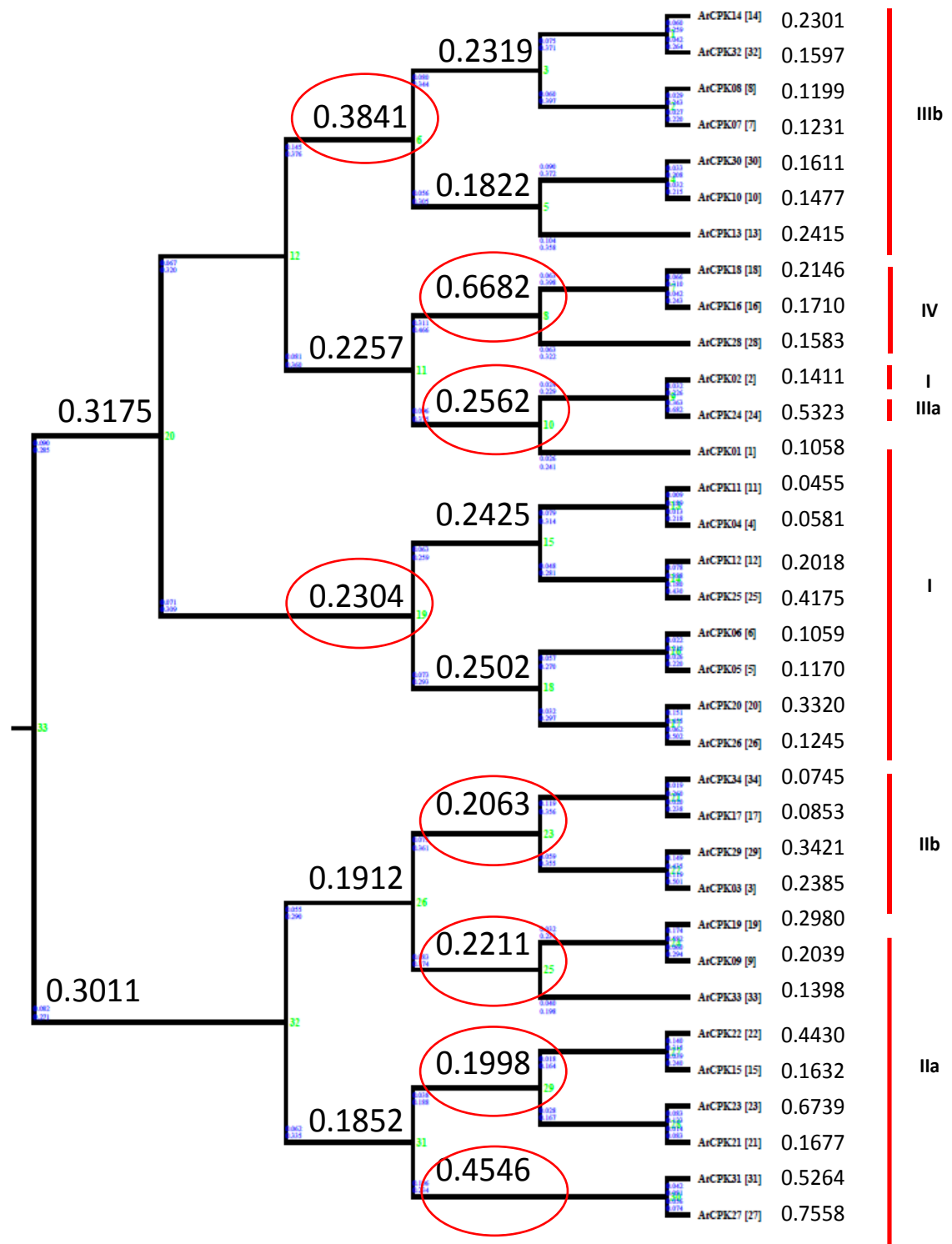
**Figure 3.7 Ka/Ks tree of Arabidopsis CPKs.** The tree was constructed using the $K_a/K_s$ Calculator (http://services.cbu.uib.no/tools/kaks). Decimal numbers in black font indicate $K_a/K_s$ ratios of the branches below them. Numbers in green font indicate node number. Numbers in blue font indicate $K_a$ and $K_s$ values for the corresponding branch. Detailed figure shown in Appendix 18.

### 3.3.3.2 $K_a/K_s$ analysis in Arabidopsis, rice, grape, potato and moss CPKs

To further investigate, the $K_a/K_s$ ratio calculations were also carried out combining Arabidopsis, rice, grape, potato and moss CPKs (Figure 3.8, detailed figure and analysis in Appendix 20 and 21). Similar to the AtCPKs, the general topology of the tree resembles the phylogenetic trees constructed in this thesis and previous authors (Asano et al. 2005; Li et al. 2008a). Most of the evolutionary groups maintained the same groupings, except for group III which formed a paraphyletic group at the second branch at node 117. There was also one case where OsCPK09, which belongs to IIIb, has grouped with members of group IIIa. In this tree (Figure 3.8), it appeared that group I and II separated from group III and IV, instead of group IV or group II initially separating from the three other groups (Figure 3.2 and 3.6). Also, the branching of the subgroups and smaller clades within the groups differed from the previous trees constructed.

This tree also supported the inference that CPKs are highly conserved, as all the $K_a/K_s$ ratios computed for all branches were less than 1, ranging from 0.0455 to 0.89 (Appendix 21), with an average of 0.1353 for all branches. There was one branch that showed a $K_a/K_s$ ratio greater than 1 (1.1357), which is the branch separating StCPK28 and StCPK16. StCPK28 was initially included in the phylogenetic analysis but was later excluded due to its similarity to a CRK rather than a CPK. It was only included in this analysis to further verify its exclusion.

The $K_a/K_s$ tree support the inference made earlier that group IIb CPKs are the most conserved among the evolutionary groups of CPKs. The branch that included all group IIb CPKs as one monophyletic group was the first branch of node 36, with a $K_a/K_s$ value of 0.1101 (Figure 3.8). This value is lower compared to that of the other groups: group I CPKs at node 74 branch 2, with a $K_a/K_s$ ratio of 0.2681; group IIa CPKs at node 36 branch 2 with a $K_a/K_s$ ratio of 0.1704; and group IV CPKs at node 91 branch 1 with a $K_a/K_s$ ratio of 0.5769. The paraphyletic clade that included all group III and IV CPKs at node 117 branch 2 had a $K_a/K_s$ ratio of 0.1815, while the clade that included all group IIIb and some group IIa CPKs at node 116 branch 2 had a $K_a/K_s$ ratio of 0.1484.

**Figure 3.8 K$_a$/K$_s$ tree of Arabidopsis, rice, grape, potato and moss CPKs.** The tree was constructed using the KaKs Calculator (http://services.cbu.uib.no/tools/kaks). Decimal numbers in black font indicate K$_a$/K$_s$ ratios of the branches below them. Numbers in green font indicate node number. Numbers in blue font indicate K$_a$ and K$_s$ values for the corresponding branch. Detailed figure shown in Appendix 20.

| CPK Evolutionary Group | I | IIa | IIb | IIIa/b* | IV |
|---|---|---|---|---|---|
| Ka/Ks ratio on main branch forming one clade | 0.2681 | 0.1704 | 0.1101 | 0.1484 | 0.5769 |

To further support the inferences made, the average $K_a/K_s$ ratios of CPKs in each evolutionary group were also calculated and compared (Figure 3.9, Appendix 21). Among AtCPKs, group IIIa showed the highest average $K_a/K_s$ ratios with an average of 0.5323, followed by group IIa with an average of 0.3714 (Figure 3.9a). Group I, IIb, IIIb and IV relatively had lower $K_a/K_s$ ratios, with averages of 0.1649, 0.1327, 0.1690 and 0.1813, respectively. Group IIb AtCPKs had the lowest average $K_a/K_s$ ratio; however its difference from groups I, IIIb and IV was not significant based on the standard error (SE) of the mean. With the combined analysis of Arabidopsis, rice, grape, potato and moss CPKs, group IIa showed the highest average $K_a/K_s$ ratios with an average of 0.2479, followed by group IIIa with an average of 0.2307 (Figure 3.9b). However, the difference between these two were not significant based on the SE of the mean. Group I, IIb, IIIb and IV relatively had lower $K_a/K_s$ ratios, with averages of 0.1360, 0.0943, 0.1258 and 0.1106, respectively. Group IIb CPKs in this analysis also had the lowest average $K_a/K_s$ ratio, although its error bars shortly overlap with that of group IV.
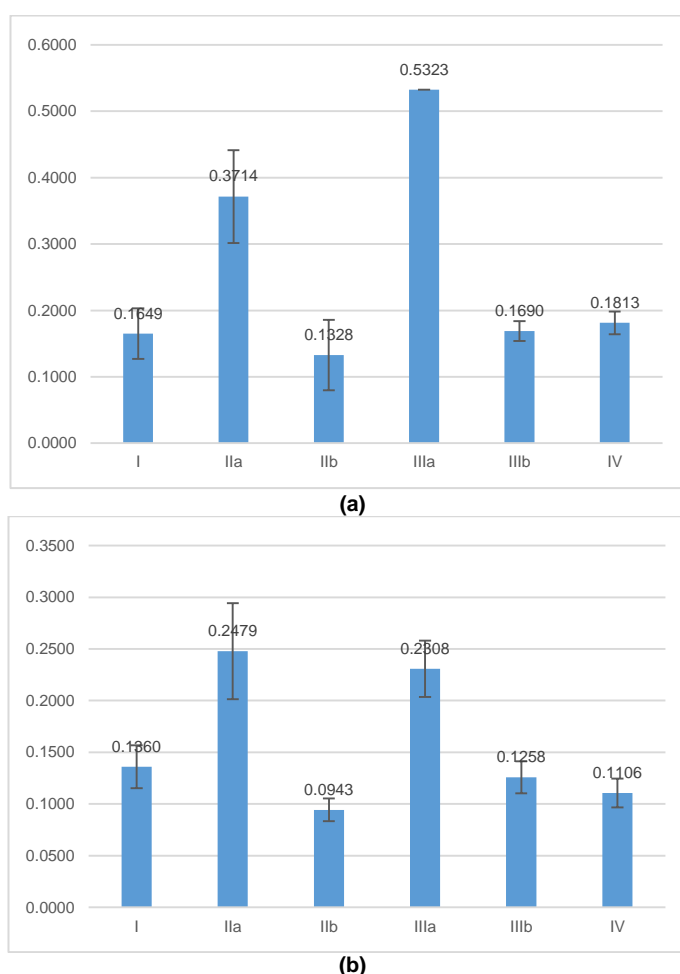


**Figure 3.9 Average $K_a/K_s$ ratios of Group I, IIa, IIb, IIIa, IIIb and IV CPKs.** (a) $K_a/K_s$ analysis of Arabidopsis CPKs. (b) $K_a/K_s$ analysis of Arabidopsis, grape, potato and moss CPKs. Error bars indicate SE of the mean.

**3.4 Discussion**

The 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), this chapter presented how CPK genes have evolved in Viridiplantae. Is the evolution of CPKs in Viridiplantae 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).

### 3.4.1 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 calmodulin (Harmon et al. 2000; Harper et al. 2004; Zhang and Choi 2001) but it is unclear how this ancestral CPK gene diversified into multiple gene family members among plants. Apicomplexan protists generally have only about seven to ten CPK genes, while plants have up to fifty, 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 aa sequence according to Zhang and Choi (2001). Previous authors noted that independent CPK gene expansion events occurred between protists and plants (Billker et al. 2009; Nagamune and Sibley 2006). Apicomplexan CPKs divide into four groups, Api1–4 (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 the phylogenetic analysis in Figure 3.2, section 3.3.2); Api3 and Api4 are more similar to animal CaMK (Nagamune and Sibley 2006). In the phylogenetic analysis shown in Figure 3.2, section 3.3.2, 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, the data presented supports 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 (Figure 3.2, *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 artefact of

94

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 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. The phylogenetic analysis provides an indication that green algae and land plant CPKs had a common ancestral gene, but the diversification of CPKs between these taxa were independent of one another.

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

In contrast with 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 or adaptation of plants into 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 (Figure 3.3). 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; Parfrey et al. 2011; Zimmer et al. 2007) and the first appearance of land plants (estimated between 400 and 700 MYA) (Gensel 2008; Raven and Edwards 2001) 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 characterised by a dramatic increase in spore and vascular plant diversity (a characteristic feature among land plants), as evidenced 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 (Figure 3.2) and molecular clock analysis (Figure 3.3) 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 Figure 3.3). 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 are 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 signalling 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 acclimatisation (Contreras-Porcia et al. 2011), organisation 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 specialisations. One study of moss showed a CPK to be upregulated 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 MAMP signalling and MAPK activation (Boudsocq and Sheen 2013; Boudsocq et al. 2010). 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 type 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.

### 3.4.3 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 fewer than ten genes among extant green algae, to about 10 to 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 were assigned into clusters (i.e. Ia, Ib, Ic.1 and Ic.2 described in Figure 3.2) 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 alkalinisation factors (RALF) gene family (Cao and Shi 2012). The massive expansion of CPK genes among

angiosperms therefore is primarily a passive effect of polyploidisation events that occurred from the last common ancestors of seed plants, angiosperms, monocot and eudicots.

CPK sequences among all the land plants included in the phylogenetic analysis were highly conserved, particularly in the PK, AJ and CAD domains. The average aa 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 included in the analysis were 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 α in plants, which have >50% identity between animals, yeasts and plants (Immanuel et al. 2012). Moreover, the $K_a/K_s$ ratio among Arabidopsis CPKs in these regions demonstrate evolutionary pressure for these sequences to be maintained. Among all the 34 AtCPKs, the $K_a/K_s$ ratios ranged from 0.0455 to 0.6682. With the combined analysis of CPKs in Arabidopsis, rice, grape, potato and moss, the $K_a/K_s$ ratios were still less than one, ranging from 0.0455 to 0.89. This indicates that these CPK genes are under purifying (stabilising) selection. Even though there has been significant expansion of CPK genes, the evolutionary pressure to maintain high sequence conservation in the PK, AJ and CAD regions contributes to various examples of functional similarity, redundancy and overlap among many CPKs. Sequence evolution that brought about distinctive function and localisation in many extant CPKs may have occurred mostly within N-VD and CT.

Several processes drive the preservation of duplicated genes among extant organisms, including functional retention, pseudogenisation (nonfunctionalisation), neofunctionalisation, and subfunctionalisation (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 localisation (Konrad et al. 2011; Ray et al. 2007). 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). Neofunctionalisation refers to the acquisition of novel functions by a duplicated gene, while subfunctionalisation 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 was observed between major CPK evolutionary groups. Some CPKs respond exclusively to certain types of developmental, abiotic, or biotic 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 had 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 along successive ice ages. As the evolutionary groupings can be considered as the outcome of duplications among ancient genes, it can be hypothesised that the ancestral CPK genes of plants had multiple functions, both in the development of reproductive structures and in the maintenance of cellular homeostasis; but due to multiple duplication events, these have subfunctionalised into either developmental or osmotic stress response and neofunctionalised 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 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 abiotic and biotic 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 in flower and pollen. Further analysis of these genes and their orthologues in other species and observation of their effect when ectopically expressed in the organ of their subgroup counterpart may help elucidate the diversification of function among CPKs.

### 3.4.4 What is the most conserved CPK?

As discussed in section 2.4, the identification of the most conserved CPK genes is important because it may lead to the following: the discovery of elements that are important in abiotic and biotic stresses; the identification of motifs important in predicting the tertiary structure and function of other CPKs without known tertiary structures; and the development of molecular approaches to abiotic and biotic stress diagnosis and management in a broad range of plant species. Very few studies have explored the identification of most conserved members of a gene family (Deveaux et al. 2008; Gogarten 1994; Roberti et al. 2006). These studies have used phylogenetic analysis, gene structure analysis, and $K_a/K_s$ ratio calculations in measuring and comparing gene sequence conservation.

In sections 3.3.1 and 3.3.3, the most conserved members of the CPK gene family in green algae and plants were identified using two approaches: phylogenetic analysis and $K_a/K_s$ ratio calculations. Gene structure analysis was performed in section 3.3.1; however, this approach was not used in identifying the most conserved CPKs since there was not enough gene structure data for *P. patens.* The branch lengths in a phylogenetic tree indicate how different or similar a sequence is to its common ancestor; or the amount of evolution estimated to have occurred between them (Baldauf 2003; Bininda-Edmonds 2009). Shorter branches indicate that a sequence has a higher degree of similarity to its common ancestor compared to other sequences; and thus indicate higher degree of sequence conservation. $K_a/K_s$ ratios assumes that if there are more synonymous changes ($K_s$) than nonsynonymous changes ($K_a$) between sequences ($K_a/K_s$ less than 1), there is a selective force that causes the protein to be maintained, or that the protein is under purifying selection. On the other hand, if there are more nonsynonymous ($K_a$) than synonymous ($K_s$) changes between sequences ($K_a/K_s$ greater than 1), there is a selective force that causes the protein to change and evolve, or that the protein is under positive selection. When the amount of synonymous changes equal that of nonsynonymous changes, the protein is in neutral selection (Hasanuzzaman et al. 2013; Heidarvand and Maali Amiri 2010; Walley et al. 2007).

As discussed in section 3.3.1 and 3.3.3, Group II CPKs had the shortest main branch *(e2,* Figure 3.2), and the lowest average branch length (0.547). Within that clade, Group IIb CPKs also included members from all lower plant genomes used in the analysis. As it has the shortest main branch and the shortest average branch length, it can be inferred that CPKs belonging to this

evolutionary group have undergone the least amount of evolution from the common ancestor of CPKs. This was further verified by the $K_a/K_s$ ratio analysis. The $K_a/K_s$ ratios calculated were less than 1 for all of the branches in the generated $K_a/K_s$ tree, which is common among functional proteins as these tend to be highly conserved in evolution (Hasanuzzaman et al. 2013; Walley et al. 2007). In this regard, CPKs are under purifying selection. The amount of evolution in each CPK group was compared using the $K_a/K_s$ ratios in their 'main branch'--- the branches that led to a monophyletic clade that included all members of a CPK group, or a paraphyletic group that includes most of the members of a CPK group. For the two $K_a/K_s$ analyses done (one involving Arabidopsis CPKs only and the other including Arabidopsis, rice, grape, potato and moss CPKs), Group IIb CPKs have demonstrated the lowest $K_a/K_s$ ratios in their main branch compared to the other CPK evolutionary groups. The average $K_a/K_s$ ratio for each evolutionary group was also the lowest in Group IIb CPKs in both the analyses done. However, its difference with groups IIIb and IV did not show significant statistical difference. The average $K_a/K_s$ ratios in this provide weak support to the hypothesis made, but further analysis including more CPK sequences may improve this statistical support. Nevertheless, the data presented in this chapter which includes the phylogenetic analysis and $K_a/K_s$ ratios of the main branches will be enough to suggest that group IIb CPK are the most conserved evolutionary group of CPKs.

### 3.4.5 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 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 orthologues of CPKs between species, and to predict the function of newly identified CPK sequences within a genome.

The phylogenetic analysis and review of expression and functional information of CPKs, presents a detailed view of potential CPK orthologues among agriculturally important plant species (Figure 3.2 and 3.5; Appendix 15-16). Close CPK homologues with highly similar functions are present among the monocot-eudicot clades. The 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 AtCPK 4 and 11 compared with AtCPK12. These three sequences high homology in terms of protein sequence. AtCPKs 4 and 11 have very similar intron/exon as well, but AtCPK 12 gene structure show some differences from these two. In terms of function, AtCPK 4 and 11 show functional similarity, but not AtCPK12. These findings indicate the potential use of gene structure and phylogenetic analysis together to predict functional specificity between paralogues. Highly homologous sequences with similar gene structure will most probably show similarity in function. However, this approach may not be appropriate to find functional orthologues, as the intron/exon patterns may greatly vary between different genomes.

It must be noted that the phylogenetic analysis involved only the conserved regions (PK, AJ and CAD) of CPKs. Since no functional pattern was identified 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 domains (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 orthologues may also reflect the scarcity of functional information about CPKs, particularly in biotic stress responses (Figure 3.5c). There is very little information on the role of CPKs in pathogen defence, 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.

The evolution of CPKs appears to have occurred in parallel with the terrestrial transition of plants. CPK evolution is characterised 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 signalling 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. This suggests that CPK gene explosion among higher plants is largely a result of the polyploidisation events that occurred along plant evolution. CPKs have subfunctionalised and neofunctionalised into different developmental and stress responses. The sequence evolution of the PK, AJ and CAD domains, 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 their function in response to a stimulus or stress response be predicted? 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 orthologues. 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.

# Chapter Four

# What is the role of the most conserved CPKs in plant stress and pathogen responses?

**4.1 Introduction**

CPKs have been reported to function in multiple biological processes in plants, including responses to stress, pathogens, hormones and developmental stimuli. CPKs have been coined as 'hubs' in plant stress signalling and development because they participate in both short-term and long-term responses within one or several different signalling pathways (Schulz et al 2013). In Chapter 3 and as published in Valmonte et al. (2014), it has been presented that CPK functions are redundant and overlapping; that the main evolutionary groups have no unique functions and that there are very few subgroups that show uniformity in function. The majority of the information about the role of CPKs in plant stress and pathogen infection comes from expression data from microarray experiments; although recently there are an increasing number functional experiments that involve *in vivo* and *in vitro* phosphorylation studies, quantitative reverse transcriptase PCR (qRT-PCR), reverse genetics approaches, and promoter function studies.

Stress- and pathogen-responsive genes have been explored as targets for molecular approaches in managing the risks that pathogen and abiotic stress pose to agriculture and the environment. Since CPKs appear to be stress and pathogen responsive, these genes can potentially be utilised as markers for the presence of infection and specific pathogens, as well as the selection of certain cultivars that can be highly susceptible, resistant or tolerant to a plant disease or adverse environmental conditions.

This chapter explores the role of the most conserved CPKs in response to abiotic and biotic stresses for three main reasons. Firstly, this can provide insights with regards the functional diversification of CPKs. Secondly, since these CPKs are highly conserved, their potential use as molecular indicators for plant infections can be applicable to a wide range of plant species. Thirdly,

this may also provide ideas for the use of these genes in selective breeding approaches for crops with higher resistance to environmental stress and pathogen infections.

This chapter aims to answer the question: What is the role of the most conserved CPKs in response to plant stress and pathogen responses? To address this question, this chapter has the following specific objectives: (1) to determine the transcript accumulation of the most conserved CPK in Arabidopsis plants in response to abiotic stresses including drought and high salinity and biotic stresses including specific bacterial, viral and fungal infections; (2) to determine the physical status and stress responsiveness of Arabidopsis plants when the expression of the most conserved CPK gene is ablated (knocked-out) or increased (overexpressed); and (3) to establish whether this function is conserved amongst important crops belonging to different families: rice (*O. sativa*) and kiwifruit (*A. chinensis*).

## 4.2 Materials and Methods

### 4.2.1 *In silico* approach

#### 4.2.1.1 Identifying AtCPK3, 17 and 34 orthologues in rice and kiwifruit

The orthologues of Group IIb AtCPKs (AtCPK3, 17 and 34) in rice and kiwifruit were identified using phylogenetic analyses. The CPK genes that were most closely related to these CPKs in the phylogenetic trees, showing monophyly and dichotomy, were considered as their orthologues. For rice, the orthologous genes were identified directly from the phylogenetic trees reported by Asano et al (2005), Li et al (2008b) and in chapter 3 of this thesis (Figure 2.7). For kiwifruit, the orthologous genes were identified from a phylogenetic tree constructed from predicted CPK coding sequences, since there was no previous report identifying CPKs from its genome. This process was performed in collaboration with a post-doctoral research project at PFR (Arthur et al. 2012).To identify potential CPK sequences in kiwifruit, a BLAST search was carried out using all 34 AtCPK coding sequences as query sequences against three sets of databases available from the Plant and Food Research NZ Genome Server between May 2011 and May 2012: (1) *A. chinensis* EST library; (2) *A. chinensis* CK15_02 Genome Scaffolds; and (3) *A. chinensis* CK51F3_01 Hybrid Gene Models. EST library entries that were highly similar to CPKs (alignment score >=200) were collected and assembled to each of the 34 AtCPKs. The cDNA clones corresponding to the most 5' EST sequences that map to a particular CPK were sent for full DNA sequencing analysis to Macrogen Inc. (South Korea). BLAST hits and full sequences from EST libraries were examined to identify CPKs using the criteria described in section 3.2. Kiwifruit sequences were named AcCPK1 to AcCPK21, reflecting the species name of kiwifruit. Nucleotide and protein sequences of the 21 AcCPKs and 34 AtCPKs were aligned using the ClustalW program (Larkin et al. 2007) in GeneiousPro 5.6. NJ and ML trees were constructed from these alignments using the same software.

The orthologous genes that were identified from the phylogenetic trees are shown in Table 4.1. All these genes were studied in the succeeding *in silico* approaches, while the *in planta* approaches focused only on AtCPK3 and its orthologues in rice and kiwifruit.

**Table 4.1. Group IIB CPKs studied in the subsequent *in silico* and *in planta* approaches**

| Arabidopsis | Rice | Kiwifruit |
|---|---|---|
| AtCPK03 (At4g23650) | OsCPK01(Os01g43410) | AcCPK16 (DB Acc No. 5527801) |
| | OsCPK15 (Os05g50810) | |
| AtCPK17 (At5g12180) | OsCPK02 (Os01g59360) | AcCPK11 (DB Acc No. 5526785) |
| AtCPK34 (At5g19360) | OsCPK14 (Os05g41270) | |
| | OsCPK25 (Os11g04170) | |
| | OsCPK26 (Os12g03970) | |

### *4.2.1.2 Literature and expression database search*

Experimental information from the literature and expression databases was searched for each of the genes listed in Table 4.1. Literature, sequence information and expression data were collected from several biological data resources, namely: National Center for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/), The Arabidopsis Information Resource (TAIR, http://www.arabidopsis.org/), Arabidopsis eFP Browser (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi), Rice Genome Annotation Project (RGAP, http://rice.plantbiology.msu.edu/), The European Bioinformatics Institute (EBI, http://www.ebi.ac.uk/), Plant Expression Database (PLEXdb, http://plexdb.org/), Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg/), and Kiwifruit Genome Information (The New Zealand Institute for Plant and Food Research).

### *4.2.1.3 Expression analysis using Genevestigator*

To collect and analyse gene expression metadata, a publicly available online platform Genevestigator V3 (https://www.genevestigator.com/gv/index.jsp) (Hruz et al. 2008) was utilised. Transcript accumulation of Group IIb CPKs across developmental stages, among tissue types, and upon biotic and abiotic stresses were analysed using the Development Tool, Anatomy Tool and Perturbations Tool, respectively. All analyses were carried out using default parameters, except for setting filters for the Log (2) ratio differences at 1.5 or greater.

### 4.2.2 *In planta* approach

#### 4.2.2.1 Plant material and growth conditions

Wild-type *A. thaliana* ecotype Columbia (Col-0) and *A. chinensis* (Hort 16A) were obtained from existing seed stocks at PFR. Arabidopsis T-DNA insertion knock-out lines of AtCPK3 and 34 and SAIL overexpressor lines of AtCPK3 were obtained as seed stocks from the Nottingham Arabidopsis Stock Centre (NASC, UK). *O. sativa* L. cv Nipponbare seeds were obtained from the National Institute of Agrobiological Sciences (NIAS) in Japan. *A. thaliana* plants that overexpress AtCPK3 from the constitutive 35S CaMV promoter were developed at PFR (described below in section 4.2.2.6). *A. chinensis* plants that are knockouts or overexpressors of AcCPK16 were developed with the assistance of the Breeding and Genomic Team at PFR (described below in section 4.2.2.7).

All plants were grown in a Physical Containment Level 2 (PC2) containment glasshouse or growth cabinet at Plant and Food Research (Mt Albert) at 22-26 °C with a 16hr light and 8 hr dark cycle. Table 4.2 summarises all plant materials used, their source IDs and related controls, approvals or documentation.

**Table 4.2. Plants utilised in this study and their sources**

| Plant Species | Genetic Modification | Source/Construct ID | Approvals and related documents |
|---|---|---|---|
| *A. thaliana* **ecotype Columbia (Col-0)** | wild-type | N/A | EPA Approval GMD101124/ GMD02088 |
| | *atcpk3-1* (T-DNA knockout) | NASC ID N655814 SALK_106720C | |
| | *atcpk3-2* (T-DNA knockout) | NASC ID N522682 SALK_022862 | |
| | *atcpk3* (T-DNA knockout) | NASC ID N595134 SALK_095134 | |
| | *ATCPK3* (T-DNA overexpressor) | NASC N871308 SAIL_120_H09 | |
| | *ATCPK3* (T-DNA overexpressor) | pHEX2AtCPK3full3 (developed at PFR as part of this thesis) | |
| ***O. sativa* L. cv Nipponbare** | None | JP No. 222429 | Import Permit 155.02.05 Phytosanitary Certificate No. 250-93-000696 and No. 250-93-000798 |
| ***A. chinensis* (Hort 16A)** | None | N/A Kind donation from Dr Alison Duffy (PFR) | GMD101131 |
| | *accpk16* (knockout) | Developed by the Breeding and Genomics team (PFR) | GMD101131 |
| | *ACCPK16* (overexpressor) | | |

### 4.2.2.2 Stress and pathogen treatments for analysing expression changes of AtCPK3 and its orthologues

#### 4.2.2.2a A. thaliana

Drought and salt stress treatments were carried out for wild-type Arabidopsis grown in soil (Table 4.3) and in tissue culture tubs (Table 4.4). For the Arabidopsis plants grown in soil, seedlings were grown with normal watering until three weeks. Drought treatment was then performed by eliminating watering for up to 21 days, while salt treatment was carried out by placing the pots in plant growth trays flooded with salt solutions (100 mM or 200 mM NaCl) for up to 14 days. Leaf tissue samples were collected from treated plants and controls at 15 min, 30 min, 1 h, 4 h, 24 h, 48 h, 7 days and 14 days for salt treatment and at 7, 14 and 21 days for drought treatment. For the Arabidopsis plants grown in tissue culture, seedlings were grown on ½ Murashige and Skoog (MS) agar medium (Appendix 22) for two weeks. Tubs were flooded with either 200 mM NaCl or 200 mM mannitol to mimic drought conditions. Leaf and root tissue samples were collected at 15 min, 30 min, 1 h, 4 h, 24 h, and 48 h. There were three biological replicates for each treatment and time point.

**Table 4.3. Experimental design for abiotic stress treatments of Arabidopsis grown in soil**

| TREATMENT TIME | Control (no treatment) | 100mM NaCl | 200mM NaCL | Drought |
|---|---|---|---|---|
| 15 mins. | 🌱🌱🌱 | 🌱🌱🌱 | 🌱🌱🌱 | - |
| 30 mins. | 🌱🌱🌱 | 🌱🌱🌱 | 🌱🌱🌱 | - |
| 1 h | 🌱🌱🌱 | 🌱🌱🌱 | 🌱🌱🌱 | - |
| 4 h | 🌱🌱🌱 | 🌱🌱🌱 | 🌱🌱🌱 | - |
| 24 h | 🌱🌱🌱 | 🌱🌱🌱 | 🌱🌱🌱 | - |
| 48 h | 🌱🌱🌱 | 🌱🌱🌱 | 🌱🌱🌱 | - |
| 7 days | 🌱🌱🌱 | 🌱🌱🌱 | 🌱🌱🌱 | 🌱🌱🌱 |
| 14 days | 🌱🌱🌱 | 🌱🌱🌱 | 🌱🌱🌱 | 🌱🌱🌱 |
| 21 days | 🌱🌱🌱 | - | - | 🌱🌱🌱 |

**Table 4.4. Experimental design for abiotic stress treatments of Arabidopsis grown in MS agar**

| TREATMENT TIME | Control (no treatment) | 200mM NaCL | 200mM Mannitol |
|---|---|---|---|
| 15 mins. | | | |
| 30 mins. | | | |
| 1 h | | | |
| 4 h | | | |
| 24 h | | | |
| 48 h | | | |

For pathogen treatments (biotic stress), two to four-week old wild-type Arabidopsis grown in soil were subjected to a fungal pathogen *Botrytis cinerea*, a bacterial pathogen *Pseudomonas syringae pv. tomato* DC3000 (Pto DC3000 or Pst DC3000) and five viral pathogens, namely *Cauliflower mosaic virus* (CaMV), *Tobacco mosaic virus* (TMV), *Tomato spotted wilt virus* (TSWV), *Turnip mosaic virus* (TuMV), and *Turnip yellow mosaic virus* (TYMV). *B. cinerea* and Pto DC3000 are known pathogens that were used as biotic stress treatments in previously reported Arabidopsis microarray datasets (Winter et al. 2007). The viral pathogens selected for this study were known to readily infect Arabidopsis. There were three biological replicates for each treatment and time point.

For the *B. cinerea* treatment (Table 4.5), isolate REB 702-1 was obtained from Saadiah Arshed, PFR. Two-week old cultures grown on potato dextrose agar (PDA) (Appendix 22) were used to inoculate Arabidopsis leaves. Inoculation was done as described by Govrin and Levine (2002). Plates were flooded with 30 mM $K_2HPO_4$, 0.05% glucose to collect spores and incubated at 20-22 °C for 3 hours. Spore suspensions were diluted to contain 1 x $10^5$ spores/mL. Four-week old Arabidopsis were inoculated by placing a 5 µL drop of the spore suspension onto three rosette leaves. Mock inoculation was also performed with the buffer used for making the spore suspension. Leaf samples were taken from inoculated and mock-inoculated plants at 0, 1, 2, 6 and 10 dpi. Preparation of *B. cinerea* cultures and inoculum were carried out under PC1 conditions while plant inoculation was done under PC2 conditions. Infections were identified by the presence of lesions and sporulation on the leaves from the point of inoculation.

**Table 4.5. Experimental design for *B. cinerea* treatments of Arabidopsis**

| TREATMENT TIME (dpi) | Control (no treatment) | Mock inoculated | *Botrytis cinerea* (REB-702-1) |
|---|---|---|---|
| 1 | | | |
| 2 | | | |
| 6 | | | |
| 10 | | | |

For the Pto DC3000 treatment (Table 4.6), stab culture stocks of *Pseudomonas syringae* van Hall 1902 (ICMP18429, MPI Import Permit No. 2010039160) were obtained from the International Collection of Microorganisms from Plants (ICMP, Landcare Research New Zealand. A liquid culture was grown from this stock using Luria-Bertani (LB) broth (Appendix 22) at 28 °C for 48 hrs. This was adjusted to $1 \times 10^8$ colony-forming units (CFU)/mL and used as inoculum. Four-week old Arabidopsis were inoculated by placing a 5 µL drop of the bacterial suspension to three rosette leaves. Mock inoculation was also performed using LB broth. Leaf samples were taken from inoculated and mock-inoculated plants at 1, 2, 6 and 10 dpi. Handling of Pto DC3000 and inoculation of plants were all carried out under PC2 conditions. Infections were identified by the presence of lesions on the leaves growing from the point of inoculation.

**Table 4.6. Experimental design for *Pto DC3000* treatments of Arabidopsis**

| TREATMENT TIME (dpi) | Control (no treatment) | Mock inoculated | *Pto* DC3000 |
|---|---|---|---|
| 1 | | | |
| 2 | | | |
| 6 | | | |
| 10 | | | |

Infection of Arabidopsis plants with the five viruses CaMV, TMV, TSWV, TuMV and TYMV (Table 4.7) was previously carried out by a collaborator at PFR as part of a PhD project (Lilly et al. 2011). Further information about these viruses is shown in Appendix 23. To describe the process briefly, virus-infected tissue was homogenised in a virus inoculation buffer (Appendix 22). Inoculation on three-week old Arabidopsis seedlings containing six to eight rosette leaves without initial bolts was carried out since infection rate with viruses is potentially higher at this stage of development (Biddington 1986). To inoculate, cotton buds were dipped in the inoculation suspension and carborundum (600 grit, BDH) and then rubbed gently on to three leaves. The inoculated leaves were pierced with a small needle for identification. Mock inoculations were also performed with the inoculation buffer. Leaf samples were taken at 2, 3, 7, 14, 21, 28 and 35 dpi. Agdia Immunostrips® monoclonal antibody strip systems were used to detect TMV and TSWV (Agdia Inc., Illinois USA) following the manufacturer's instructions. All five viruses were also detected by RT-PCR using virus-specific primers, as described by Lilly et al. (2011).

**Table 4.7. Experimental design for virus treatments in Arabidopsis**



### 4.2.2.2b O. sativa

Drought and salt stress treatments were carried out for wild-type *O. sativa* (rice) grown in soil (Table 4.8), with three biological replicates for each treatment and time point. Rice seedlings were grown with normal watering until three weeks. Drought treatment was performed by eliminating watering in the succeeding 14 days, while salt treatment was carried out by placing the pots in plant growth trays flooded with 200 mM NaCl for 14 days. Leaf tissue samples were collected from treated plants and controls at 15 min, 30 min, 1 h, 4 h, 24 h, 48 h, 7 days and 14 days for salt treatment and at 7 and 14 days for drought treatment.

**Table 4.8. Experimental design for abiotic stress treatments in rice**

| TREATMENT TIME | Control (no treatment) | 200mM NaCL | Drought |
|---|---|---|---|
| 15 mins. |  |  | - |
| 30 mins. |  |  | - |
| 1 h |  |  | - |
| 4 h |  |  | - |
| 24 h |  |  | - |
| 48 h |  |  | - |
| 7 days |  |  |  |
| 14 days |  |  |  |

For pathogen treatments (biotic stress), three week old rice seedlings grown in soil were subjected to a fungal pathogen *Magnaporthe grisea*, a bacterial pathogen *Pseudomonas syringae pv. syringae* (Pss) and a viral pathogen *Cymbidium mosaic virus* (CymMV). The pathogens selected for this study as these were known to infect rice and were available for use in research in New Zealand. Further information about these pathogens is shown in Appendix 23. There were three biological replicates for each treatment and time point.

For *M. grisea* and Pss treatments, detached leaf assays (Table 4.9) were performed instead of whole plant infections due to restrictions in the usage of PC2 glasshouses. *M. grisea* (ICMP14481, MPI Import Permit No. 2001012667) and Pss (ICMP4265) stock cultures were obtained from ICMP (Landcare Research, NZ). Handling of these organisms and detached leaf assays were all done under PC2 conditions. *M. grisea* was subcultured on PDA plates, double bagged in ziplock bags and grown at 24-26 °C in the dark for 4 days and then under 12 hrs light/dark cycle for 7 days. Pss was subcultured in Kings medium B agar plates (Appendix 22) and grown at 24-26 °C for 48 hrs.

Spot inoculation in detached rice leaves was performed with modifications from Jia et al (2003). To prepare the inoculum for *M. grisea,* established plates were flooded with 0.25% gelatine, 0.02% Tween 20 solution and were filtered using sterile cheesecloth. The spore suspension was adjusted to 1 x $10^4$ spores/mL. To prepare the inoculum for *Pss,* an overnight liquid culture was prepared from established cultures using LB broth. The bacterial suspension was adjusted to 1 x $10^5$ CFU/mL. The youngest leaves from each rice plant were selected and

cut into 5 cm segments. The detached leaf segments were immediately placed into petri dishes lined with moist filter paper. Each leaf segment was spot-inoculated with seven 5 µL droplets of either control, mock, conidial, or bacterial suspension. The petri dishes were sealed, placed in a ziplock bag and maintained at 21 to 24°C under continuous fluorescent light (10 to 12 µEm$^{-2}$ s$^{-1}$). Sterile deionised water was added every day to the filter paper to maintain moisture levels and avoid desiccation of detached leaves during incubation. Samples were taken at 2, 6 and 10 dpi. Infections were identified by the presence of lesions on the leaves growing from the point of inoculation.

**Table 4.9. Experimental design for *Magnaporthe* and *Pss* detached leaf assays in rice**

| TREATMENT TIME (dpi) | Control (no treatment) | Mock 1 (Magnaporthe) | Magnaporthe grisea | Mock2 (*Pss*) | *Pss* |
|---|---|---|---|---|---|
| 2 | | | | | |
| 6 | | | | | |
| 10 | | | | | |

For the virus infection in rice (Table 4.10), fresh CymMV-infected vanilla leaf tissue was obtained from Prof. Mike Pearson (University of Auckland). This was used as inoculum for an initial batch of rice plants in order to propagate CymMV-infected rice tissue. Virus inoculation was carried out as described in section 4.2.2.2a. Leaf samples were taken at 2, 7, 14, 21, 28 dpi. Agdia Immunostrips® monoclonal antibody strip systems (Agdia Inc., Illinois USA) were used to detect CymMV, following the manufacturer's instructions.

**Table 4.10. Experimental design for virus treatment in rice**

| TREATMENT TIME (dpi) | Control (no treatment) | Mock | CymMV |
|---|---|---|---|
| 2 | | | |
| 7 | | | |
| 14 | | | |
| 21 | | | |
| 28 | | | |

### 4.2.2.2c *A. chinensis*

Drought and salt stress treatments were carried out for wild-type *A. chinensis* Hort16A grown in soil (Table 4.11), with three biological replicates for each treatment and time point. Before planting, kiwifruit seeds were stratified by soaking overnight in 10 ppm giberellic acid (GA3) to increase germination rate. Kiwifruit seedlings were grown with normal watering until four weeks. Drought and salt treatment was carried out as described in section 4.2.2.2b. Leaf tissue samples were collected from treated plants and controls at 30 min, 4 h, 24 h, 48 h, 7 days and 14 days for salt treatment and at 7 and 14 days for drought treatment.

**Table 4.11. Experimental design for abiotic stress treatment in kiwifruit**

| TREATMENT Time | Control (no treatment) | 200mM NaCL | Drought |
|---|---|---|---|
| 30 mins. | 🌱🌱🌱 | 🌱🌱🌱 | - |
| 4 h | 🌱🌱🌱 | 🌱🌱🌱 | - |
| 24 h | 🌱🌱🌱 | 🌱🌱🌱 | - |
| 48 h | 🌱🌱🌱 | 🌱🌱🌱 | - |
| 7 days | 🌱🌱🌱 | 🌱🌱🌱 | 🌱🌱🌱 |
| 14 days | 🌱🌱🌱 | 🌱🌱🌱 | 🌱🌱🌱 |

For the fungal pathogen *B. cinerea* treatment in kiwifruit (Table 4.11), inoculum preparation and inoculation was performed as described in section 4.2.2.2a. Inoculation of three leaves was done for each of the four-week old kiwifruit seedlings. Leaf samples were taken from infected and mock-inoculated plants at 2, 6 and 10 dpi. Infections were identified by the presence of lesions and sporulation on the leaves from the point of inoculation.

**Table 4.12. Experimental design for *B. cinerea* treatment in kiwifruit**

| TREATMENT Time (dpi) | Mock | *B. cinerea* |
|---|---|---|
| 2 | 🌱🌱🌱 | 🌱🌱🌱 |
| 6 | 🌱🌱🌱 | 🌱🌱🌱 |
| 10 | 🌱🌱🌱 | 🌱🌱🌱 |

For the virus infection of kiwifruit (Table 4.12), freeze-dried leaf tissue from a *Cucumber mosaic virus* (CMV)-infected *Delphinium* plant (isolate 03/76) was obtained from Kate Olliver (PFR). Inoculation of two leaves of each of the four-week old kiwifruit seedlings was performed as described in section 4.2.2.2a. Leaf samples were taken from inoculated and mock-inoculated plants at 7, 14, 21, 28 and 35 dpi. Agdia Immunostrips® monoclonal antibody strip systems (Agdia Inc., Illinois USA) were used to detect CMV, following the manufacturer's instructions.

**Table 4.13. Experimental design for virus treatment in kiwifruit**

| TREATMENT | Mock | CMV |
|---|---|---|
| Time (dpi) | | |
| 7 | 🌿🌿🌿 | 🌿🌿🌿 |
| 14 | 🌿🌿🌿 | 🌿🌿🌿 |
| 21 | 🌿🌿🌿 | 🌿🌿🌿 |
| 28 | 🌿🌿🌿 | 🌿🌿🌿 |
| 35 | 🌿🌿🌿 | 🌿🌿🌿 |

Inoculation of kiwifruit with bacteria in the genus *Pseudomonas* could not be performed due to biological safety restrictions (only trained and permanent PFR staff were allowed to perform such experiments) and to experimental limitations in the facility at that time.

### 4.2.2.3 RNA extraction, quality analysis and cDNA synthesis

Total RNA was extracted from the leaf or root samples collected in section 4.2.2.2 using a Spectrum TM Plant Total RNA Kit (Sigma-Aldrich, St. Louis, MO, USA) for Arabidopsis and rice and a modified cetyltrimethylammonium bromide (CTAB) extraction procedure by White et al (2008) for kiwifruit. RNA samples were treated with DNase I (amplification grade; Invitrogen, San Diego, CA, U.S.A.) to remove any potential genomic DNA (gDNA) contamination. RNA concentration and purity was measured using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies Inc., Wilmington, DE, U.S.A.) while RNA integrity was analysed using a Bioanalyzer 2100 RNA Nano LabChip 6000 (Agilent Technologies, Santa Clara, CA, U.S.A.). RNA Integrity (RIN) values were assigned by the Bioanalyzer software algorithm, which determines the quality of a total RNA sample from the 28S:18S ribosomal RNAs ratio and from the entire electrophoretic profile (Schroeder et al. 2006). RIN values range from 1 to 10, with 10 being a fully intact RNA and ≥ 7.0 being acceptable. However, the algorithm used by the software version was designed for mammalian RNA and did not consider chloroplast RNA in plants. RIN values were then adjusted with visual inspection of the peaks for 25S, 23S and 16S RNAs. All RNA samples used for the succeeding experiments were ensured to have an absorbance ratio (A260/280) between 1.8 and 2.2, and adjusted RIN value of 7.0 or greater. RNA samples were reverse transcribed into cDNA using a SuperScript® VILO™ cDNA synthesis kit (Life Technologies-Invitrogen, San Diego, CA, U.S.A.) following the manufacturer's protocol. The total amount of RNA transcribed to cDNA was adjusted to 2 μg in Arabidopsis and 1 μg in rice and kiwifruit, in a total volume of 40 μL.

## 4.2.2.4 Reverse transcriptase-quantitative polymerase chain reaction (RT-qPCR) to measure transcript accumulation of AtCPK3, OsCPK1, OsCPK15 and AcCPK16

### 4.2.2.4a Reference gene selection, primer design and testing

*Selection of reference genes*

Reference genes, or housekeeping genes were used as internal controls to ensure cDNA quality and as standards for quantifying stress and other stimulus-responsive genes. For Arabidopsis, the reference genes that were selected for evaluation in this study were the following: Elongation factor- 1 α (EF-1α, At5g60390), SAND family protein (SAND, At2g28390), Protodermal factor 2 (PDF2, At1g13320) and F-Box family protein (F-BOX, At5g15710). These genes have been listed as some of the stably expressed genes in various Arabidopsis experiments (Czechowski et al. 2005) and have been identified and validated as the most consistent upon virus infection (Lilly et al. 2011). For rice, the reference genes that were selected for evaluation were the following: TBC1 domain family member 22A (OsTBC, LOC_Os09g34040), Tumour protein homologue (OsTPH, LOC_Os11g43900.1), RNA-binding protein (OsRBP, LOC_Os03g46770.1) and Expressed protein 1 (OsEP1, LOC_Os07g02340.1). These were reported as some of the most stable reference genes in development, biotic and abiotic stress in rice (Maksup et al. 2013; Narsai et al. 2010). For kiwifruit, the reference genes selected for evaluation were: Actin mRNA 1 (AdACT1, orthologue of At5g09810), Ubiquitin-conjugating enzyme 9 (UBC9, orthologue of At4g27960) and Protein Phosphatase 2A regulatory unit (PPPRSA, orthologue of At1g13320). These were utilised and verified to be stable in previous RT-qPCR studies in kiwifruit (Bulley et al. 2009; Li et al. 2013). All reference genes were analysed using GeNORM, an algorithm that calculates a gene stability value (M value) based on pairwise comparisons and geometric averaging of transcript abundance (Q values) among reference genes in different biological samples (Vandesompele et al. 2002).

*Primer design and testing*

Gene specific primers near the 3' end were designed for each of the targeted Group IIB CPK genes in Arabidopsis, rice and kiwifruit. Design of forward and reverse primers was performed using the software Primer3Plus (Untergasser et al. 2007) with the following criteria to ensure primer specificity and efficiency: (1) melting temperature (Tm) of 60±3 °C; (2) primer length of 20 to 27 base pairs (bp); (3) GC content of 45-55%, and (4) amplicon size of 130-200 bp. The target regions of the forward and reverse primers spanned an intron (to detect genomic DNA

contamination). Primers for the reference genes were adapted from previous reports mentioned above for Arabidopsis (Lilly et al. 2011), Table 4.14 presents all primers used in this study and relevant information with regards them.

**Table 4.14. Primers used for RT-qPCR analysis of target and reference genes**

| Gene ID / locus | Annotation | Sequence (5' to 3') | | Tm (°C) | % GC | Amplicon size (bp) |
|---|---|---|---|---|---|---|
| At5g15710 | FBOX | F1463 | GGCTGAGAGGTTCGAGTGTT | 59.5 | 55.0 | 140 |
| | | R1602 | GGCTGTTGCATGACTGAAGA | 60.0 | 50.0 | |
| At5g60390 | EF1-α | F922 | CACCACTGGAGGTTTTGAGG | 60.5 | 55.0 | 137 |
| | | R1158 | TGGAGTATTTGGGGGTGGT | 60.0 | 52.6 | |
| At2g28390 | SAND | F1763 | GTTGGGTCACACCAGATTTTG | 60.3 | 47.6 | 127 |
| | | R1896 | GCTCCTTGCAAGAACACTTCA | 60.6 | 47.6 | |
| At1g13320 | PDF2 | F1622 | TCATTCCGATAGTCGACCAAG | 60.1 | 47.6 | 104 |
| | | R1726 | TTGATTTGCGAAATACCGAAC | 60.0 | 38.1 | |
| At4g23650 | AtCPK3 | F1526 | CATTGCTGAAGTAGACACCG | 56.9 | 50.0 | 115 |
| | | R1640 | GATCTCTCACATTCTGCGTC | 55.8 | 50.0 | |
| At5g12180 | AtCPK17 | F1736 | AAGAGAGTTACACACAGGGG | 54.2 | 50.0 | 112 |
| | | R1847 | CCTCCCTTAAAGATCTCCTCC | 58.3 | 52.4 | |
| At5g19360 | AtCPK34 | F1422 | CATGAACGATGGCAGAGAC | 57.7 | 52.6 | 120 |
| | | R1541 | GGATTAGGATCTGGGTTTCC | 57.4 | 50.0 | |
| Os09g34040 | OsTBC1 | TBCF | TGGTCATGTTCCTTCAGCAC | 59.7 | 50.0 | 111 |
| | | TBCR | GACTTGGCGAGCTTTTGAAC | 60.0 | 50.0 | |
| Os11g43900.1 | OsTPH | TPHF | CATTGGTGCCAACCCATC | 60.8 | 55.6 | 113 |
| | | TPHR | AAGGAGGTTGCTCCTGAAGA | 59.0 | 50.0 | |
| Os03g46770.1 | OsRBP | RBPF | ATGTCGAGTACCGCTGCTTC | 60.4 | 55.0 | 120 |
| | | RBPR | TCTCCCTGTCGTTGATGATCT | 59.7 | 47.6 | |
| Os07g02340.1 | OsEP1 | EP1F | AGGAACATGGAGAAGAACAAGG | 59.6 | 45.5 | 112 |
| | | EP1R | CAGAGGTGGTGCAGATGAAA | 59.8 | 50.0 | |
| Os01g43410 | OsCPK1 | F1531 | ATGGGGGACGATAAAACGAT | 60.4 | 45.0 | 123 |
| | | R1653 | GTTGGGAGCAATCTCAGGGT | 61.4 | 55.0 | |
| Os005g50810 | OsCPK15 | F1581 | ACATGGGTGATGAAGCGACA | 59.7 | 55.0 | 139 |
| | | R1720 | AAACATCCGCCGTCGATTTG | 59.6 | 50.0 | |
| EST197478 | AcActin1 | F456 | CCAAGGCCAACAGAGAGAAG | 60.0 | 55.0 | 197 |
| | | R653 | GACGGAGGATAGCATGAGGA | 60.0 | 55.0 | |
| EST6156473 | UBC9 | F | CCATTTCCAAGGTGTTGCTT | 60.0 | 45.0 | 190 |
| | | R | TACTTGTTCCGGTCCGTCTT | 59.6 | 50.0 | |
| EST447684 | PPPRSA | F | GCAGCACATAATTCCACAGG | 59.2 | 50.0 | 110 |
| | | R | TTTCTGAGCCCATAACAGGAG | 59.3 | 47.6 | |
| A9gene044348.1 | AcCPK16 | F1423 | GCCCTCAAGAAGTACAACATGG | 59.3 | 50.0 | 148 |
| | | R1571 | CGGTGTCTGTTCGTGACCAA | 60.5 | 55.0 | |

Primers were initially tested for their ability to amplify product of correct size by employing end-point RT-PCR reaction consisting of 1.0 µL cDNA synthesised from 2µg RNA, 12.5 µL of GoTaq®Green Master Mix (Promega Corp., Madison, WI, U.S.A.), 0.5 µL of each 10 µM forward and reverse primer, and 10.5 µL of UltraPure™ DNase/RNase-free distilled water (LifeTechnologies – Invitrogen San Diego, CA, U.S.A.) to a total volume of 25 µL. The following PCR conditions were employed: initial denaturation at 94°C for 5 min; followed by amplification with 35 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 30 s and extension at 72°C for 30 s; followed by a final extension at 72°C for 5 min. Negative control reactions, omitting cDNA template were also prepared for each set of reaction.

### 4.2.2.4b RT-qPCR experiments and analysis

To quantify the transcript accumulation of the target and reference genes from each sample, qPCR reactions were performed using a LightCycler 480 Real-Time PCR system (Roche Applied Science, Branchburg, NJ, U.S.A.). Reactions were in a 10 µL total volume containing 1 µL of primer pair (2 µM forward and reverse primer), 4 µL of cDNA and 5 µL of LightCycler 480 SYBR Green I Master mix reagent. A Biomek 3000 Robot (Beckman Coulter, Fullerton, CA, U.S.A.) was used to aliquot all reagents, primers, and samples into 384-well plates, with two technical replicates and three biological replicates for each sampling timepoint. The qPCR reaction consisted of pre-incubation at 95°C for 5 mins and amplification with 45 cycles of denaturing at 95°C for 10 s, annealing at 60°C for 10 s and extension at 72°C for 10 s. Fluorescence acquisition was set up at the end of each cycle. The amplification step was followed by a melting curve analysis, with one cycle of 95°C for 5 s, 65°C for 1 min and a ramp to 97°C at a rate of 0.11°C/s. Five fluorescence acquisitions per °C were taken. Samples were cooled at 40° for 10s.

Fluorescence data per cycle were exported from the LightCycler 480 software into a *.csv file using Python 2.6.3 (Python Software Foundation; custom script by Jeremy McRae, PFR). Baseline correction, log transformation and primer PCR efficiency calculation from linear regression were done using the software LinRegPCR 11.1 (Ruijter et al. 2009). Initial expression values or transcript abundance (Q) of each gene for each sample were calculated using the formula shown in Figure 4.1. This calculation employs the comparative ΔCq method and rescaling of the data based on the calculated PCR efficiency and the relative accumulation values for each gene. The M value for each reference gene and Normalisation Factor for each Q value was calculated using GeNorm v3.5 analysis software. Reference genes with an M value less than 1 were considered as acceptable for use in the normalisation of qPCR data. Transcript accumulation of the targeted CPK genes were normalised using two to three reference genes (three for Arabidopsis and kiwifruit; two for rice) with the formulae shown in Figure 1.

<div style="border:1px solid">

**Initial expression values:**
**Q value** = PCR_efficiency $^{(lowest\_Cq\_value - current\_Cq)}$

**Normalisation:**

**For two reference genes:**
**SD factor** = normalisation factor*$((SD\_ref1/(2*mean\_ref1))^2+(SD\_ref2/(2*mean\_ref2))^2)$
**For three reference genes**
**SD factor** = normalisation factor*
$((SD\_ref1/(3*mean\_ref1))^2+(SD\_ref2/(3*mean\_ref2))^2+(SD\_ref3/(3*mean\_ref3))^2)$

**Normalised mean** = mean / normalisation factor
**Normalised SD** = normalised mean*((SD factor/normalisation factor)^2+(SD/mean)^2)^0.5
**Rescaled mean** = current cell / which ever sample mean you want to be = 1
**Rescaled SD** = current cell / which ever sample mean you set to be = 1
**n** = number of replicate wells in the PCR
**rescaled SE** = rescaled SD / (SQRT(n))

</div>

**Figure 4.1. Formulae used to calculate initial expression values and normalised values**

### 4.2.2.5 Design and testing of antibody to detect AtCPK3 protein accumulation

In order to detect and quantify AtCPK3 accumulation from crude leaf samples, an attempt to design and produce antibodies that will specifically detect AtCPK3 protein was done. Since the N-termini of CPKs have the highest sequence variation, this region was selected for searching peptide targets for antibody design. The peptide target and antibody was designed and produced by GenScript Antibody Service (GenScript USA, Inc.) (Table 4.15).

**Table 4.15. Properties of the AtCPK3 antibody produced by GenScript**

| No | Start | Antigenic Determinant | Length | Antigenicity/Surface /Hydrophilicity | Coil | Amphipathic | Synthesis | Rabbit BLAST |
|---|---|---|---|---|---|---|---|---|
| 1 | 30 | CKPAGERRGSSGSGT | 14 | 2.54/1.00/0.68 | Y | Y | N | 42% |

The antibody was tested for its specificity and sensitivity using a western blot analysis. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) was first performed with leaf samples from Arabidopsis wildtype, AtCPK3 knockouts and AtCPK3 overexpressor plants. Denatured protein samples of about 10 to 30 µL were loaded onto a 10% NuPAGE® Bis-Tris SDS polyacrylamide gel (Life Technologies) and run at 130 V for 1 hr 30 min or until the dye front had reached the bottom of the gel. Gels were run in an Xcell *SureLock*® Mini-Cell (Invitrogen™) and electrophoresed in 1x MOPS buffer (Invitrogen™). The Precision Plus Protein™ All Blue Standards (Bio-Rad) was used as size determination standards. Proteins were transferred from the polyacrylamide gel to a 7 x 8.5 cm Immobilon®-P polyvinylidene fluoride (PVDF) membrane (0.2 µm, Millipore). Before the transfer, membranes were immersed in methanol for 10 sec and equilibrated in Towbin Transfer buffer. The protein transfer was set up in an Xcell II™ Blot module (Invitrogen™) at 90 mA for 14 hr using Towbin Transfer buffer. To check for successful transfer,

the membrane was stained with Ponceau S for 5 min and destained in distilled water. Upon complete destaining, the membrane was placed in a sealed plastic bag membrane and was blocked with 5% non-fat milk at room temperature for 1 hr. The solution was replaced with the primary AtCPK3 antibody diluted 1:500x in 5% non-fat milk and was incubated overnight at 4°C. Incubation with horseradish peroxidase-linked secondary antibody (1:5000x dilution) was done at room temperature for 1 hr. The membrane was washed four times for 5 min in 1x Tris-buffered saline-tween (TBS-T) between each step. Antibody-bound proteins were detected using the Western Lightning® ECL Pro (PerkinElmer) chemiluminescent substrate at a 1:1 ratio and incubated at room temperature for 2 min. To visualise the proteins, the membrane was then exposed to Amersham Hyperfilm™ ECL (GE Healthcare) for 1 to 15 min. The image was captured using a CURIX 60 Table-Top Processor (Agfa).

### 4.2.2.6 Development of Arabidopsis plants that overexpress the AtCPK3 gene

#### 4.2.2.6a Development of expression clones by Gateway® cloning

Gateway® cloning technology, which employs bacteriophage lambda site-specific recombination system, was utilised to develop expression clones that constitutively expressed the full AtCPK3 gene. The full AtCPK3 gene was PCR amplified from a healthy Arabidopsis leaf cDNA (obtained as described in section 4.2.2.3) using AtCPK3 gene specific primers that were flanked with Gateway attB sequence: forward primer GatewayAtCPK3F1-5'GGGGACAAGTTTGTACAAAAAAGCAGGCTATGGGCCACAGACACAGCAAGTCCA3' and reverse primer GatewayAtCPK3R1634: 5'GGGGACCACTTTGTACAAGAAAGCTGGGTTCACATTCTGCGTCGGTTTGGCACC3'. The end-point PCR reaction consisted of 1.0 µL healthy Arabidopsis leaf cDNA, 12.5 µL of GoTaq®Green Master Mix (Promega Corp., Madison, WI, U.S.A.), 0.5 µL of each 10 uM GatewayAtCPK3 forward and reverse primer, and 10.5 µL of UltraPure™ DNase/RNase-free distilled water (LifeTechnologies – Invitrogen San Diego, CA, U.S.A.) to a total volume of 25 µL. The following PCR conditions were employed: initial denaturation at 94°C for 5 min; followed by amplification with 35 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 30 s and extension at 72°C for 1 min; followed by a final extension at 72°C for 5 min. Negative control reactions, omitting cDNA template were also prepared for each set of PCR reactions. PCR products were gel purified using QIAquick Gel extraction kit (QIAgen).

The purified PCR product was used to begin the two-step process of Gateway® cloning: a BP Reaction and an LR Reaction (Figure 4.2). The BP Reaction between the AttB-flanked AtCPK3 PCR product combined with an AttP donor vector (pDONR/Zeo, Gateway®) gave rise to an AttL entry clone. The pDONR/Zeo_AtCPK3Full entry clone was then recombined with an AttR destination vector, pHEX2 (obtained from Sakuntala Karunarietnam, PFR) to give rise to an AttB expression clone pHEX2_AtCPK3Full. Maps of the vectors used are shown in Appendix 24.
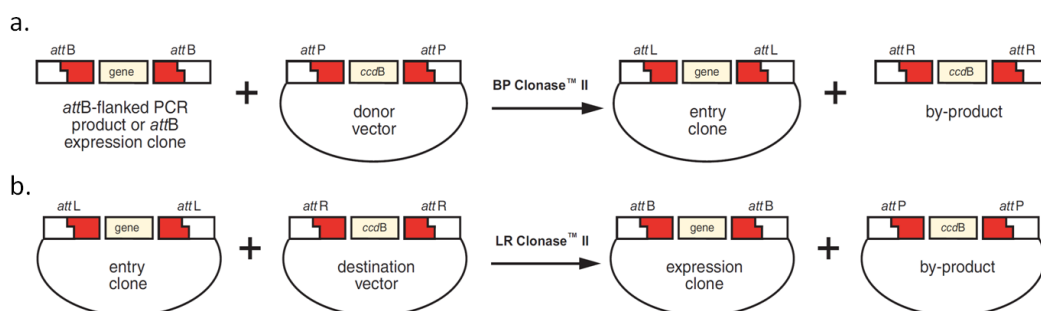


**Figure 4.2. Gateway® recombination.** A. BP Reaction – BP clonase™ II enzyme mix facilitates recombination between an attB substrate (e.g. attB-PCR product) with an attP substrate (e.g. donor vector) to create an attL-entry plasmid. B. LR Reaction – LR clonase™ II enzyme mix facilitates recombination between an attL substrate (e.g. entry-clone) with an attR substrate (e.g. destination vector) to create an attB-expression plasmid.

Bacterial transformations were carried out using OneShot® TOP10 Chemically Competent *E.coli* (Life Technologies) following manufacturer's procedure. Bacteria transformed with the entry clones were grown in media with 100 µg/mL of zeocin (appropriate for pDONR/Zeo) while destination clones were grown in media with 100 µg/mL spectinomycin (appropriate for pHEX2). Resulting clones were screened using colony PCR. Colonies were randomly picked and cells were suspended in 5 µL of distilled water. The end-point PCR reaction consisted of 1.0 µL colony suspension, 12.5 µL of GoTaq®Green Master Mix (Promega Corp., Madison, WI, U.S.A.), 0.5 µL of each 10 uM GatewayAtCPK3 forward and reverse primer, and 10.5 µL of UltraPure™ DNase/RNase-free distilled water (LifeTechnologies – Invitrogen San Diego, CA, U.S.A.) to a total volume of 25 µL. The PCR conditions employed were the same as above.

Plasmids were extracted from positive colonies and sent for sequencing. Plasmid extractions were done using GenElute™ HP Plasmid Miniprep Kit following the manufacturer's instructions. Plasmids were sent for full sequencing to Macrogen Inc (South Korea) using GatewayAtCPK3 forward and reverse primers as well as M13 forward (5'GTAAAACGACGGCCAG3') and M13 reverse (5'CAGGAAACAGCTATGAC3') sequencing primers to check if the sequence and direction of sequence was correct.

### 4.2.2.6b Transformation of expression clones (pHEX2_AtCPK3Full) into Agrobacterium tumefaciens

Electrocompetent *A. tumefaciens* GV3101 was obtained from Tracey Immanuel (PFR). Cells were prepared from a single colony grown in 2 mL of Yeast Extract Peptone (YEP) broth with 25 mg/L rifampicin, and 50 mg/L gentamycin at 28°C overnight with 250 rpm shaking. The culture was grown into large scale by inoculating into 200 mL YEP broth and incubating at 28°C with 250 rpm shaking until the absorbance at 600nm was 0.3 (4-5 hrs). The culture was aliquoted into four 50 mL Falcon tubes and centrifuged at 4,000 x g at 4°C for 10 minutes using a Sorvall® RC-5C Plus centrifuge (DuPont). Cell pellets were collected and resuspended in 20 mL 1 mM sterile filtered HEPES pH7 for each tube. This washing step was repeated twice and the pellets were finally resuspended in 2 mL ice cold 10% (v/v) sterile glycerol. Cells were aliquoted into 40 µL volumes and stored at -80°C.

To transform *A. tumefaciens* GV3101 with pHEX2_AtCPK3Full, electroporation was employed. Electrocompetent cells were thawed on ice and 1 µL of 100 ng/µL plasmid DNA was added to each of two 40 µL cell aliquots. Cells were electroporated in the Eppendorf Eporator at 1440 volts using 1 mm cuvettes and incubated in 500 µL LB broth at 28°C for 3 hr with shaking at 250 rpm. The broth culture was plated LB agar with 100 mg/mL spectinomycin at 28°C for 48 hrs.

### 4.2.2.6c Transformation of Arabidopsis plants using A. tumefaciens

Arabidopsis plants were transformed with the pHEX2_AtCPK3Full constructs in *A. tumefaciens* using the floral dip method (Clough et al., 1998). A single colony of *A. tumefaciens* GV3101 containing pHEX2_AtCPK3Full was inoculated in 2 mL of LB containing 100mg/mL spectinomycin. This was grown at 28°C overnight with 250 rpm shaking and further grown in large scale by adding a 100 µL aliquot into 200 mL LB broth containing 100 mg/mL spectinomycin broth and again incubating overnight with the same conditions. The culture was divided into four 50 mL Falcon tubes and centrifuged at 6,000 xg (Sorvall® RC-5C Plus centrifuge, Du Pont). The culture pellets were pooled and resuspended in 200 mL 5% sucrose solution. Immediately before carrying out the floral dipping, 80 µL of Silwet L-77 (Lehle Seeds) was added to the suspension.

Three pots of Arabidopsis with young inflorescence (approximately four weeks since sowing) were transformed with the cell suspension prepared. The inflorescences were

submerged in the cell suspension for 30-45 sec (first two trials were done at 30 sec and the succeeding trials 45 sec). Soft paper towels were used to remove excess fluid after dipping the plants. Plastic floral sleeves were used to cover each plant and sealed at the top to retain humidity for three days. The seal was removed and plants were allowed to develop seeds for approximately three weeks.

### 4.2.2.6d Seed collection and handling

The floral dipped plants, as well as the knock-out and overexpressor seeds purchased from NASC, were transferred to a seed drying facility and allowed to dry for three to four weeks. The seeds were collected manually using fine sieves to remove dirt and other plant materials. Seeds were collected in 1.5 mL microcentrifuge tubes and stored in a box at room temperature in a PC1 laboratory.

To select for successful transformants or to verify the T-DNA lines, the seeds were surface sterilised and grown in ½ MS agar with 100 mg/mL kanamycin. Seed sterilisation was done by immersing a 100 µL volume of seed in 50% (v/v) bleach, 1% (v/v) Triton X-100 for 20 mins with shaking. The seeds were then rinsed three to five times with sterile distilled water. Sterilised seeds were then grown in media tubs at 20°C with a 16 hr light, 8 hr dark cycle. Non-transformant seeds germinated but became bleached and died from the antibiotic in the medium, whereas transformant seeds and T-DNA lines germinated and developed normally. Upon rooting, the successful transformants were transferred into soil. To retain humidity, plastic floral sleeves were used to cover each plant and sealed at the top for three to five days or until the plants appeared healthy on soil. The seal on top was opened and plants were allowed to develop seeds for approximately three weeks.

For the pHEX2_AtCPK3Full transformant plants, seed collection and plant selection cycles were carried out until the fourth (T4) generation to ensure homozygosity. For the T-DNA knockout and over-expression lines, seeds were collected from three to five plants that grew from the initial selection and kept for use in the subsequent experiments.

### 4.2.2.6e Verification of AtCPK3 overexpression or knock out

End-point RT-PCR was carried out using gene-specific AtCPK3 primers amplifying the full sequence. Using the VILO Superscript cDNA synthesis kit (Life Technologies) cDNA was synthesised from 1µg of RNA in a 20 µL-total volume of cDNA synthesis reaction. The cDNA was

tested in a PCR reaction containing 1 µL cDNA, 12.5 µL of GoTaq®Green Master Mix (Promega Corp., Madison, WI, U.S.A.), 0.5 µL of each 10 µM forward and reverse primer, and 10.5 µL of UltraPure™ DNase/RNase-free distilled water (LifeTechnologies – Invitrogen San Diego, CA, U.S.A.) to a total volume of 25 µL. The following PCR conditions were employed: initial denaturation at 94°C for 5 min; followed by amplification with 35 cycles of denaturation at 94°C for 30 s, annealing at 60°C for 30 s and extension at 72°C for 1 min; followed by a final extension at 72°C for 5 min. Negative control reactions omitting cDNA template were also prepared for each set of PCR reactions.

### 4.2.2.7 Development and verification of kiwifruit plants that either overexpress or are knockouts of the AcCPK16 gene

The transformation procedure to obtain AcCPK16 overexpressors and knockouts was carried out by the Breeding and Genomics team (PFR) based on the previous report by Wang et al. (2007). Transformation was carried out in *in vitro* tissue culture that were previously established (Wang et al. 2006). These tissue cultures were made from winter dormant canes of cultivar Hort16A kiwifruit (*A. chinensis*) that collected from Plant and Food Research orchards at Te Puke, New Zealand. *A. tumefaciens* strain EHA105 (Hood et al. 1993) harbouring binary plasmids pSAK778S_304838 (for overexpression) and pTKO2S_304838 (for knockout) was used in transformation. Overnight culture of bacteria were made in 15 mL MGL liquid medium (Tingay et al. 1997) containing 100 mg/L spectinomycin dihydrochloride at 28°C in an orbital shaker at 250 rpm. Cell pellets were obtained by centrifugation at 5000 xg for 10 min. Cell pellets were re-suspended in 10 mL MS liquid medium (Appendix 22) with vitamins (Duchefa) twice. Final bacterial pellets were re-suspended in 10 mL MS liquid medium supplemented with 100 µM acetosyringone. This was used as the inoculum.

Leaf strips of about 2 to 5 mm in size were excised from young leaves of *in vitro* grown shoots and were inoculated with suspension cultures of *A. tumefaciens* for 30 min. Leaf strips were then blotted dry with sterile filter paper (Whatman, Schleicher and Schuell). Inoculated leaf strips were transferred onto co-cultivation medium M1 (Appendix 22). Leaf strips were incubated at 24°C, with cool white fluorescent light (~120 $\mu mol^{-2}$ m $s^{-1}$) for two days at 16 h photoperiod. The leaf strips were then transferred to regeneration and selection medium M2 (M1 containing 150 mg/L of kanamycin and 300 mg/L of timentin (SimthKline Beecham, Australia, Pty Ltd)). In four to six weeks, further selection and regeneration was performed on the calli formation.

Adventitious buds that formed were transferred to shoot elongation medium M3 (Appendix 22) and elongated shoots (>2 cm) were transferred to rooting medium M4 (Appendix 22). Rooted transgenic plants were transferred to a ½ litre pot with potting mix composed of peat, pumice and vermiculite. Plants were placed in a misting chamber for three weeks and were progressively acclimatised to ambient light and temperature conditions in a containment greenhouse.

End-point RT-PCR was carried out using gene-specific AcCPK16 primers amplifying the full sequence, following the methodology described in section 4.2.2.6e.

### 4.2.2.8 Abiotic and biotic stress treatments of Arabidopsis for comparing phenotypic responses among wild-type, overexpressors and knockouts of AtCPK3

To determine if AtCPK3 performs a vital role in a selection of biotic and abiotic stress responses, knockout and overexpressor lines of AtCPK3 were compared with wild-type Arabidopsis in terms of phenotypic responses (Table 4.16). All treatments were performed on three-week old seedlings, except for virus treatments which were performed on 4-6 leaf stage seedlings. Phenotype measurement parameters are listed in the datasheet shown in Appendix 25. Treatments were performed as described in section 4.2.2.2a. Measurements were taken at day 14 for drought, days 6 and 10 for *B. cinerea*, and at 7, 14, 21 and 28 dpi for TYMV.

**Table 4.16. Experimental design for comparing phenotypic responses in AtCPK3 knockouts, overexpressors and wild-type Arabidopsis.** Number of days specified with the treatments indicate the last day of exposure.

| Lines | CONTROL (No treatment) | Drought (14 d) | *Botrytis cinerea* (7 d) | Mock virus (28 d) | TYMV (28 d) |
|---|---|---|---|---|---|
| WT col-0 | 10 x | 10 x | 10 x | 10 x | 10 x |
| *atcpk3-1* T-DNA KO | 10 x | 10 x | 10 x | 10 x | 10 x |
| *atcpk3-2* T-DNA KO | 10 x | 10 x | 10 x | 10 x | 10 x |
| *atcpk3-3* T-DNA KO | 10 x | 10 x | 10 x | 10 x | 10 x |
| pHEx2AtCPK3Full | 10 x | 10 x | 10 x | 10 x | 10 x |
| ATCPK3-3 OX (SAIL_120_H09) | 10 x | 10 x | 10 x | 10 x | 10 x |

### 4.2.2.9 Abiotic and biotic stress treatments of kiwifruit for comparing phenotypic responses among wild-type, overexpressors and knockouts of AcCPK16

Three overexpressor lines, three knock-out lines and one vector-only line were developed by the Breeding and Genomics team at PFR as described above. Each line had about 10-17 plantlets available for transplanting. Kiwifruit plants from tissue culture were transferred into soil and grown in a PC2 establishment room for four weeks. Plants were acclimatised in the designated glasshouse unit for a week. Some of the plants died during this acclimatisation period.

Due to the small sample size, there were only two to three replicates for each of the four treatments: control, drought, mock, CMV and *B. cinerea* detached leaf assay (Table 4.17). For drought, measurements of plant height and severity scores were taken at days 0, 7 and 14 d, and dry weight at 14 d. For *B. cinerea,* spot inoculation was performed in detached leaves as described in section 4.2.2.2b and measurements of fungal growth or leaf lesion were taken at 2 and 7 dpi, with two leaves for each plant line. For CMV, measurements at 7, 14, 21 and 28 dpi were planned; however, the CMV infection did not appear to be systemic, and had varying symptoms, so measurements were not taken and no phenotype analysis was carried out for the CMV response.

**Table 4.17. Experimental design for comparing phenotypic responses in AcCPK16 knockouts, overexpressors and vector-only kiwifruit.** Number of days specified with the treatment indicated the last day of exposure.

| Lines | No. of plantlets potted from tissue culture | Plants alive at the start of analysis | CONTROL (No treatment) | Drought (14 d) | Mock (virus) (28 d) | CMV (28 d) | *Botrytis cinerea* (detached leaf assay) (7 d) |
|---|---|---|---|---|---|---|---|
| **WT** | 16 | 12 | 🌱🌱 | 🌱🌱 | 🌱🌱 | 🌱🌱🌱 | 🌱🌱🌱 |
| **KO 1** (pTKO2s_304838_E05) | 17 | 13 | 🌱🌱 | 🌱🌱 | 🌱🌱🌱 | 🌱🌱🌱 | 🌱🌱🌱 |
| **KO 2** (pTKO2s_304838_E10) | 16 | 12 | 🌱🌱 | 🌱🌱 | 🌱🌱 | 🌱🌱🌱 | 🌱🌱🌱 |
| **KO 3** (pTKO2s_304838_E11) | 14 | 14 | 🌱🌱 | 🌱🌱🌱 | 🌱🌱🌱 | 🌱🌱🌱 | 🌱🌱🌱 |
| **OX 1** (pSAK778s_304838_E05) | 16 | 12 | 🌱🌱 | 🌱🌱 | 🌱🌱 | 🌱🌱🌱 | 🌱🌱🌱 |
| **OX 2** (pSAK778s_304838_E06) | 13 | 8 | 🌱🌱 | 🌱🌱 | 🌱🌱 | 🌱🌱 | - |
| **OX 3** (pSAK778s_304838_E07) | 14 | 9 | 🌱🌱 | 🌱🌱 | 🌱🌱 | 🌱🌱🌱 | - |

### 4.2.2.10 Statistical analysis

Statistical support to results was determined using Analysis of Variance (ANOVA) and follow-up tests such as Tukey's test and Fisher's LSD. Statistical support was considered as strong ($P \leq 0.01$), good ($0.01 < P \leq 0.05$) or weak ($0.05 < P < \sim 0.10$). Levene's test was initially done before the ANOVA test to determine if the values have equal variance. The statistical software Minitab was used to perform all statistical tests (Minitab 17 Statistical Software 2010).

**4.3 Results**

**4.3.1 What are the orthologues of Arabidopsis Group IIB CPKs in rice and kiwifruit?**

The orthologues of AtCPK3, 17 and 34 were identified in rice and kiwifruit based on sequence homology determined by phylogenetic analysis. The orthologues in rice were identified from a previous report that identified all the CPK gene family members of from the rice genome. For kiwifruit, there was no previous report reporting all CPK gene family members. The draft genome sequence of kiwifruit has been published (Sousa et al. 2013); however, it still required complete assembly and gene annotations. ESTs, genome scaffold data and hybrid genome information available internally from the PFR genome database were screened for the presence of potential CPK sequences. Potential kiwifruit CPKs (AcCPK) were identified based on the criteria given in section 3.2.1. A total of 21 sequences were gathered and considered as potential AcCPKs. This number, however, cannot be considered as the complete set of CPK gene family members in kiwifruit because some CPK subgroups do not have representing AcCPKs from the search performed. Figure 4.3 shows the phylogenetic tree constructed including all Arabidopsis, rice and kiwifruit CPKs identified. Group IIB CPKs form one clade at the bottom of the tree (highlighted in green). Table 4.18 shows the percent aa identities of Group IIb CPKs from these three species.

In rice, there were two genes most closely related to AtCPK3 (OsCPK1 and OsCPK15) while there were four genes most closely related to AtCPK17 and 34 (OsCPK2, 14, OsCPK25 and OsCPK26). AtCPK3 is a singleton; however, it is orthologous to a closely related pair in rice, OsCPK1 and 15, which have 83.79% aa identity across the whole length of the protein. These genes have 72.59 % and 70.30% aa identity with AtCPK3, respectively. On the other hand, AtCPK17 and 34 are a closely related pair having 92.99% aa identity to each other. Their orthologues in rice are two gene pairs: OsCPK2 and 14 with 85.42% aa identity to each other and OsCPK25 and 26 with 99.82% aa identity to each other. These genes have 76.18%, 74.72%, 70.24% and 70.24% aa identity with AtCPK17, respectively and 77.04%, 75.61%, 70.24% and 70.24% aa identity with AtCPK34, respectively. These rice CPK genes were found to be present in the duplicated regions of chromosomes 1, 5, 11 and 12, respectively, and therefore are considered to have arisen via genome segmental duplication events (Asano et al. 2005). OsCPK25 and 26 were considered as products of a recent duplication event as evident from their high aa and nucleotide sequence (99.3%) identities.

**Figure 4.3. Phylogenetic analysis of Arabidopsis, rice, and kiwifruit CPKs.** Group IIB CPKs (at the bottom of the tree) are highlighted in green.

**Table 4.18. Percent aa identity of Group IIb CPKs in Arabidopsis, rice and kiwifruit.** Matrix generated using Geneious 8.0

| | AtCPK 03 | OsCPK 01 | OsCPK 15 | AcCPK 16 | AcCPK 03 | AtCPK 17 | AtCPK 34 | OsCPK 02 | OsCPK 14 | OsCPK 25 | OsCPK 26 | AcCPK 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **AtCPK03** | | 72.59 | 70.30 | 69.89 | 45.39 | 64.53 | 65.09 | 62.34 | 61.01 | 58.86 | 58.86 | 62.85 |
| **OsCPK01** | | | 83.79 | 71.40 | 45.92 | 67.49 | 66.35 | 63.19 | 62.41 | 61.04 | 61.04 | 63.48 |
| **OsCPK15** | | | | 71.40 | 46.63 | 63.54 | 62.43 | 59.74 | 60.11 | 59.78 | 59.78 | 63.30 |
| **AcCPK16** | | | | | 58.46 | 64.96 | 65.34 | 63.45 | 61.80 | 61.74 | 61.74 | 67.36 |
| **AcCPK03** | | | | | | 43.09 | 43.09 | 43.01 | 42.53 | 42.91 | 42.91 | 44.77 |
| **AtCPK17** | | | | | | | 92.99 | 76.18 | 74.72 | 70.24 | 70.24 | 78.59 |
| **AtCPK34** | | | | | | | | 77.04 | 75.61 | 70.24 | 70.24 | 78.77 |
| **OsCPK02** | | | | | | | | | 85.42 | 71.59 | 71.59 | 74.86 |
| **OsCPK14** | | | | | | | | | | 70.75 | 70.75 | 72.69 |
| **OsCPK25** | | | | | | | | | | | 99.82 | 69.78 |
| **OsCPK26** | | | | | | | | | | | | 69.78 |
| **AcCPK11** | | | | | | | | | | | | |

The genes that appeared orthologous to AtCPK3 in kiwifruit were AcCPK16 and AcCPK3. AcCPK16 has 69.89% aa identity with AtCPK3, while AcCPK3 only has 45.39% amino identity. AcCPK3 appears to be an anomalous sequence because its 5' half did not align well with either AcCPK16 or AtCPK3 (Appendix 26). The 3' half aligned well with AcCPK16, with 94.6% identity in that region (aa position 331 to 626). It is possible that this sequence is a gene pair of AcCPK16, or a variant of AcCPK16. The great dissimilarity in the 5' half could be due to issues with the splicing algorithms utilised in gene prediction software. For the purpose of this research, the functional and bioinformatic analysis focused on AcCPK16 and did not include AcCPK3.

There appeared to be a single gene (AcCPK11) orthologous to AtCPK17 and 34 in kiwifruit. It is interesting to note that while there were two orthologous gene pairs in rice, there was only one identified in kiwifruit. However, there may be other kiwifruit CPK genes unidentified because of the limitations on the availability of a fully assembled genome.

**4.3.2 How is the expression of Group IIB CPKs in Arabidopsis affected by biotic and abiotic stresses? Is this similar in other monocot and dicot plants?**

*4.3.2.1 In silico approach.*

Divergence into two main groups of function was observed among Group IIB CPKs. This is based on previous literature and gene expression databases mentioned in section 3.3.2 (Figure 3.6 and Appendices 14 to 16). Group IIb.1 CPKs appeared to be responsive to stress, pathogens and environmental stimuli whereas Group IIb.2 CPKs appeared to be exclusively important in floral development. This is also supported by microarray information from TAIR showing their transcript accumulation across the plant's anatomy, developmental stages and different physiological conditions (Figures 4.4 to 4.8).

### *4.3.2.1a Group IIb CPKs in development and plant anatomy*

Transcripts of Group IIb.1 CPKs (AtCPK3, OsCPK1 and 15) were present in all developmental stages (Figure 4.4) and throughout the plants' anatomy (Figure 4.5). On the other hand, Group IIb.2 CPKs (AtCPK17 and 34 and OsCPK 2, 14, 25 and 26) were only present during the stages of floral development. OsCPK14 appeared to have two splice variants, which have similar transcript profiles. Group IIb.2 CPKs were concentrated in the stamen while Group IIb.1 CPKs have lower expression levels in this organ. Moreover, Group IIb.1 CPKs were reported to change in transcript accumulation in response to various stresses, while there was no stress response among Group IIb.2 CPKs. However, no transcript information, was available for AcCPK11 and AcCPK16.

The accumulation of Group IIb.2 CPK transcripts in the whole plant microarray was poorly detectable in Arabidopsis throughout development, while they are highly detectable in rice during early floral development stages alone (Figure 4.4). Based on this figure, the amount of AtCPK17 and 34 detected in Arabidopsis only increased slightly as flowers started to develop. Although these mRNAs are highly accumulated in the stamen, the low detection may be due to the small size and ratio of floral tissue compared to the total anatomy of Arabidopsis plants. In contrast, rice floral tissues are more abundant in relation to the plant's total anatomy. Group IIb.2 CPKs were present in the stamen and flower in very high amounts, but Group IIb.1 CPKs were present in this organ in low to medium amounts only.
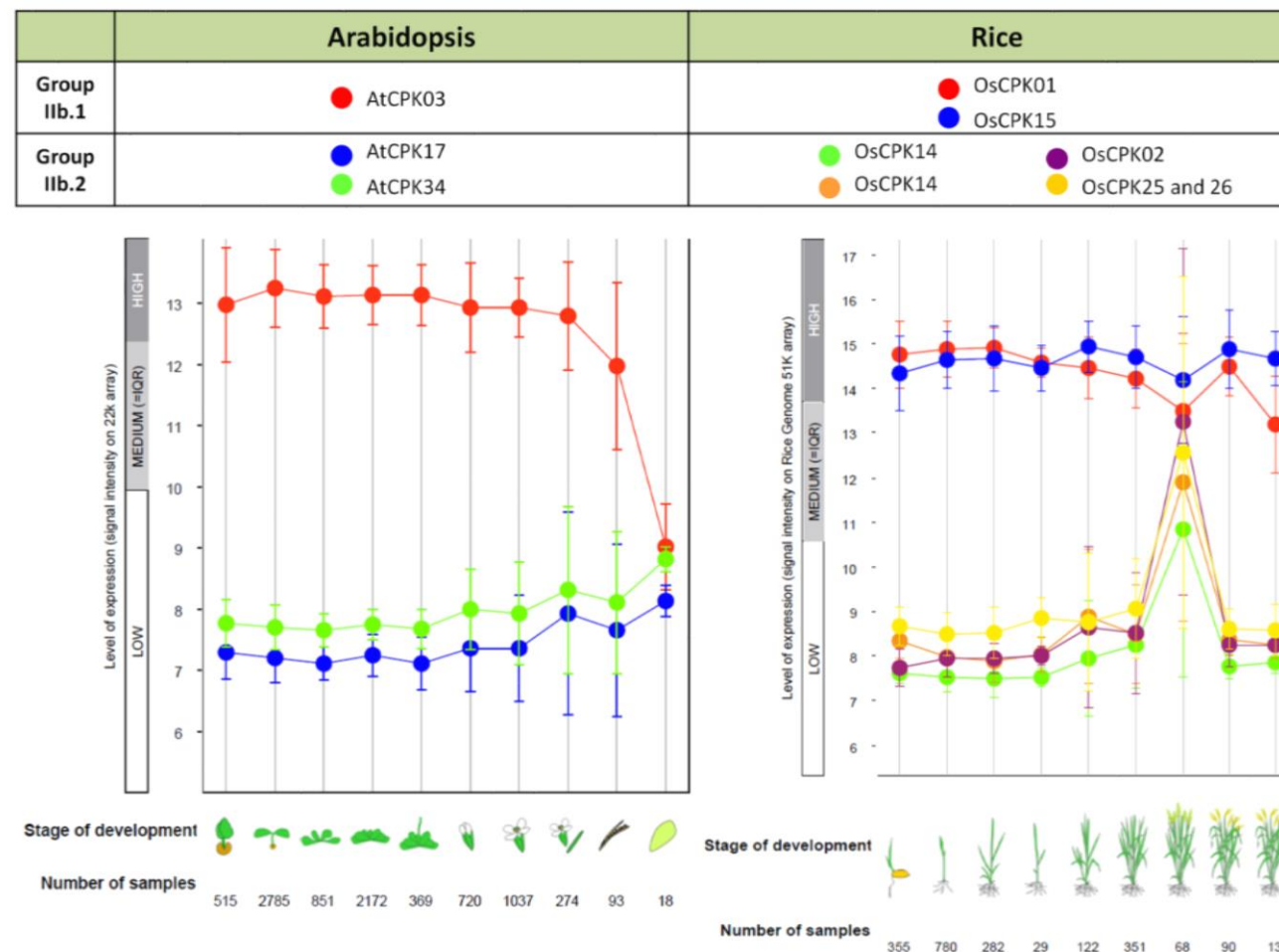
**Figure 4.4. Summary of Group IIB CPK transcript accumulation across developmental stages.** Figure generated using Genevestigator V3 (Hruz et al. 2008).
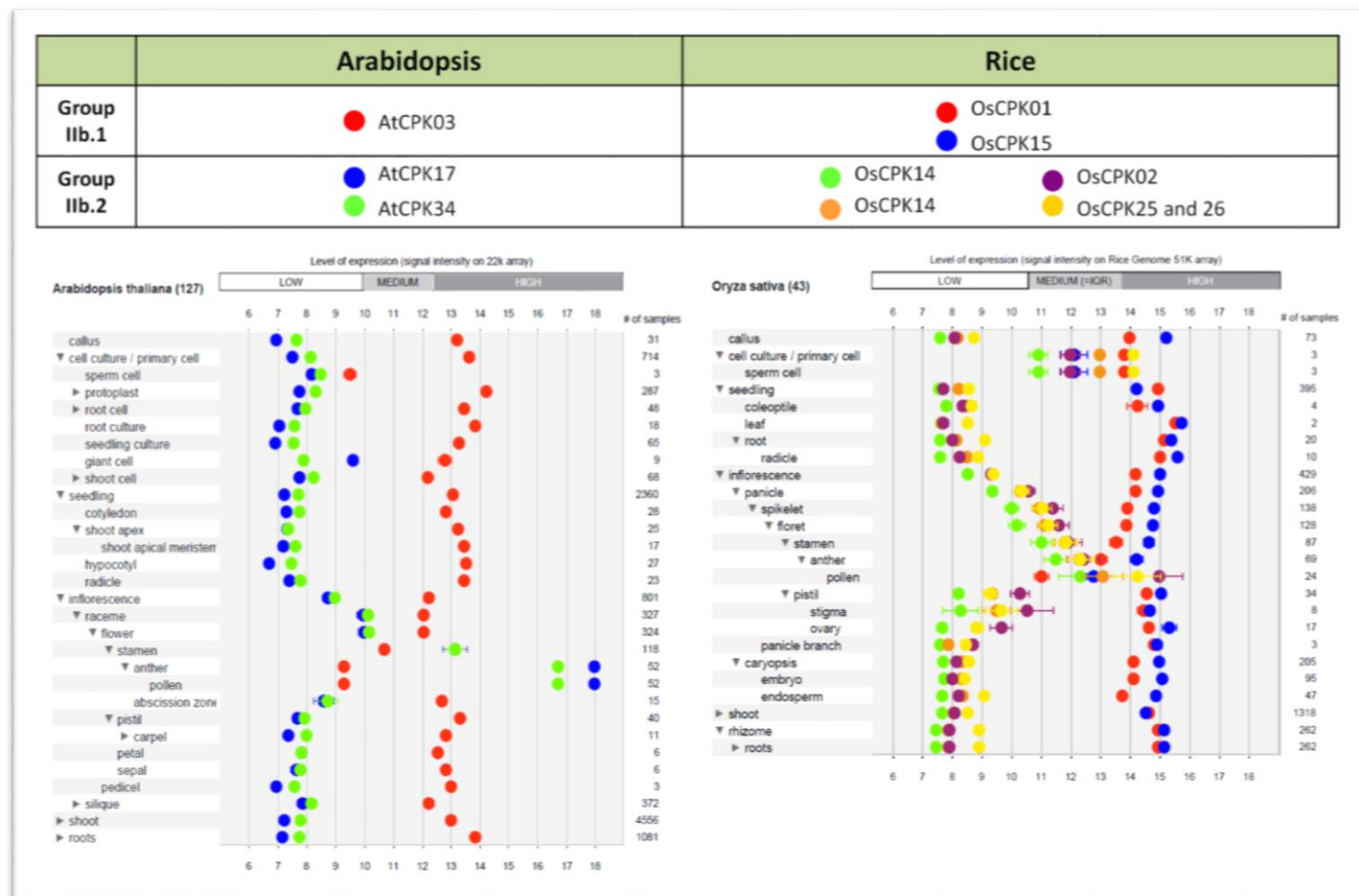
**Figure 4.5. Summary of Group IIB CPK transcript accumulation among tissue types.** Figure generated using Genevestigator V3 (Hruz et al. 2008).

### 4.3.2.1b Abiotic stress responses of Group IIb CPKs

The transcript accumulation of AtCPK3 changes in response to various abiotic stresses. Based on the collected expression information from Genevestigator V3 (Figure 4.6), AtCPK3 transcript increased by about two-fold (one unit in Log2 ratio) in response to the following: cold treatment in rosettes and drought treatment in roots among Col-0 ecotype Arabidopsis plants; drought treatment in whole plants with srk2dei mutation; and hypoxia treatment among Col-0 ecotypes and anac102 mutants. It increased by two to four-fold in response to cold among plants that overexpress RPS4, including those that also have rrs1 and eds1 mutation. On the other hand, AtCPK3 transcript decreased by two-fold in response to heat among hsf1 mutants and ws ecotypes, in response to osmotic stress (ecotype not mentioned), and upon labelling with a photoactivatable ribonucleoside analogue 4-thiouridine (4SU) at 17 °C and 27 °C.

Based on a 24-hr time series experiment data available from TAIR and Arabidopsis eFP browser, AtCPK3 levels fluctuate in response to drought, salt and mannitol, both in root and shoot tissue samples (Figure 4.7). In response to drought treatment (air stream for 15 minutes with loss of approximately 10% weight), there was no significant difference in AtCPK3 expression between control and treated shoot samples in the first hour, but a decrease of about 20% was observed at 3 hrs. The response to drought treatment was similar in root samples; however, there was a small increase at 1 hr, decrease between 3 to12 hrs and no difference from control by 24 hrs. In response to salt treatment (150 mM NaCl), AtCPK3 increased slightly between 3 to 24 hr in shoot samples but decreased continuously between 30 minutes and 24 hrs in root samples. In response to mannitol treatment (300mM mannitol) which induces osmotic stress, AtCPK3 decreased only at 24 hr in shoots but continuously decreased between 1 and 24 hr in roots. AtCPK17 and AtCPK34 did not show any significant change in any of the treatments.
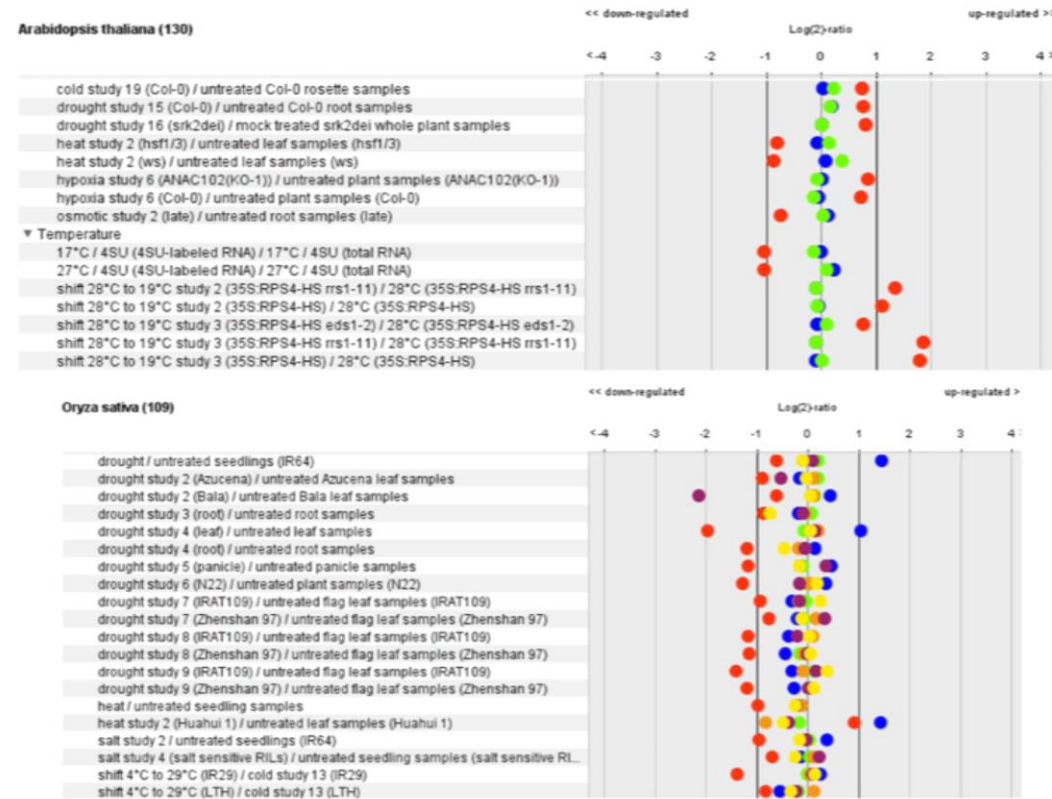
**Figure 4.6. Summary of Group IIB CPK transcript accumulation in response to abiotic stress.** Figure generated using Genevestigator V3 (Hruz et al. 2008).
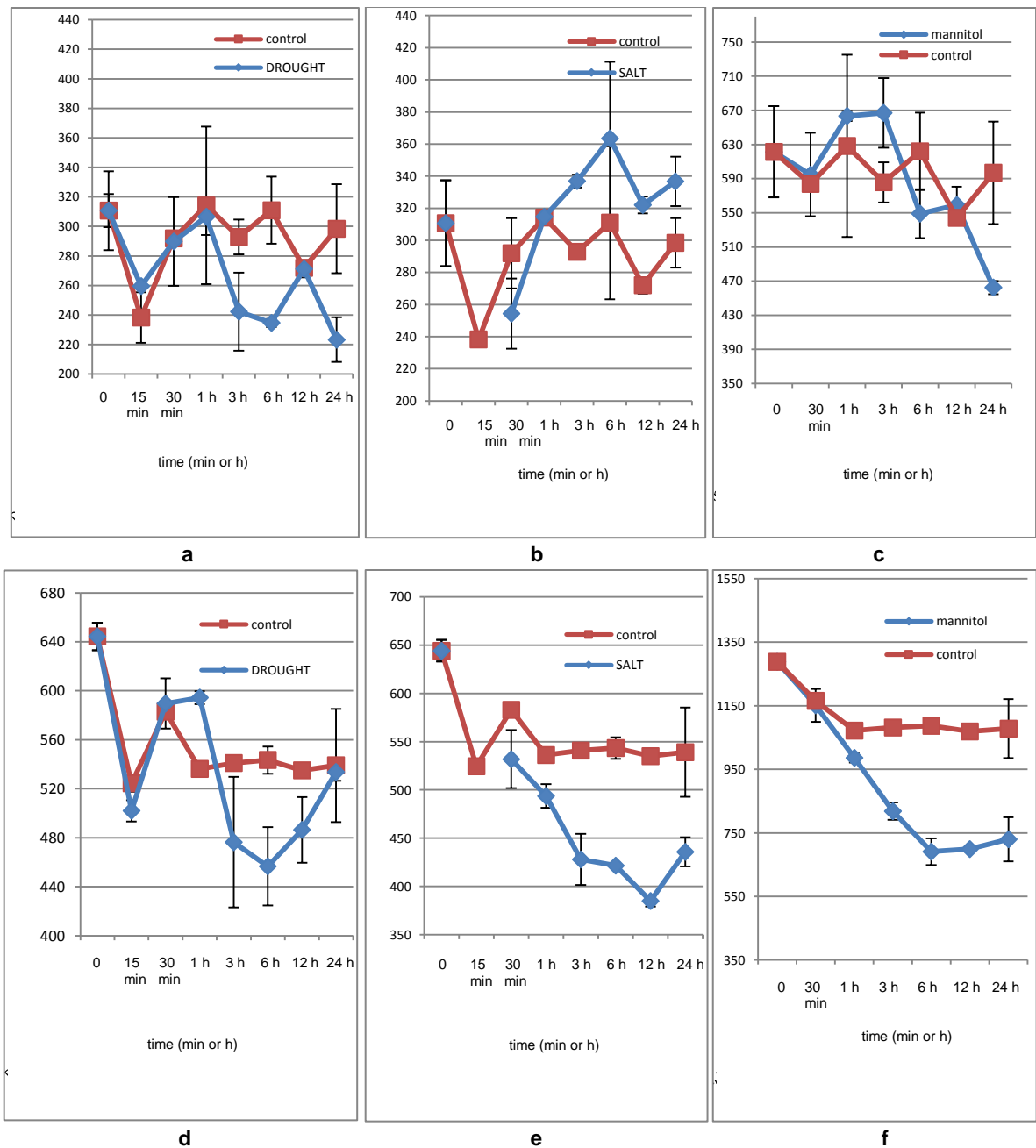
**Figure 4.7. AtCPK3 transcript accumulation in response to drought, salt and osmotic stress; time series experiment.** (a) Shoot, drought treatment. (b) Shoot, salt treatment. (c) Shoot, mannitol treatment. (d) Root, drought treatment. (e) Root, salt treatment. (f) Root, mannitol treatment. Publicly available data of expression values were taken from the Arabidopsis eFP Browser (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi). Values shown are normalised signal intensity readings measured from samples taken at 15 mins, 30 mins, 1 h, 3 h, 6 h, 12 h and 24 h.

In rice, abiotic stress treatments were reported in different cultivars such as IR64, Azucena, Bala, N22, IRAT109, Zhenshan97, and other varieties (Figure 4.6). In several cultivars, it was shown that in response to drought OsCPK1 decreased by approximately two-fold in leaf, panicle and root samples, while in one study OsCPK1 decreased by approximately four-fold. On the other hand, OsCPK15 increased by two-fold in only two drought studies (IR64 and unknown variety) and did not change significantly in the other cultivars. In response to heat, OsCPK1 decreased by two-fold in one study while OsCPK15 did not show any change. In another heat study with the Huahui1 cultivar, both OsCPK1 and OsCPK15 increased by two-fold. In response to salt, OsCPK1 decreased by about two-fold while OsCPK15 showed a very slight increase. OsCPK 14, 2, 25 and 26 did not show significant changes in transcript accumulation in response to any of the stresses.

### 4.3.2.1c Biotic stress responses of Group IIb CPKs

In response to different biotic stresses, AtCPK3, OsCPK1 and 15 showed changes in transcript accumulation in a very few treatments (Figure 4.8). Approximately two-fold increase was shown for AtCPK3 in response to the plant viruses *Cabbage leaf curl virus* (CaLCuV) and *Turnip mosaic virus* (TuMV). A 1.5-fold increase in AtCPK3 was observed in response to five pathovars of the bacterial pathogen *Pseudomonas syringae* and to *Xanthomonas campestris,* while a two-fold increase was observed in response to the fungus *Alternaria brassicicola* at 6 hours post inoculation (hpi)*.* On the other hand, approximately 1.8-fold decrease in AtCPK3 was observed in response to a mutant *P. syringae* pathovar (pv. Tomato DC3118 with Cor-hrpS mutation) and in the oomycete *Phytophthora parasitica* (at 6 and 10 hpi). A two-fold decrease was also observed in response to the fungus *Alternaria brassicicola* (at 6 hpi)*.* OsCPK1 and 15 both increased by approximately two-fold in response to *Agrobacterium tumefaciens* among calli samples of Nipponbare and Zhenshan rice cultivars in four studies, while only OsCPK1 increased in two other studies of the Nipponbare cultivar. In response to another bacterial pathogen, *Xanthomonas oryzae,* OsCPK1 increased between two- and four-fold in all reported studies except for one that infected IR24 cultivars with *X. oryzae pv. oryzae* PXO99A, where it decreased by two-fold. In the same study OsCPK15 did not show significant changes in response to infections among IR24, IRBB5 and IRBB7 cultivars with approximately a 1.5-fold increase. In response to the fungal pathogen *Magnaporthe grisea,* infections on the Taipei 309 cultivar showed a two-fold decrease in both OsCPK1 and OsCPK15 while infections on TP-Pi54-15

showed a two-fold decrease in OsCPK1 only. Infection with a closely related fungus *M. oryzae* showed a two-fold increase (at 2 dpi). Treatment with the brown planthopper *Nilaparvata lugens* showed about a two-fold decrease in OsCPK1 in the stems of two rice varieties. OsCPk2, 14, 25 and 26 did not show significant changes.

Reverse mutation and gene over-expression studies in Arabidopsis (Table 4.19) support the role of AtCPK17 and 34 in pollen development (Myers et al. 2009; Zhou et al. 2009) and the importance of AtCPK3 in osmotic and salt stress (Mehlmer et al. 2010; Mori et al. 2006), herbivore attack (Kanchiswamy et al. 2010), hormone signalling (Munemasa et al. 2011), and flg22 (bacterial) responses (Boudsocq et al. 2010). These functions correlate with the tissue localisation of these genes and changes in transcript accumulation as detected in microarray studies. However, no studies have yet focused on the expression levels of Group IIb.1 CPKs in response to different stresses and the direct effect of the genes' absence or overexpression in the phenotype of the whole plant. In the succeeding sections of this chapter, this study focuses on members of Group IIb.1 CPKs in the model dicot plant, Arabidopsis, a model monocot plant, rice, and another dicot plant in the asterid family, kiwifruit.

**Figure 4.6. Summary of Group IIB CPK transcript accumulation in response to biotic stress.** Figure generated using Genevestigator V3 (Hruz et al. 2008).

**Table 4.19. Previously reported functions of AtCPK3, 17 and 34**

| Questions | | AtCPK3 | AtCPK17 | AtCPK34 |
|---|---|---|---|---|
| **Mutation experiments** | **1.What happens to the plants phenotype when this gene is disrupted (LOSS-OF-FUNCTION mutants)?** | - Does not impair MeJA-induced stomatal closure (Munemasa et al. 2011)- MeJA independent<br>- Lower transcript levels of PDF1.2 compared to WT plants during herbivore attack (Kanchiswamy et al. 2010)<br>- Stomatal closure, ABA and $Ca^{2+}$ activation of slow-type anion channels, and ABA activation of plasma membrane $Ca^{2+}$-permeable channels are impaired (Mori et al. 2006)<br>-plants are salt-sensitive; germination rate under salt stress is decreased (Mehlmer et al. 2010) | cpk17 and 34 double mutants result in 350-fold reduction in pollen transmission efficiency and three-fold reduction in tube growth rate (Myers et al. 2009) | |
| | **2.What happens to the plants phenotype when this gene is overexpressed (OVER-EXPRESSION mutants)?** | - In overexpression mutant protoplasts, CPK3 kinase activity was induced by salt and other stresses (Mehlmer et al. 2010)<br>- germination rate under salt stress is increased (Mehlmer et al. 2010) | Transient overexpression did not affect pollen tube tip growth (Zhou et al. 2009) | Transient overexpression induce depolarisation of pollen tube growth (reduced elongation, increase width) (Zhou et al. 2009) |
| **Physiology/ Interaction with signals** | **3.What signals (hormone, light, and other genes) ARE AFFECTED BY THE EXPRESSION/ACTIVITY of this gene?** | - When constitutively active, it could induce/activate more than five-fold a flg22 reporter NHL10-LUC in mesophyl protoplasts (Boudsocq et al. 2010) phosphorylates ERF1, HsfB2a, and CZF1/ZFAR1 in the presence of Ca2+(Kanchiswamy et al. 2010) | Due to very low transcript levels, significant changes are undetectable | |
| | **4.What signals (hormone, light, and other genes) ARE NOT AFFECTED BY/DOES NOT AFFECT THE EXPRESSION/ ACTIVITY of this gene?** | - Transcriptional induction of salt stress and MAPK dependent marker genes are not affected by CPK3 (Mehlmer et al. 2010)<br>- MeJA (Munemasa et al. 2011)<br>- ABA and IAA* | | |

### 4.3.2.2 In planta approach

#### 4.3.2.2a Testing of primers designed or selected for qPCR

Primers that were designed or selected for use in the RT-qPCR experiments were initially tested by end-point PCR to check for primer specificity and sensitivity. These primers were listed earlier in Table 4.14, section 4.2.2.4a. The primer sets were also designed against an intron-spanning target site to readily determine if genomic DNA is present in the cDNA samples without needing a minus RT control. If genomic DNA was present in a sample, a product of higher molecular weight was expected to be seen in the samples along with the expected PCR product.

*AtCPK3, 17 and 34 qPCR primers*

The AtCPK3, 17 and 34 RT-PCR products showed bands of the expected size: 115 bp for AtCPK3 (Figure 4.9a), 112 bp for AtCPK17 and 120 bp for AtCPK34 (Figure 4.9b). No genomic DNA was present in the samples as indicated by the presence of only one band of the expected size. As shown in Figure 4.9 no extra bands were seen in the leaf, flower or floral samples for each primer set (lanes 2 to 5 in Figure 4.9a and lanes 3, 4, 6 & 8 in Figure 4.9b). For AtCPK3, a minus RT set up was initially setup to confirm the absence of genomic DNA in the RNA samples used. These RNA samples were used in the subsequent RT-PCR experiments to test the other primers.
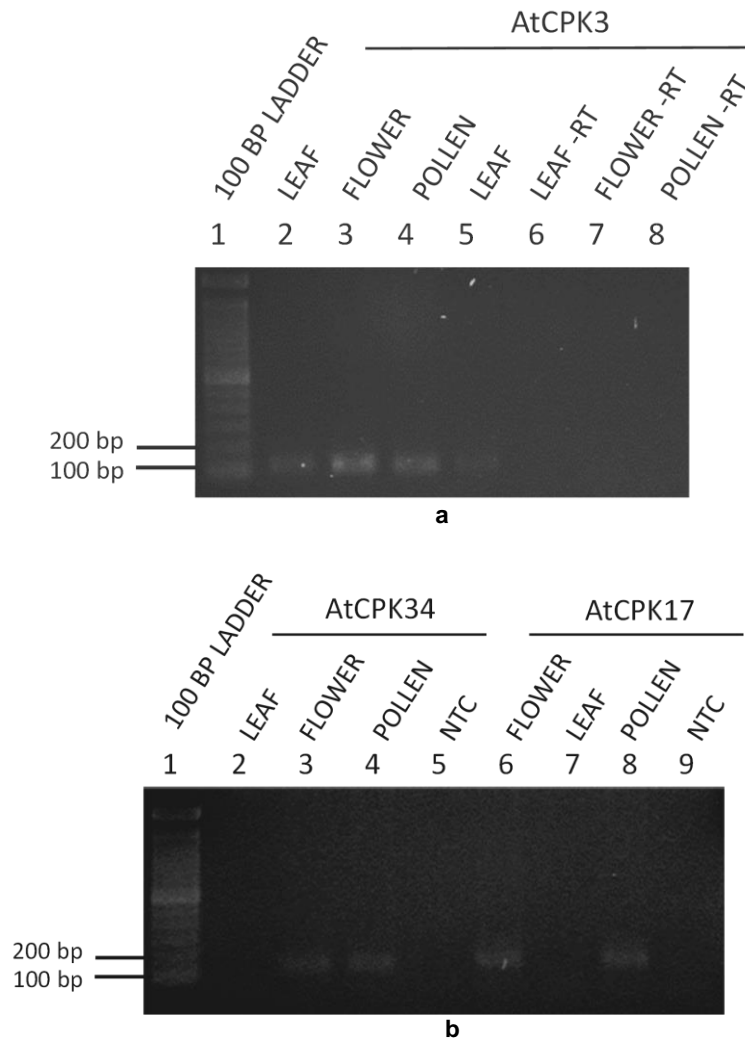
**Figure 4.9. PCR products from Arabidopsis leaf, flower and pollen cDNA using AtCPK3, 17 and 34 qPCR primers. (a) AtCPK3 qPCR primer.** Lane 1: 100 bp ladder (Invitrogen); Lanes 2 to 5: AtCPK3 RT-PCR products from Arabidopsis leaf, flower and pollen RNA, with RT; Lanes 6 to 8: AtCPK3 RT-PCR products from Arabidopsis leaf, flower and pollen RNA, minus RT. **(b) AtCPK17 and 34 qPCR primers.** Lane 1: 100 bp ladder (Invitrogen); Lanes 2 to 4: AtCPK34 PCR products from Arabidopsis leaf, flower and pollen cDNA, with RT; Lane 5: AtCPK34 RT-PCR product, NTC; Lanes 6 to 8: AtCPK17 RT-PCR products from Arabidopsis leaf, flower and pollen cDNA, with RT; Lane 9: AtCPK17 RT-PCR product, NTC.

*Arabidopsis reference genes qPCR primers*

In Arabidopsis, four reference gene primer sets were available from a collaborator who carried out a previous study on identifying and validating reference genes for normalisation of transcripts in virus-infected Arabidopsis (Lilly et al. 2011). All the reference genes showed RT-PCR products with bright bands of the expected sizes: 137 bp for EF1-α, 140 bp for FBOX, 127 bp for SAND and 104 bp for PDF2 (Figure 4.10). To optimise the use of qPCR materials having four genes per plate (including target gene) only three out of the four genes were selected for use in the subsequent RT-qPCR experiments. These were EF1-α, FBOX and SAND, which were the top three most stably expressed in virus infection (Lilly et al. 2011).
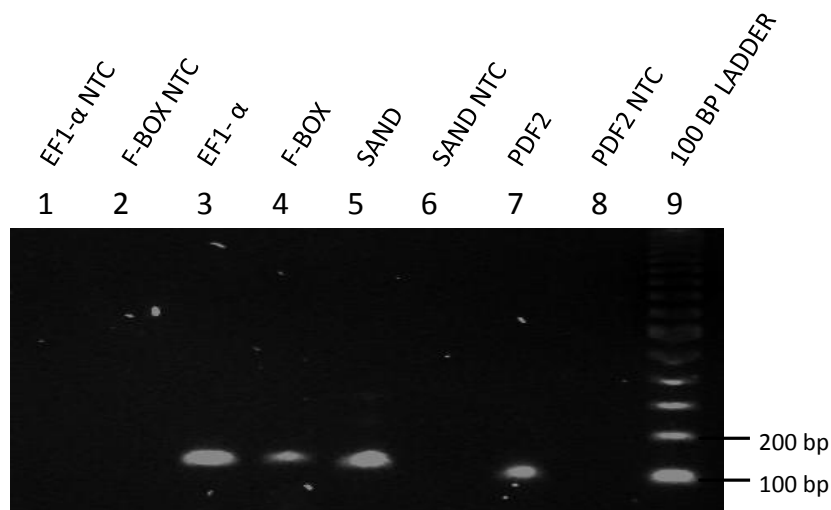


**Figure 4.10. Reference genes selected for evaluation for qPCR experiments in Arabidopsis.** Lanes 1 and 2: EF1 α and F-BOX RT-PCR products, NTC; Lane 3: EF1 α RT-PCR product from Arabidopsis leaf RNA (137 bp), Lane 4: F-BOX RT-PCR product from Arabidopsis leaf RNA (140 bp); Lane 5: SAND RT-PCR product from Arabidopsis leaf RNA (127 bp); Lane 6: SAND RT-PCR product, NTC, Lane 7: PDF2 RT PCR product from Arabidopsis leaf cDNA (109 bp); Lane 8: PDF2 RT-PCR product, NTC; Lane 9: 100 bp ladder (Invitrogen).

*OsCPK1 and 15 primers*

The OsCPK1 and OsCPK15 qPCR primers designed were also tested for their ability to amplify products of the expected size in four rice leaf tissue cDNA (Figure 4.11). All RT-PCR products showed intense bands of the expected sizes: OsCPK1 at 123 bp and OsCPK15 at 139 bp. Similar to the Arabidopsis samples, no genomic DNA was observed in the rice leaf cDNA because only one band of the expected size was observed in all samples. RT-PCR products from rice leaf sample one appeared to be a result of degraded RNA template. This leaf RNA sample was not included in succeeding experiments.
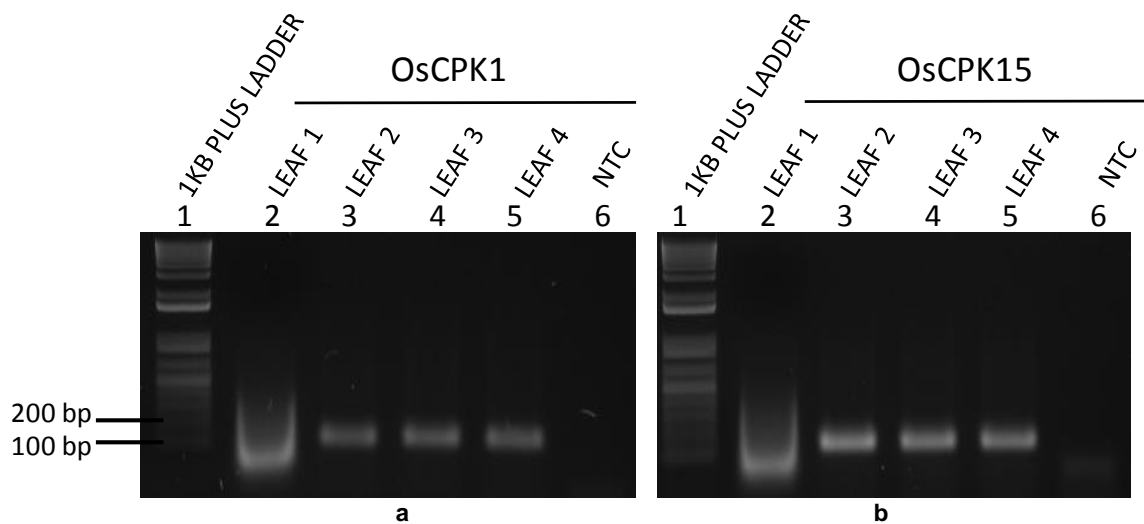


**Figure 4.11. Testing of OsCPK1 and OsCPK15 qPCR primers. (a) OsCPK1 qPCR primer.** Lane 1: 1 Kb Plus ladder (Invitrogen); Lanes 2 to 5: OsCPK1 RT-PCR products from rice leaf RNA samples, with RT (123 bp); Lane 6: OsCPK1 RT-PCR product, NTC; **(b) OsCPK15 qPCR primer.** Lane 1: 1 Kb Plus ladder (Invitrogen); Lanes 2 to 5: OsCPK15 RT-PCR products from rice leaf RNA samples, with RT (139 bp); Lane 6: OsCPK15 RT-PCR product, NTC.

*Rice reference genes primers*

In rice, four reference gene primer sets were selected from previous reports that identified reference genes showing stable expression under biotic and abiotic stress treatments (Maksup et al. 2013; Narsai et al. 2010). Three of the reference genes showed RT-PCR products from rice leaf samples of the expected sizes: 112 bp for OsEP1, 111 bp for OsTBC and 113 bp for OsTPH (Figure 4.12). No RT-PCR product was observed for OsRBP primers (120 bp). OsTBC showed more intense bands than OsEP1 and OsTPH, assuming equal amounts of template. To optimise the use of qPCR materials, only two out of the four genes, OsTBC and OsEP1 were selected for use in the subsequent qPCR experiments. OsTBC was selected because it had the lowest M-

value (0.389) and highest PCR efficiency (1.697) in an initial RT-qPCR run with random samples of rice cDNA from abiotic and biotic stress treatment. OsEP1 was selected because it was identified as a good reference gene in both of the previous studies (Maksup et al. 2013; Narsai et al. 2010) and because it had the second highest PCR efficiency (1.653) and a good M-value (0.537) in the initial qPCR run.
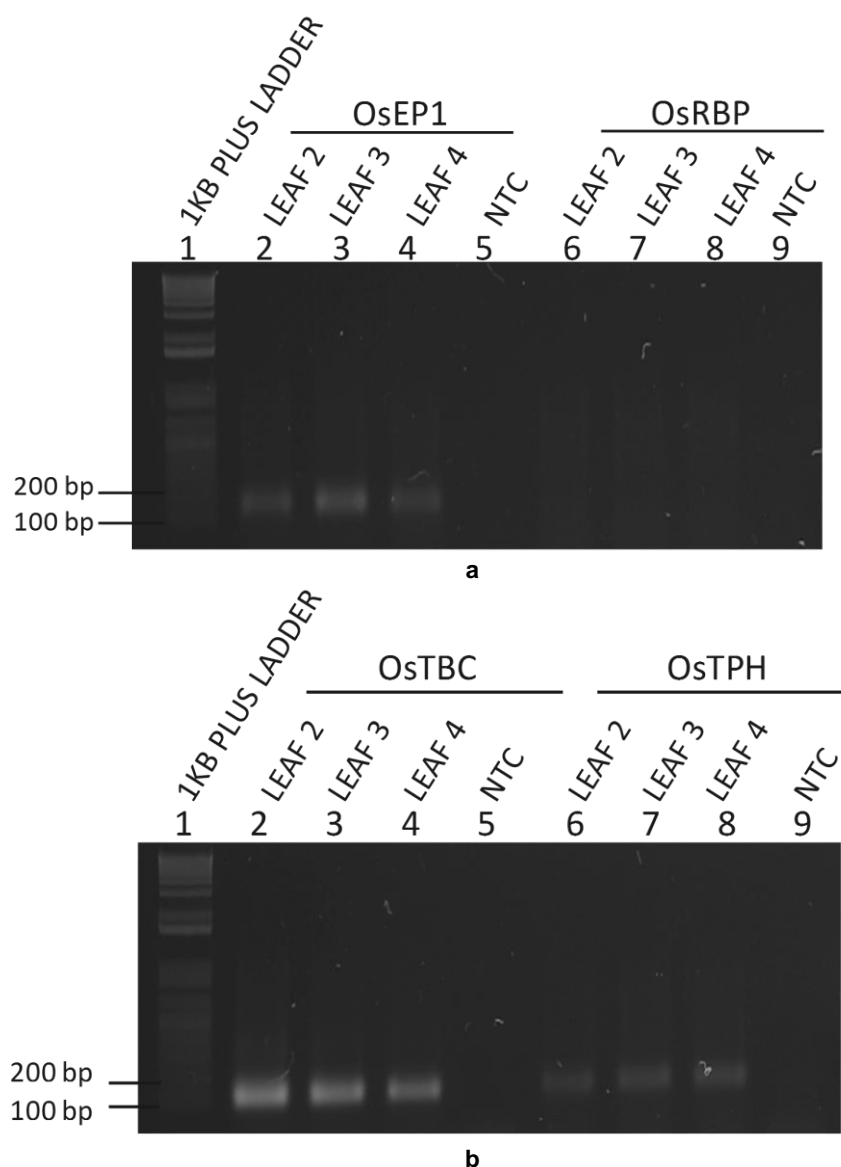


**Figure 4.12. Reference genes selected for evaluation for qPCR experiments in rice. (a) OsEP1 and OsRBP qPCR primer.** Lane 1: 1 Kb Plus ladder (Invitrogen); Lanes 2 to 4: OsEP1 RT-PCR products from rice leaf RNA samples, with RT (112 bp); Lane 5: OsEP1 RT-PCR product, NTC; Lanes 6 to 8: OsRBP RT-PCR products from rice leaf RNA samples, with RT (120 bp expected, but no product seen); Lane 9: OsRBP RT-PCR product, NTC; **(b) OsTBC and OsTPH qPCR primer.** Lane 1: 1 Kb Plus ladder (Invitrogen); Lanes 2 to 4: OsTBC RT-PCR products from rice leaf RNA samples, with RT (111 bp); Lane 5: OsTPH RT-PCR product, NTC; Lanes 6 to 8: OsTPH RT-PCR products from rice leaf RNA samples, with RT (113 bp); Lane 9: OsTPH RT-PCR product, NTC.

*AcCPK16 primer*

The AcCPK16 primer set designed was also tested for its ability to amplify the product of correct size (Figure 4.13, lane 10). The RT-PCR product was of the expected size at 148 bp. The presence of a single RT-PCR product of expected size indicated the absence of genomic DNA.

*Kiwifruit reference genes primers*

In kiwifruit, nine reference gene primer sets were selected for evaluation from previous reports: AdUBQ11 (142 bp), AdTUA (201 bp), AdActin (Wu) (197 bp), AdActin (Zhang) (size not mentioned), UBC9 (190 bp), AdActin (Li) (155 bp), PPC2A (size not mentioned), GAPDH (size not mentioned) and PPPRSA (110 bp) (Bulley et al. 2009; Li et al. 2013; Li et al. 2010; Walton et al. 2009; Wu et al. 2012; Yin et al. 2009; Zhang et al. 2006). All of the primer sets showed distinct RT-PCR products of the expected size (Figure 4.13). However, the bands for AdActin (Zhang) and GAPDH primers were faint compared to the other RT-PCR products (Figure 4.13). To optimise the use of qPCR materials, only three out of the nine genes, AdActin (Wu), UBC9, and PPPRSA were selected for use in the subsequent qPCR experiment, because these were the most stably expressed among the other genes (Bulley et al. 2009; Li et al. 2013; Wu et al. 2012).
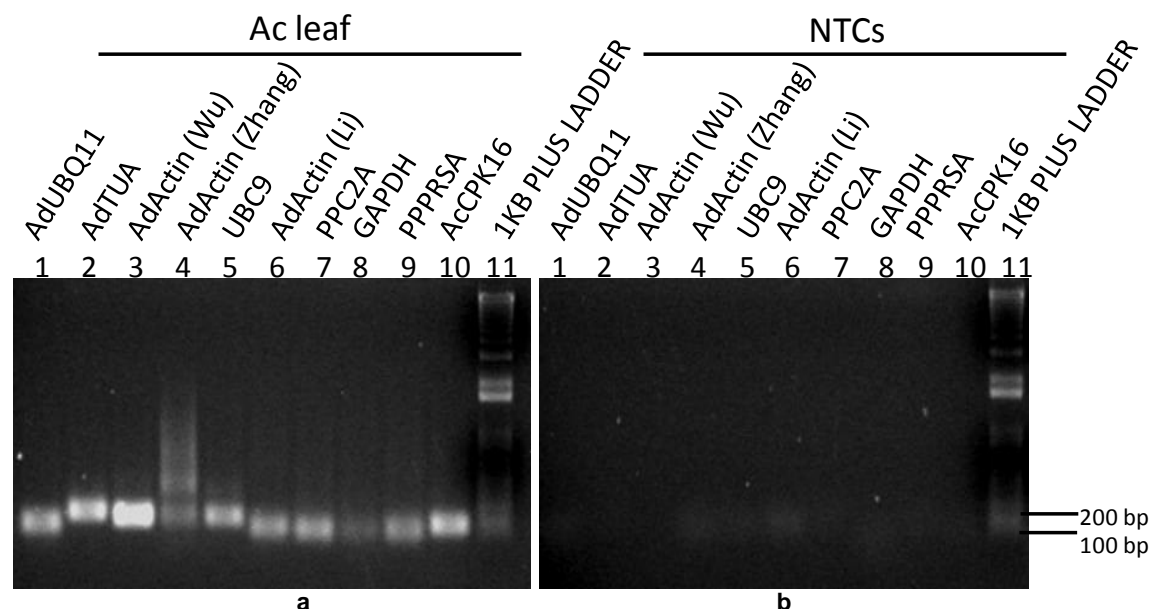


**Figure 4.13. Reference genes used for qPCR experiments in kiwifruit. (a) RT-PCR products from kiwifruit reference genes and AcCPK16 primers**. Lanes 1 to 9: RT-PCR products of tested reference gene primers from kiwifruit leaf RNA, with RT (142 bp for AdUBQ11, 201 bp for AdTUA, 197 bp for AdActin (Wu), ~ 200 bp for AdActin (Zhang), 190 bp for UBC9, 155 bp for AdActin (Li), ~140 bp for PPC2A and GAPDH and 110 bp for PPPRSA; Lane 10: AcCPK16 RT-PCR product from kiwifruit leaf RNA (148 bp); Lane 11: 1 Kb plus ladder; **(b) NTCs of the same kiwifruit primer sets.** Lanes same as (a) but with no template.

*Expression stability of reference genes*

The expression stability (M-value) of the reference genes were calculated for each qPCR run using an algorithm called GeNORM. The total number of qPCR runs (or 384-well qPCR plates) performed to determine mRNA accumulation of Group IIb.1 CPKs under normal controls and under abiotic and biotic was eleven for Arabidopsis, five for rice, and three for kiwifruit. Reference genes showing M-values lower than 1.5 are considered as stable and acceptable for qPCR analysis (Gu et al. 2011; Lilly et al. 2011; Lovdal and Lillo 2009; Mascia et al. 2010; Migocka and Papierniak 2011; Paolacci et al. 2009). All the reference genes used for Arabidopsis, rice and kiwifruit qPCR experiments showed highly acceptable M-values, with averages ranging between 0.42 and 0.62.
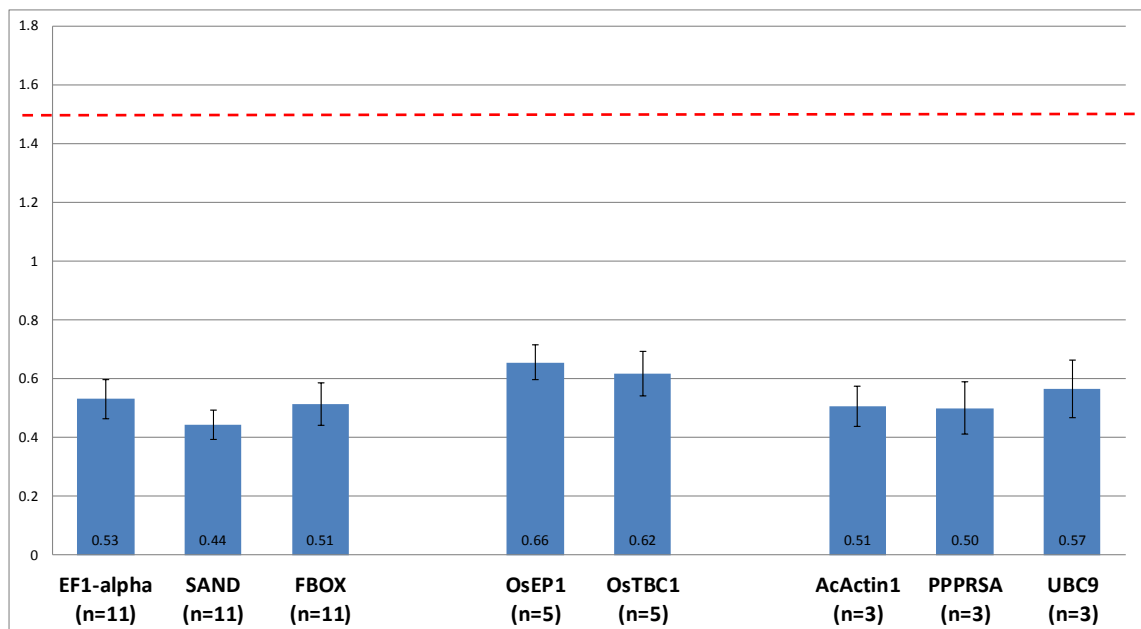


**Figure 4.14. Mean M-values of the reference genes used in Arabidopsis, rice and kiwifruit.** Error bars are the SE of the mean values between qPCR plate runs. n= the number of qPCR plate runs performed, with about 48 samples (in duplicates) per plate. An M-value of 1.5 or less is considered stably expressed and acceptable for use as reference gene for qPCR experiments (red broken line).

### 4.3.2.2b Abiotic stress responses of Group IIb.1 CPKs

In the time series experiments carried out, AtCPK3 and its orthologues in rice and kiwifruit showed changes in transcript accumulation in response to drought, salt and mannitol treatments, but only at particular time points. The amount of changes detected in the qPCR experiments were mostly less than two-fold, but the values come from three biological replicates with averages having marked differences (based on SE calculations) and many having statistically significant differences (based on one-way ANOVA and posthoc tests such as Tukey's method and Fisher's LSD method) (Appendix 35). The results correspond with some of the microarray data presented in section 4.3.2.1, although there were also contrasting responses.

In Arabidopsis grown in soil and in agar, AtCPK3 showed marked differences in transcript accumulation in leaves in response to salt and drought at the following time points: about a 1.5-fold decrease in response to drought at 7 d and 14 d in soil (Figure 4.15a), about 1.3-fold decrease in response to 200 mM mannitol at 48 h (Figure 4.16a), and a 1.2 to 1.5-fold increase in response to 200 mM salt in MS agar between 15 mins, 1 h and 4 h (Figure 4.16b). There was no support for a significant decrease in AtCPK3 transcript accumulation between 7 d drought and 7 d control (Tukey's $P=0.773$ Fisher's LSD $P=0.172$) and between 14 d drought and 14 d control (Tukey's $P=0.671$, Fisher's LSD $P=0.127$). However, strongly significant difference was determined between 7 d drought and 15 min control (Tukey's $P=0.008$, Fisher's LSD $P=0.001$) and between 14 d drought and 15 min control (Tukey's $P=0.008$, Fisher's LSD $P=0.001$), while all other time points did not exhibit significant difference from 15 min control and from each other. This may provide very weak statistical support to indicate that AtCPK3 transcript in Arabidopsis plants decreases about 1.5 fold in response to drought within 14 d. For mannitol response, a significant difference was only determined using Fisher's LSD, between 15 min mannitol and 1 h control (Fisher's LSD $P=0.039$), but no statistical difference was determined to support the marked decrease of about 1.3-fold observed at 48 h. For salt response in agar, good significant difference was only detected using Fisher's LSD method, between 15 min salt and 15 min control (Fisher's LSD $P= 0.034$), and between 1 h salt and 1 h control (Fisher's LSD $P=0.028$). This supports the marked increase observed at 15 mins and 1 h, but not at 4 h.

Root tissues showed approximately 1.2 to 1.4-fold decrease in AtCPK3 transcript in response to 200mM mannitol at 15 min and 1 h (Figure 4.17a) and a 1.2-fold increase in response to 200 mM salt at 24 h (Figure 4.16b). For mannitol response, weak support for a significant

148
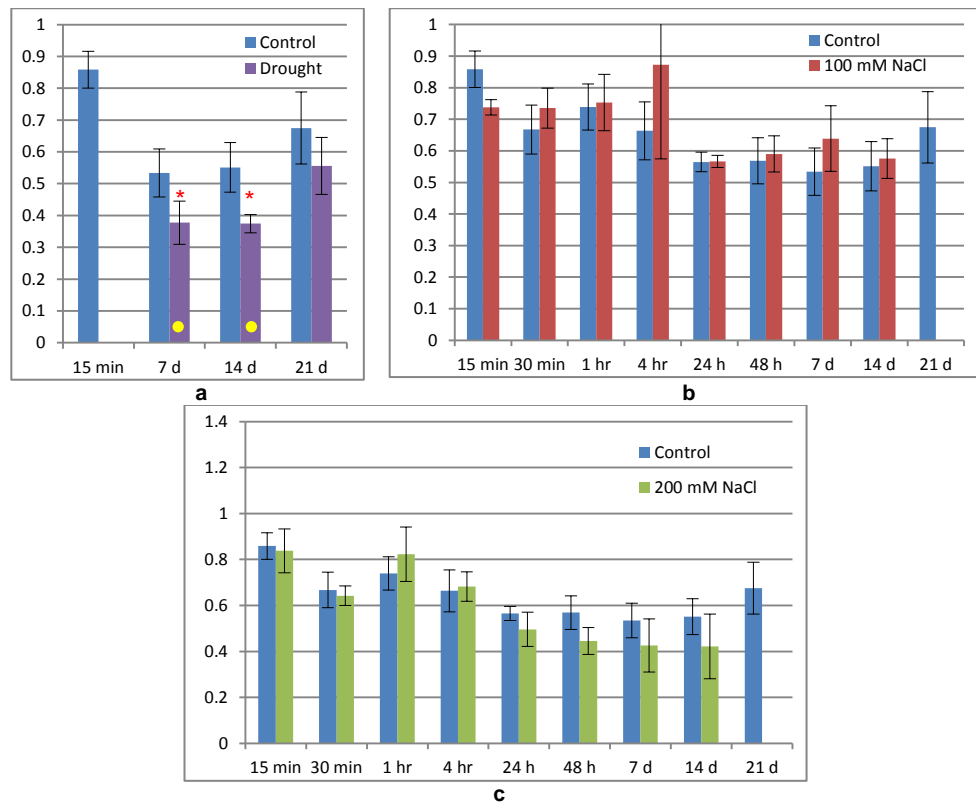
**Figure 4.15. AtCPK3 transcript accumulation in Arabidopsis leaves in response to (a) drought, (b) 100 mM salt and (c) 200 mMsalt, plants grown in soil.** Colour of bars match the treatment carried out: blue, control; purple, drought treatment; red, 100 mM NaCl treatment; green, 200 mM NaCl treatment. Line bars indicate SE of the mean. Yellow dots indicate a marked difference between control and treatment. Statistical support is indicated as: strong (***, P≤ 0.01), good (**, 0.01<P≤ 0.05) or weak (*, 0.05<P<~0.10).
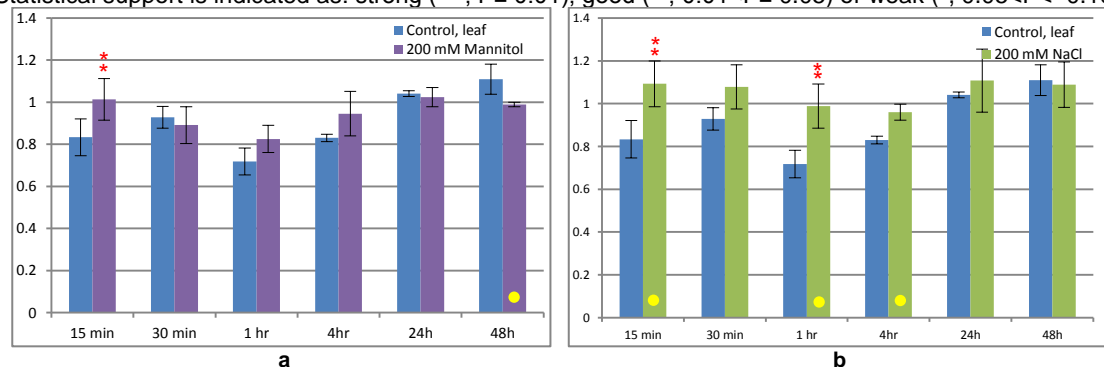


**Figure 4.16. AtCPK3 transcript accumulation in Arabidopsis leaves in response to (a) 200 mM mannitol and (b) 200 mM salt, plants grown in MS agar.** Colour of bars match the treatment carried out: blue, control; purple, mannitol treatment; green, 200 mM NaCl treatment. Line bars indicate SE of the mean. Yellow dots indicate a marked difference between control and treatment. Statistical support is indicated as: strong (***, P≤ 0.01), good (**, 0.01<P≤ 0.05) or weak (*, 0.05<P<~0.10).
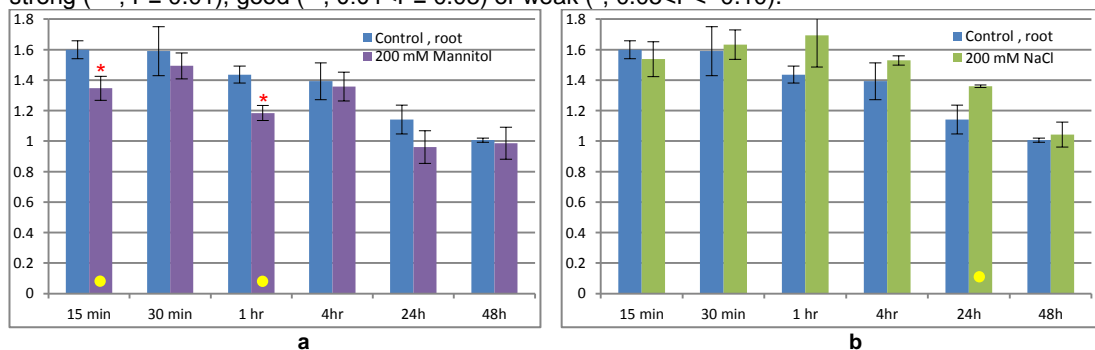


**Figure 4.17. AtCPK3 transcript accumulation in Arabidopsis roots in response to (a) 200 mM mannitol and (b) 200 mM salt, plants grown in MS agar.** Colour of bars match the treatment carried out: blue, control; purple, mannitol treatment; green, 200 mM NaCl treatment. Line bars indicate SE of the mean. Yellow dots indicate a marked difference between control and treatment. Statistical support is indicated as: strong (***, P≤ 0.01), good (**, 0.01<P≤ 0.05) or weak (*, 0.05<P<~0.10).

149

difference was determined using Fisher's LSD, between 15 min mannitol and 15 min control (P=0.073) and 1 h mannitol and 1 h control (P=0.074). For salt response, no significant difference was detected.

In rice, OsCPK1 and 15 showed similar and contrasting results with AtCPK3 (Figure 4.18a and b). As mentioned earlier, AtCPK3 showed about a 1.5-fold decrease in response to drought at 7 d and 14 d. In contrast to that, OsCPK1 showed a 1.5 fold increase in response to drought at 14d (Figure 4.18a). This had good statistical support with Fisher's LSD (P=0.024) but not with Tukey's test (P=0.220). It is however notable that there were either good or weak statistical evidences to support the difference between the 14 d drought samples and most of the controls at 30 min, 24 h, 48 h and 7d (Tukey's P=0.021, 0.116, 0.011 and 0.086; Fisher's LSD P= 0.002, 0.011, 0.001, and 0.008). In response to salt, AtCPK3 showed a 1.2 to 1.5-fold increase in response to 200 mM salt in MS agar between 15 mins, 1 h and 4 h. Similar to that, OsCPK1 showed about 1.5-fold increase in transcript accumulation, but it was only observed at 14 d in soil grown plants. This had strong statistical support with Fisher's LSD (P=0.004) but no statistical support with Tukey's test (P= 0.146). It is nevertheless notable that only the 14 d treatment samples showed significant difference to the controls at 15 min, 30 min, 4 h, 24 h, 48 h, and 7 d with Tukey's test (P=0.004, 0.002, 0.013, 0.012, 0.043, and 0.024, respectively) and Fisher's LSD (all P<0.01). This indicates good statistical support regarding the increase of OsCPK1 in response to salt within 14 d. A slight decrease in OsCPK1 was also observed in response to salt at 4 h and 24 h but did not have any statistical support (Tukey's P= 0.934 and 0.991 respectively; Fishers LSD P=0.123, and 0.231 respectively).

On the other hand, OsCPK15 showed a two-fold decrease in response to drought at 14d and a two-fold increase in response to 200mM salt at 7 and 14 d (Figure 4.18b), which was similar to the response of AtCPK3. The marked decrease in OsCPK15 at 14 d had strong statistical support with Fisher's LSD (P=0.030) but not with Tukey's test (P=0.262). Nevertheless, the difference between drought samples at 14 d and controls in other timepoints such as 30 min, 48 h and 7 d controls were also significant (Fisher's LSD P= 0.066, 0.042 and 0.089 respectively). In response to salt, the two-fold increase at 7 d had strong statistical support (Tukey's P=0.002, Fisher's LSD P=0.000), and at 14 d had good statistical support for Fisher's LSD (P= 0.028), but not for Tukey's (P=0.553).
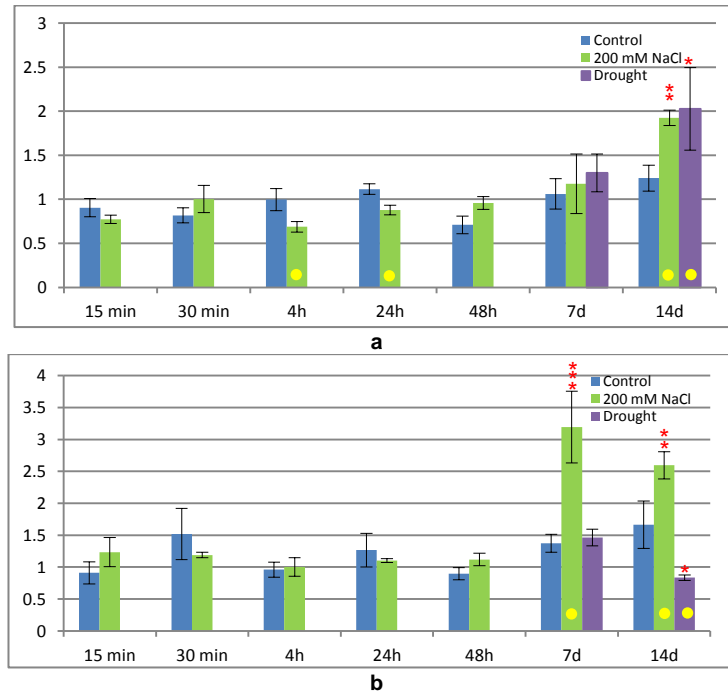
**Figure 4.18. OsCPK1 (a) and OsCPK15 (b) transcript accumulation in rice leaves in response to drought and salt, plants grown in soil.** Colour of bars match the treatment carried out: blue, control treatment samples; green, 200 mM NaCl treatment samples; purple, drought treatment. Line bars indicate SE of the mean. Yellow dots indicate a marked difference between control and treatment. Statistical support is indicated as: strong (***, P≤ 0.01), good (**, 0.01<P≤ 0.05) or weak (*, 0.05<P<~0.10).
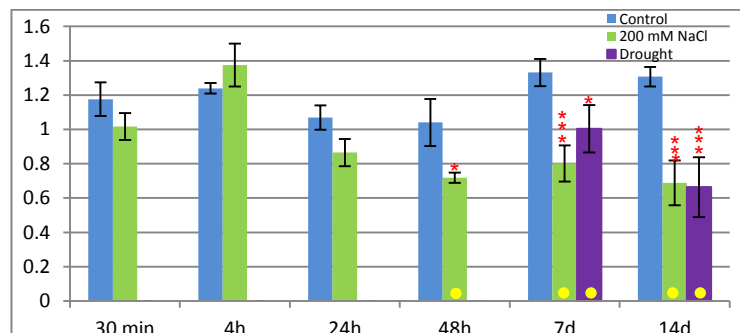


**Figure 4.19. AcCPK16 transcript accumulation in kiwifruit leaves in response to drought and salt, plants grown in soil.** Colour of bars match the treatment carried out: blue, control treatment samples; green, 200 mM NaCl treatment samples; purple, drought treatment. Line bars indicate SE of the mean. Yellow dots indicate a marked difference between control and treatment. Statistical support is indicated as: strong (***, P≤ 0.01), good (**, 0.01<P≤ 0.05) or weak (*, 0.05<P<~0.10).

In kiwifruit, AcCPK16 showed a marked decrease in transcript accumulation in response to drought at 7d (1.3-fold) and 14 d (1.5-fold). This response to drought was similar with AtCPK3 and OsCPK15, which both showed decrease in transcript accumulation. There was weak statistical support for the decrease at 7 d (Fisher's LSD P= 0.063) but strong statistical support for the decrease at 14 d (Tukey's P=0.019, Fisher's LSD P= 0.001).

In response to 200mM salt, AcCPK16 showed a marked decrease at 48h (1.5-fold), 7d (1.6-fold) and 14d (1.8-fold). This was different from AtCPK3, OsCPK1 and OsCPK15 which

showed an increase. There was strong statistical support for the decrease at 7 d (Tukey's P=0.017, Fisher's LSD P= 0.000) and 14 d (Tukey's P=0.00; Fisher's LSD P= 0.000), with only weak support at 48 h (Fisher's LSD P= 0.021).

To summarise, the Group IIb.1 CPKs in Arabidopsis, rice and kiwifruit show differential expression in response to drought and salt, but the timing and type of response differ between the genes. In response to drought, decrease in transcript accumulation was observed with AtCPK3, OsCPK15 and AcCPK16, while increase in transcript accumulation was observed with OsCPK1. In response to salt, increase in transcript accumulation was observed with AtCPK3, OsCPK1 and 16, while decrease in transcript accumulation was observed AcCPK16.

### 4.3.2.2c Biotic stress responses of Group IIb.1 CPKs

AtCPK3 and its orthologues in rice and kiwifruit showed changes in transcript accumulation in response to bacterial, fungal and viral pathogens. Similar to abiotic stresses, the amount of changes detected in the biotic stress experiments were mostly less than two-fold, although the fold differences were high (up to six and seven-fold difference) on certain time points of a particular treatment. The results also correspond with some of the microarray data presented in section 4.3.2.1. Most of the values have high SE values due to the uniqueness in the response of each biological replicate. Even though the differences appeared large, most of the values were unidirectional; i.e. either all biological replicates increase or all biological replicates decreased. In addition to the computations described in section 4.2.2.4b, a standardisation method that involves sequential corrections, log transformation, mean centering and autoscaling was done in order to reduce the high variability and draw statistically sound inferences (Willems et al. 2008).

In Arabidopsis, AtCPK3 decreased by about 1.5-fold in response to the fungal pathogen *B. cinerea* at 2 and 6 dpi (Figure 4.20). There was good statistical evidence to support this using Fisher's LSD method (P=0.004 and 0.025 respectively), but not using Tukey's test (P=0.124 and 0.450 respectively). A similar response was observed in the bacterial pathogen Pto DC3000 treatment, with a marked decrease by about 1.3-fold to 1.6 fold at 2 and 10 dpi. There was weak statistical evidence to support the decrease at 2 dpi using Fisher's LSD method (P=0.036) but not using Tukey's test (P=0.548). No statistical evidence was determined to support the marked decrease at 10 dpi.

Upon using the standardisation method by Willems et al. (2008), the fold changes increased and the variability between values were reduced (Figure 4.21). In the standardised computation for the response to *B. cinerea,* the marked differences were between 1.3 to 3.0 fold decrease at 2, 6 and 10 dpi. For Pto DC3000 the marked difference was only at 2 dpi with approximately 1.9-fold decrease. These values could not be tested using ANOVA, due to the difference in the system of computation.
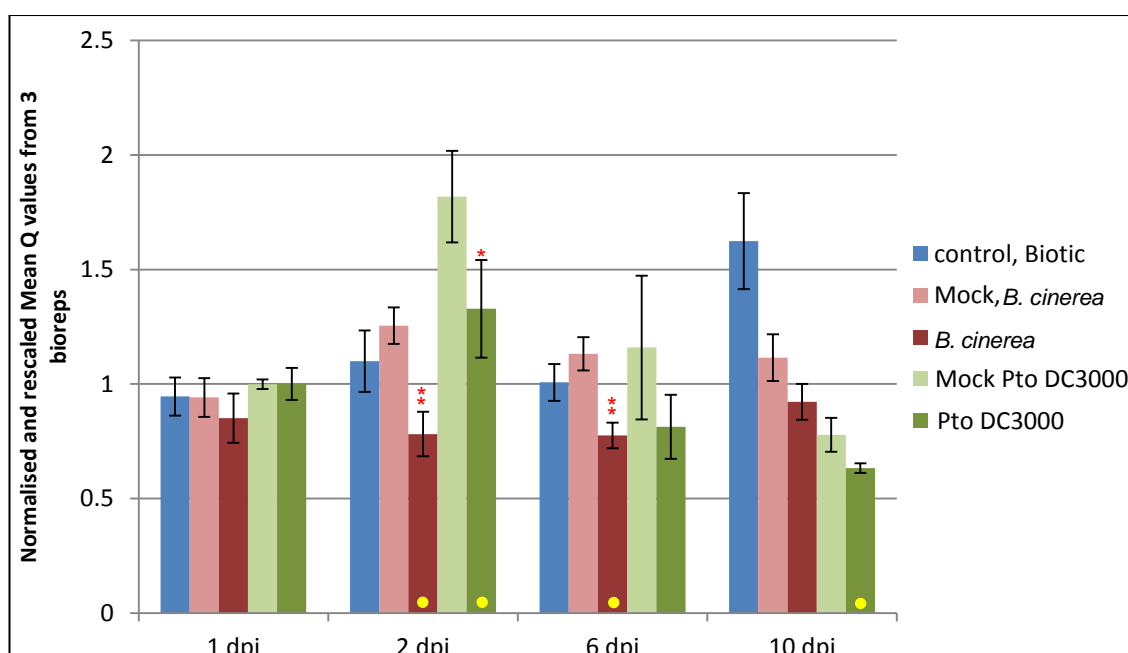


**Figure 4.20. AtCPK3 transcript accumulation in Arabidopsis leaves in response to *B. cinerea* and Pto DC3000, plants grown in soil.** Colour of bars match the treatment carried out: blue, control treatment samples (no treatment); pink, mock *B. cinerea* treatment (PDA broth); red, *B. cinerea* treatment samples; light green, mock Pto DC3000 treatment; green, Pto DC3000 treatment. Line bars indicate SE of the mean. Yellow dots indicate a marked difference between control and treatment. Statistical support is indicated as: strong (***, P≤ 0.01), good (**, 0.01<P≤ 0.05) or weak (*, 0.05<P<~0.10).
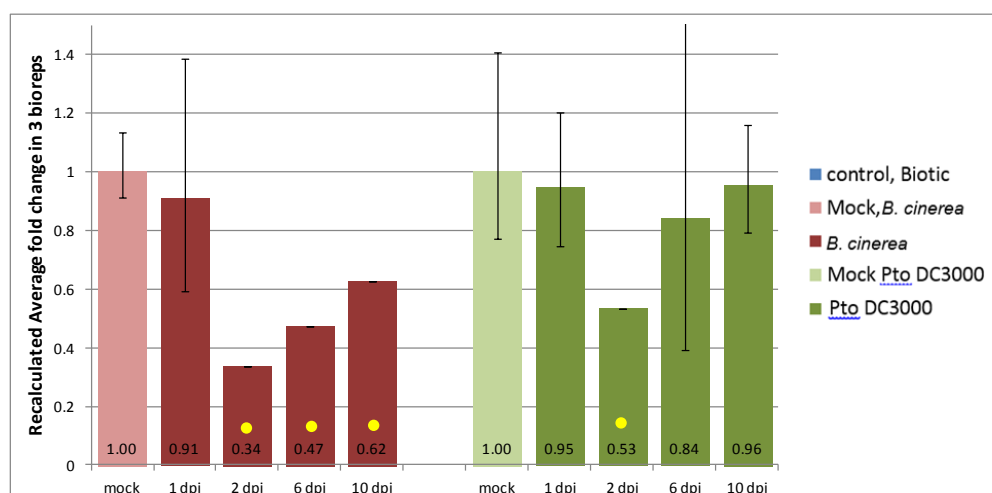


**Figure 4.21. AtCPK3 transcript accumulation in Arabidopsis leaves in response to *B. cinerea* and Pto DC3000 after being Log transformed, mean-centered and autoscaled fold changes relative to control.** Colour of bars match the treatment carried out: pink, mock *B. cinerea* treatment (PDA broth); red, *B. cinerea* treatment samples; light green, mock Pto DC3000 treatment; green, Pto DC3000 treatment. Line bars indicate SE of the mean. Yellow dots indicate a marked difference between control and treatment.

In the rice detached leaf assay, OsCPK1 decreased in response to infection with a fungal pathogen, *M. grisea* (Figure 4.22). This was similar to AtCPK3, but the decrease in OsCPK1 was only 1.3-fold at 10 dpi. There was weak statistical evidence to support this decrease using Fisher's LSD method (P=0.066) but not using Tukey's test (P=0.638). No marked difference was observed in OsCPK1 in response to Pss.

On the other hand, OsCPK15 increased about two-fold in response to *M. grisea* at 6 and 10 dpi. The marked difference at 6 dpi was supported by weak statistical evidence using Fisher's LSD method (P= 0.080) but not using Tukey's test (P=0.698).The marked difference at 10 dpi was supported by strong statistical evidence using Fisher's LSD method (P=0.009) but not using Tukey's test (P=0.176). OsCPK15 also increased about two-fold at 2 and 6 dpi and about 5.7-fold at 10 dpi in response to Pss. There was a very weak statistical evidence to support the increase at 6 dpi (Fisher's LSD P= 0.094), but there was a very strong statistical evidence to support the increase at 10 dpi (Tukey's P= 0.000; Fisher's LSD P= 0.000).



(a)



(b)
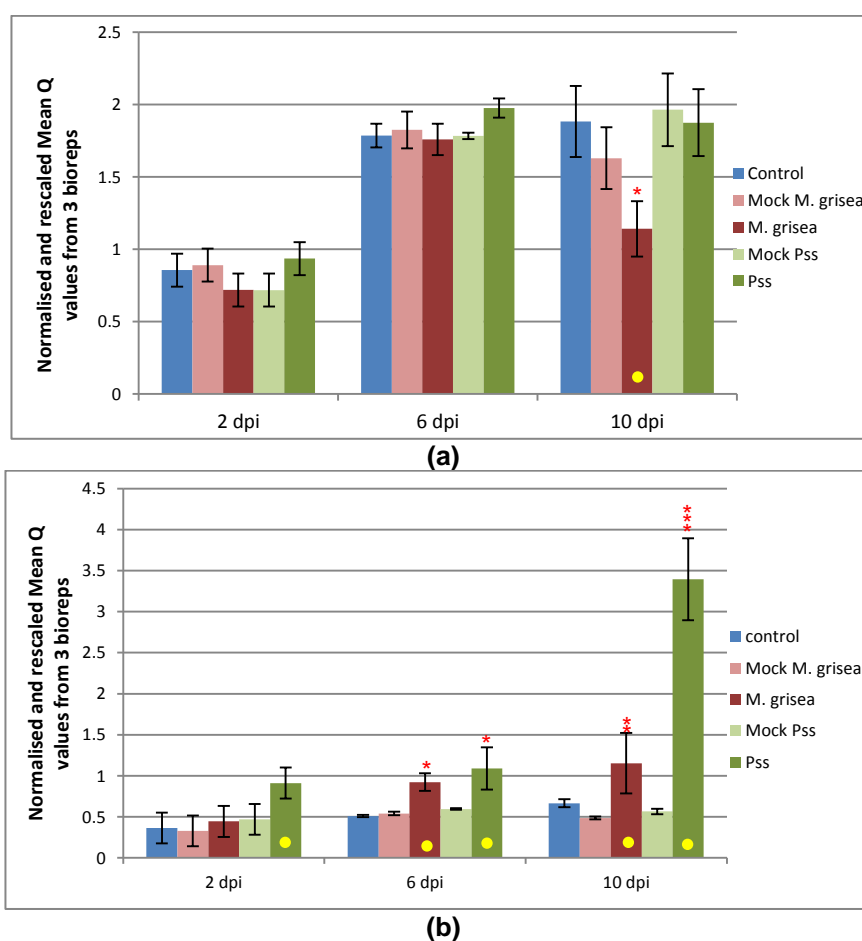
**Figure 4.22. OsCPK1 and 15 transcript accumulation in rice leaves in response to *M. grisea* and Pss, plants grown in soil. (a) OsCPK01. (b) OsCPK15.** Colour of bars match the treatment carried out. blue, control treatment samples (no treatment); light green, mock Pss treatment; green, Pss treatment; pink, mock *M. grisea* treatment (PDA broth); red, *M. grisea* treatment samples. Line bars indicate SE of the mean. Yellow dots indicate a marked difference between control and treatment. Statistical support is indicated as: strong (***, P≤ 0.01), good (**, 0.01<P≤ 0.05) or weak (*, 0.05<P<~0.10).

154

In kiwifruit, AcCPK16 only exhibited a marked difference in transcript accumulation in response to *B. cinerea* at day 10, with about 1.2-fold decrease compared to mock. However, there was no statistical evidence to support this (Tukey's P=0.641, Fisher's LSD P=0.147). As mentioned in section 4.2.2.2c, infection with *Pseudomonas sp.* or any bacteria that may infect kiwifruit was not performed due to restrictions in biological safety and limitations in the research facility.
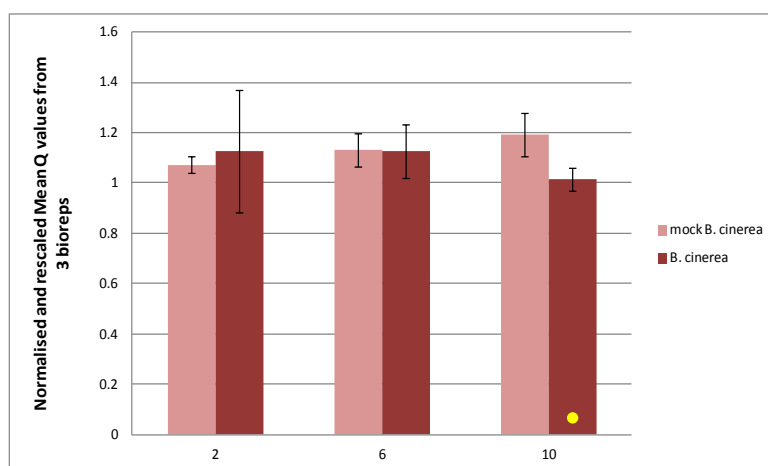


**Figure 4.23. AcCPK16 transcript accumulation in kiwifruit leaves in response to *B. cinerea*, plants grown in soil.** Colour of bars match the treatment carried out: pink, mock *B. cinerea* treatment (PDA broth); red, *B. cinerea* treatment samples Line bars indicate SE of the mean. Yellow dots indicate a marked difference between control and treatment.

In response to five viruses, AtCPK3 showed different levels of increase in transcript accumulation in leaves between 7 and 35 dpi (Figure 4.24). AtCPK3 appeared to fluctuate within the first two days of inoculation, as demonstrated by the high level of variation between the biological replicates. In response to CaMV, a marked increase was seen at 14 to 35 dpi, with 1.6-fold increase at 14 dpi, 1.8 -fold at 21 dpi, up to three-fold increase at 21 and 35 dpi. There was weak statistical evidence to support the marked increase at 14 dpi (Fisher's LSD P=0.082) while there was good evidence to support the marked increase at 21 and 25 dpi (Fisher's LSD P= 0.040 and 0.045 respectively). There was either good or strong statistical evidence to support the marked increase at 28 dpi depending on the statistical test applied (Tukey's P= 0.080, Fishers LSD P=0.001).

In response to TMV, AtCPK3 decreased by 1.6 fold at 7 dpi but increased to about two-fold at 14 dpi, 1.7-fold at 21 dpi, 1.5 fold at 28 dpi and 2.5 fold at 35 dpi. There was weak statistical

evidence to support the marked increase at 14 dpi (Fisher's LSD P=0.096) while there was strong evidence to support the marked increase at 35 dpi (Tukey's P=0.033, Fisher's LSD P=0.001). The small decrease at 7 dpi was not statistically significant.

In response to TSWV, there was a 2.9 fold increase at 7 dpi, a two-fold increase at 14 dpi, about a three-fold increase at 21 dpi, two-fold increase at 28 dpi, and four-fold increase at 35 dpi. There was statistical support for the marked increase at most of these timepoints. There was good statistical support at 7 dpi (Fisher's LSD P=0.004) and at 14 dpi (Fisher's LSD P=0.027), weak statistical support at 21 dpi (Fisher's LSD P=0.067) and strong statistical support at 35 dpi (Tukey's P=0.033, Fisher's LSD P= 0.001).

In response to TuMV, there was an increase of about 1.6 fold at 14 dpi and only 1.2 fold at 21 dpi. There was weak statistical evidence to support the marked increase at 14 dpi (Fisher's LSD P=0.070) but there was no support for the marked increase at 21 dpi. Since plants have completely died at 35 dpi under TuMV infection, no RNA of good quality was isolated from the samples.

In response to TYMV, a three-fold increase in AtCPK3 transcript was observed at 21 dpi, while 1.7-fold increase and 1.8-fold increase were observed at 28 and 35 dpi. The statistical evidence to support the marked increase was good at 21 and 35 dpi (Fisher's LSD P=0.011 and 0.042 respectively) and was strong at 28 dpi ((Tukey's P = 0.000, Fisher's LSD P= 0.000).

The standardisation method by Willems et al. (2008) has reduced the variation as measured by the SE of the mean between biological replicates (Figure 4.25). As with the normalised values, the log-transformed, mean-centred and autoscaled values of AtCPK3 transcript accumulation also generally increased in response to the five viruses starting at 14 dpi until 35 dpi. With the standardisation method, the fold differences were generally higher. Moreover, fewer SE between control and infected samples overlapped. However, these values could not be tested using ANOVA, due to the difference in the system of computation.
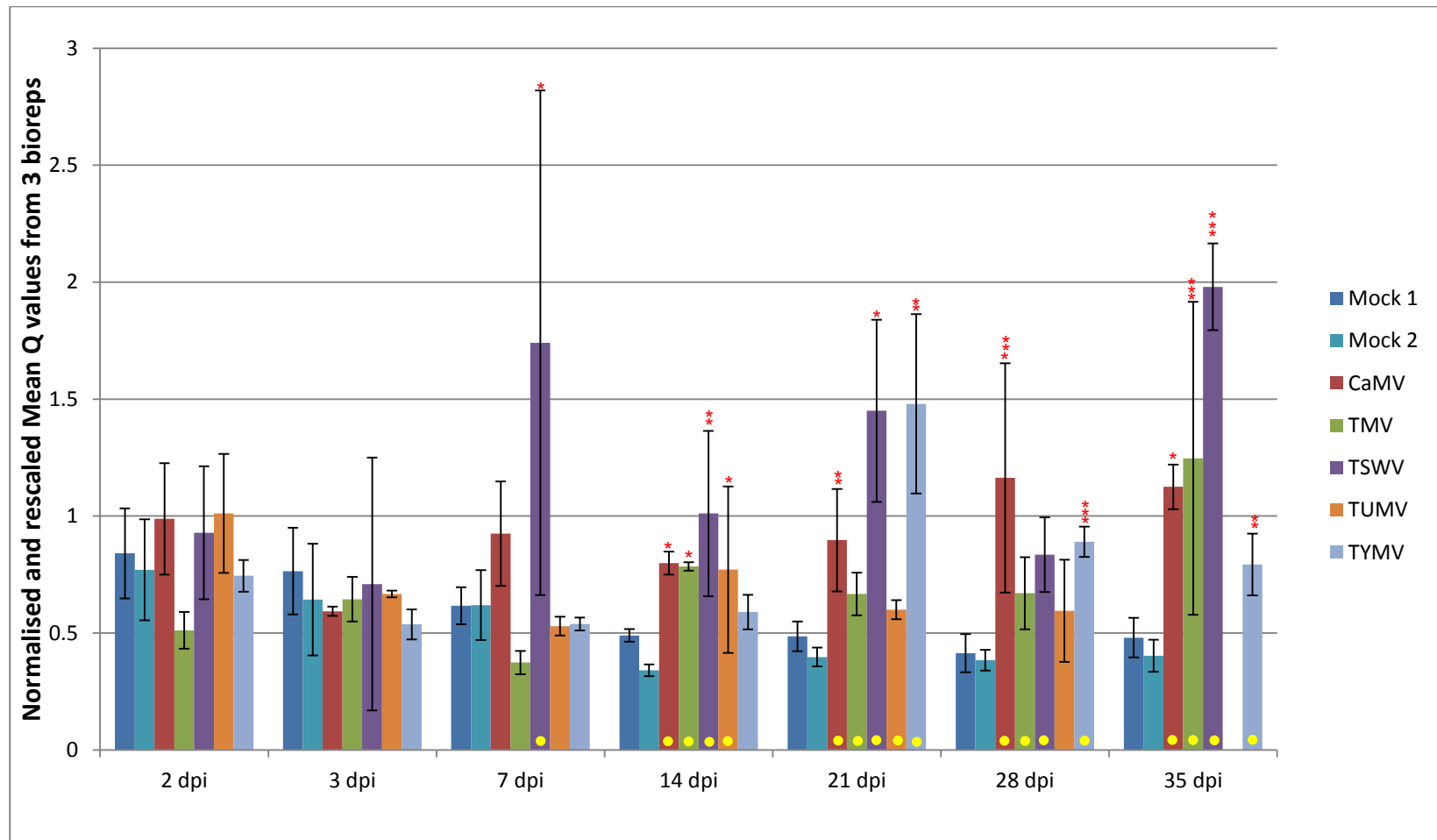
**Figure 4.24. AtCPK3 transcript accumulation in Arabidopsis leaves in response to viruses, plants grown in soil.** Normalised and rescaled mean Q values. Colour of bars match the treatment carried out: blue, mock inoculated 1; teal, mock inoculated 2; red, CaMV; green, TMV; purple, TSWV; orange, TUMV; light blue, TYMV. Line bars indicate SE of the mean. Yellow dots indicate a marked difference between control and treatment. Statistical support is indicated as: strong (***, P≤ 0.01), good (**, 0.01<P≤ 0.05) or weak (*, 0.05<P<~0.10).
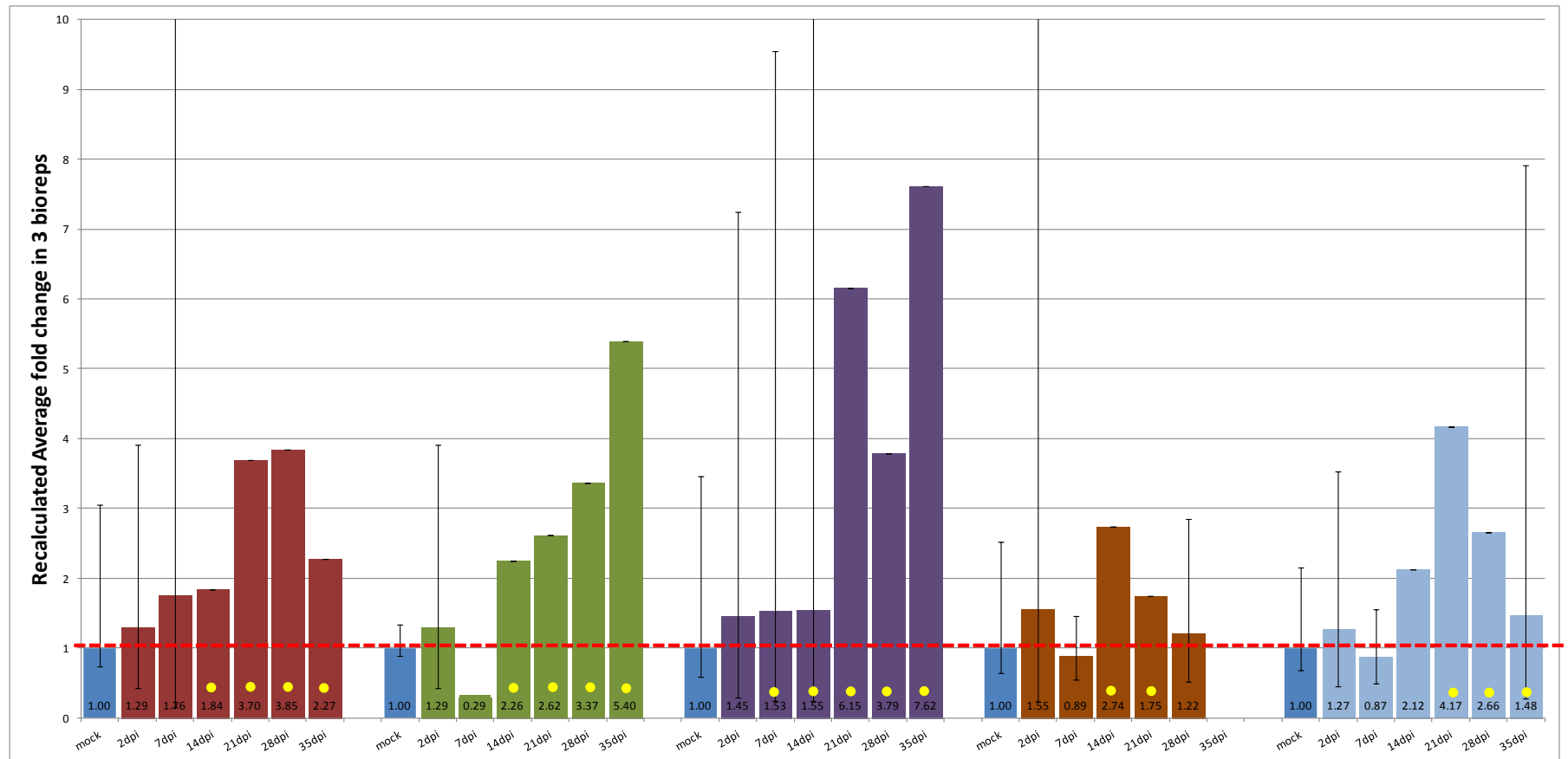
**Figure 4.25. AtCPK3 transcript accumulation in Arabidopsis leaves in response to viruses, plants grown in soil. Recalculated average fold change in three biological replicates.** Colour of bars match the treatment carried out: blue, mock inoculated; red, CaMV; green, TMV; purple, TSWV; orange, TUMV; light blue, TYMV. Line bars indicate SE of the mean. Yellow dots indicate a marked difference between control and treatment.

The five viruses used in Arabidopsis were not known to infect rice or kiwifruit. Therefore, different viruses were used in this study to infect the two selected species. CymMV, which is a common virus affecting orchids, was used for rice because this virus in present in New Zealand (Pearson et al. 2006) and can infect members of the Poaceae family to which rice belongs (Lapierre and Signoret 2004). CMV was used because this virus was reported to infect kiwifruit (Blouin et al. 2013) and can also infect other dicot species (therefore can be used in a similar study with another plant species).

OsCPK1 and OsCPK15 showed increased transcript accumulation in response to CymMV (Figure 4.26), although the fold changes were not very high. OsCPK 1 only showed a slight increase (1.4- fold) at 28 dpi, which did not have a statistical support (Tukey's P= 0.980, Fisher's LSD P= 0.255). OsCPK15, on the other hand, continuously showed increase in transcript accumulation: 1.6-fold at 14 dpi, 1.4-fold at 21 dpi and 1.8-fold at 28 dpi. There was good statistical evidence for this marked increase only at 14 dpi (Fisher's LSD P= 0.028).

Similarly, AcCPK16 showed increased transcript accumulation in response to CMV (Figure 4.27), but marked differences were only observed later during infection. An increase of 1.5-fold was observed for AcCPK16 transcript accumulation both at 28 and 35 dpi of CMV infection. These were both supported by good statistical evidence (Fisher's LSD P=0.012 and 0.007 respectively).

To summarise, Group IIb.1 CPKs in Arabidopsis, rice and kiwifruit showed differential expression in response to bacterial, fungal and viral infections. In response to bacterial infections, there was AtCPK3 appeared to be downregulated, while OsCPK1 did not change and OsCPK15 was upregulated. In response to fungal infections, AtCPK3, OsCPK1 and AcCPK16 decreased while OsCPK15 increased in transcript accumulation. In response to viruses, there is a general trend for an increase in transcript accumulation in AtCPK3, OsCPK1, OsCPK15 and AcCPK16.
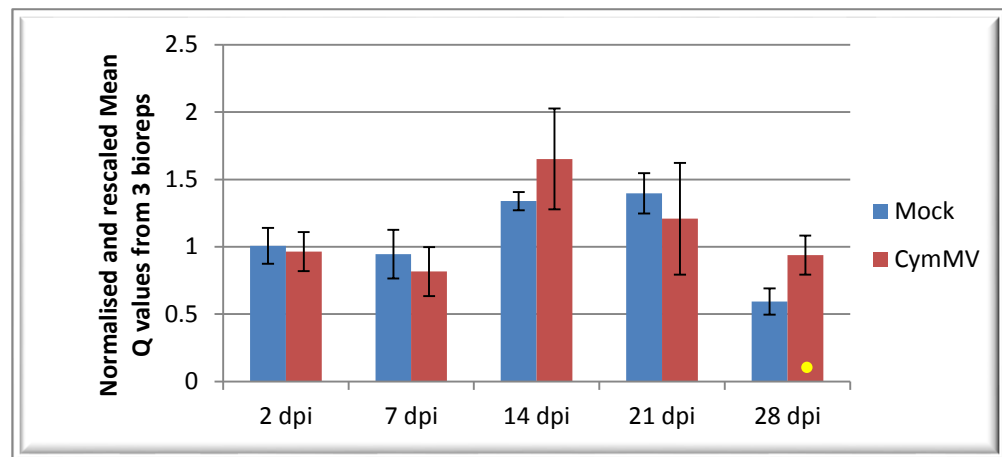
**(a)**
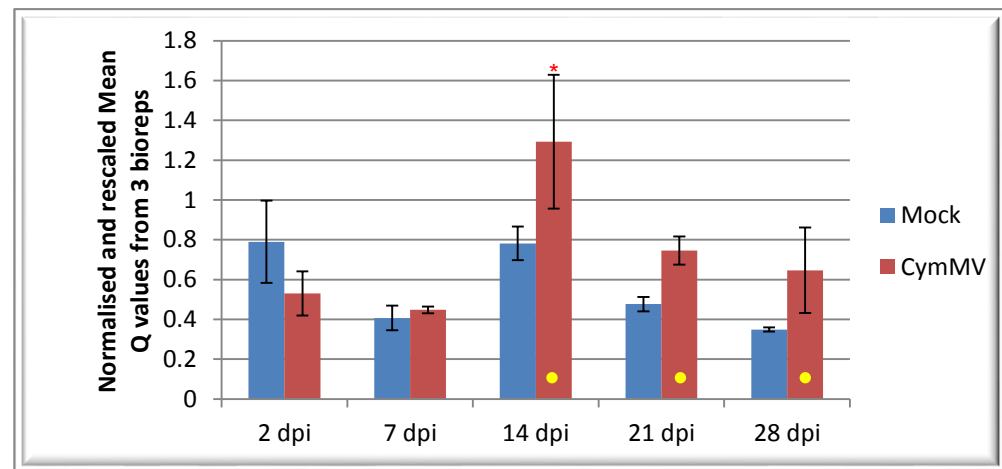


**(b)**

**Figure 4.26. OsCPK1 and 15 transcript accumulation in leaves in response to CymMV, plants grown in soil. Normalised and rescaled mean Q values.** (a) OsCPK01. (b) OsCPK15. Colour of bars match the treatment carried out: blue, mock inoculated; red: CymMV. Line bars indicate SE of the mean. Yellow dots indicate a marked difference between control and treatment. Statistical support is indicated as: strong (***, P≤ 0.01), good (**, 0.01<P≤ 0.05) or weak (*, 0.05<P<~0.10).
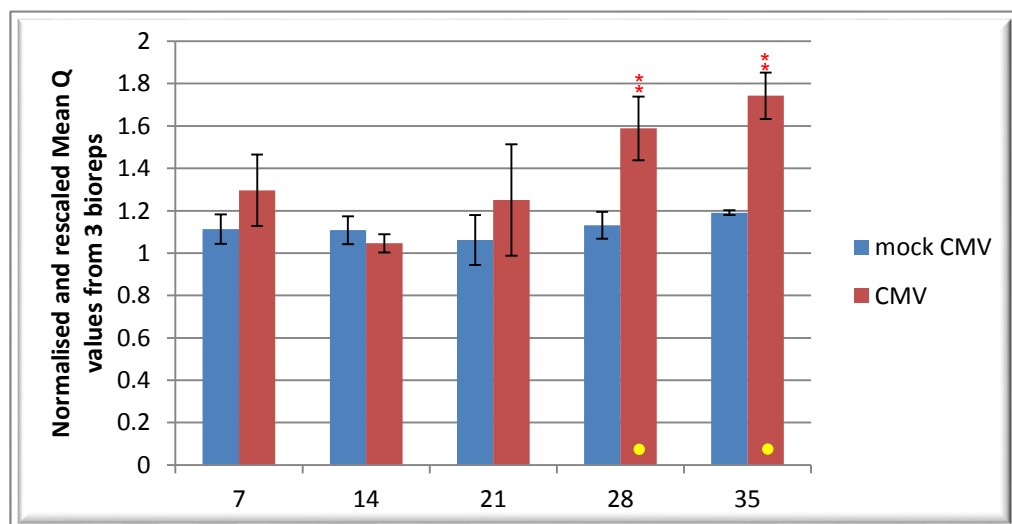


**igure 4.27. AcCPK16 transcript accumulation in leaves in response to CMV, plants grown in soil. Normalised and rescaled mean Q values.** Colour of bars match the treatment carried out: blue, mock inoculated; red: CMV. Line bars indicate SE of the mean. Yellow dots indicate a marked difference between control and treatment. Statistical support is indicated as: strong (***, P≤ 0.01), good (**, 0.01<P≤ 0.05) or weak (*, 0.05<P<~0.10).

**4.3.3 What happens to plants if AtCPK3 and its orthologues are knocked out or overexpressed?**

*4.3.3.1 Arabidopsis knockouts and overexpressors*

*4.3.3.1a Development of AtCPK3 overexpressor lines*

The full AtCPK3 gene was PCR amplified from an Arabidopsis leaf cDNA using AtCPK3 gene specific primers that were flanked with Gateway attB sequence, as described in section 4.2.2.6a. Initial attempts were unsuccessful but upon optimisation of PCR parameters and change of enzyme used (GoTaq polymerase was used instead of *Pfx* polymerase), a PCR product of the expected size (1.6 Kb) was achieved (Figure 4.28a). The PCR product was then gel-purified (QIAquick Gel extraction kit, QIAgen) to remove primer-dimers (Figure 4.28b).
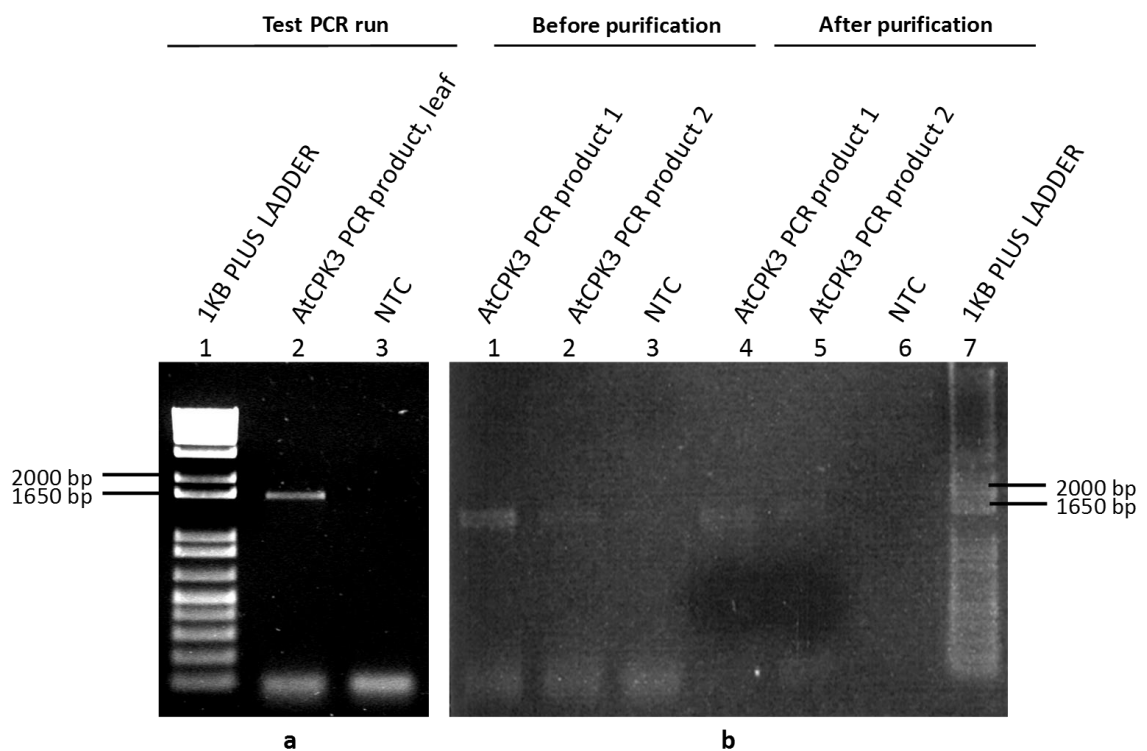


**Figure 4.28. AtCPK3-Gateway PCR product** (a) initial testing of primers; and (b) Gel purification to remove primer-dimers

The Gateway BP reaction that was carried out as described in section 4.2.2.6a was expected to produce entry clones for making AtCPK3 overexpressor lines. However, cloning the AtCPK3 into the entry plasmid was difficult. An initial cloning of the BP reaction products (as described in section 4.2.2.6a) resulted in about $4.64 \times 10^3$ CFU/µg insert DNA, which showed poor efficiency. Furthermore, a colony PCR carried out as described in section 4.2.2.6a using AtCPK3 gene specific primers was expected to result in a PCR product of 1590bp size, but no PCR product of correct size was observed in the colonies screened (Figure 4.29).
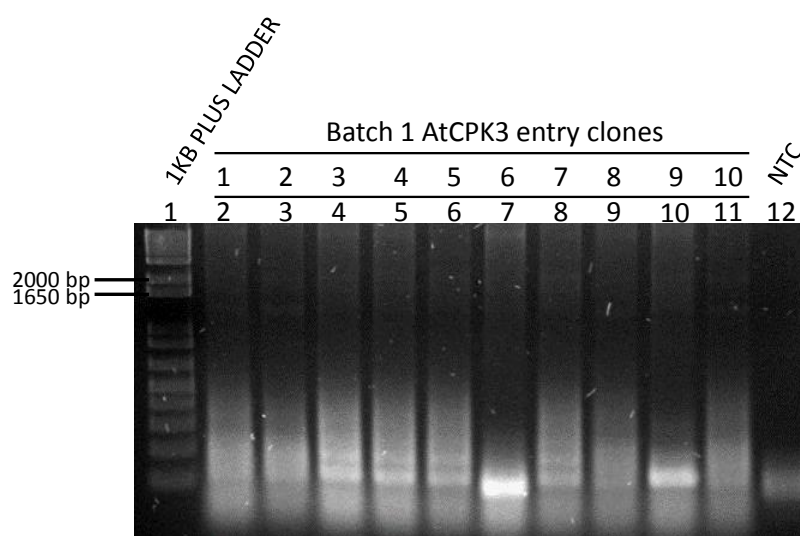


**Figure 4.29. Initial cloning of BP reaction products. 1:1 insert to vector molar ratio**. No positive clones found

The insert: vector molar ratio was then adjusted from 1:1 (ratio suggested in the manufacturer's protocol) to 1:2 and 4:5. The 1:2 insert: vector molar ratio resulted in $1.7 \times 10^4$ CFU/ug insert DNA while the 4:5 ratio resulted in $2.2 \times 10^4$ CFU/ug insert DNA. From the 4:5 ratio set-up, a total of 53 colonies were screened by colony PCR using AtCPK3-specific primers, but only one colony appeared to contain an insert approximately of the expected size of the AtCPK3 insert (1590 bp, indicated by the red arrow in Figure 4.30a, lane 6).
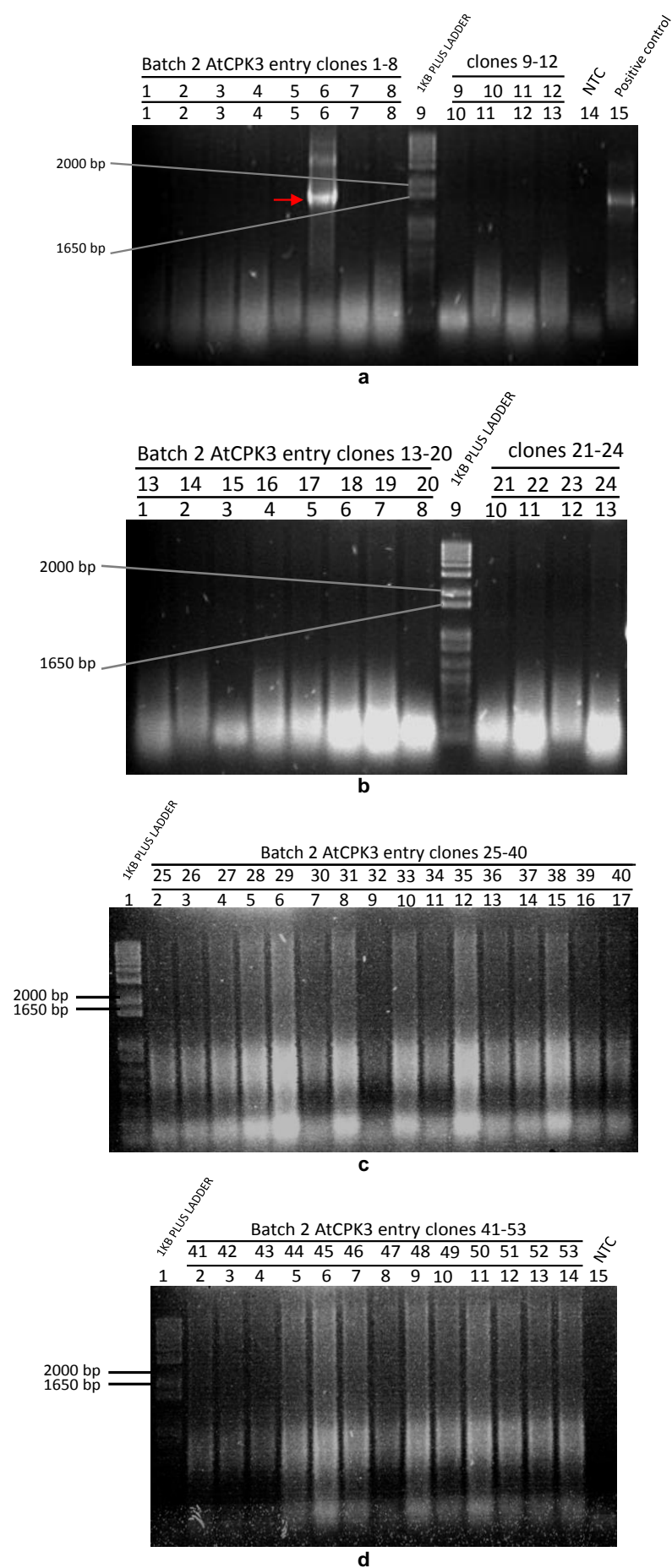
**Figure 4.30. Further cloning of BP reaction products. 4:5 insert to vector molar ratio.** (a) AtCPK3 entry clones 1 to 12. (b) AtCPK3 entry clones 13 to 24. (c) AtCPK3 entry clones 25 to 40. (d) AtCPK3 entry clones 41 to 54. The colony PCR-positive clone is indicated by the red arrow.

163

To further screen for positive colonies and to verify the presence of the AtCPK3 insert in the positive colony above, plasmid DNA isolations and end-point PCRs on the plasmids isolated were carried out as described in section 4.2.2.6a. The positive colony showed an intense PCR product about 1590 bp in size (Figure 4.31, lane 3), whereas four out of five randomly selected colonies (15, 11, 12 and 25) showed two faint bands at about 1590 bp and 2,200 bp (Figure 4.31, lanes 2 and 4 to 6).
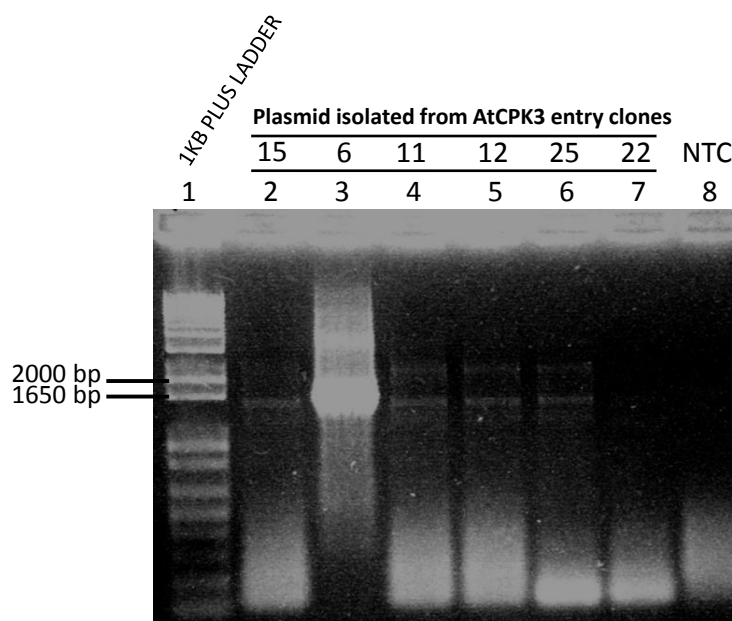


**Figure 4.31. AtCPK3 PCR products of plasmids isolated from positive colony and random colonies.**

Since these plasmids potentially have the insert as indicated by the PCR product size, they were sent for full sequencing as described in section 4.2.2.6a. Colony 15 resulted in a failed reaction while colony 6 had the correct AtCPK3 sequence, but with one base change (different from the published AtCPK3 sequence) at position 1398 (from A to G) which results in an aa change (glutamine to aspartic acid). Figure 32 shows the assemblies of the sequences of these constructs as aligned with the publicly available AtCPK3 sequence, while Figure 33 shows the mismatch between colony 6 and the published AtCPK3 sequence. These constructs were not used and further colony PCR screening was performed from the previous BP cloning (4:5 insert: vector ratio set up) products.
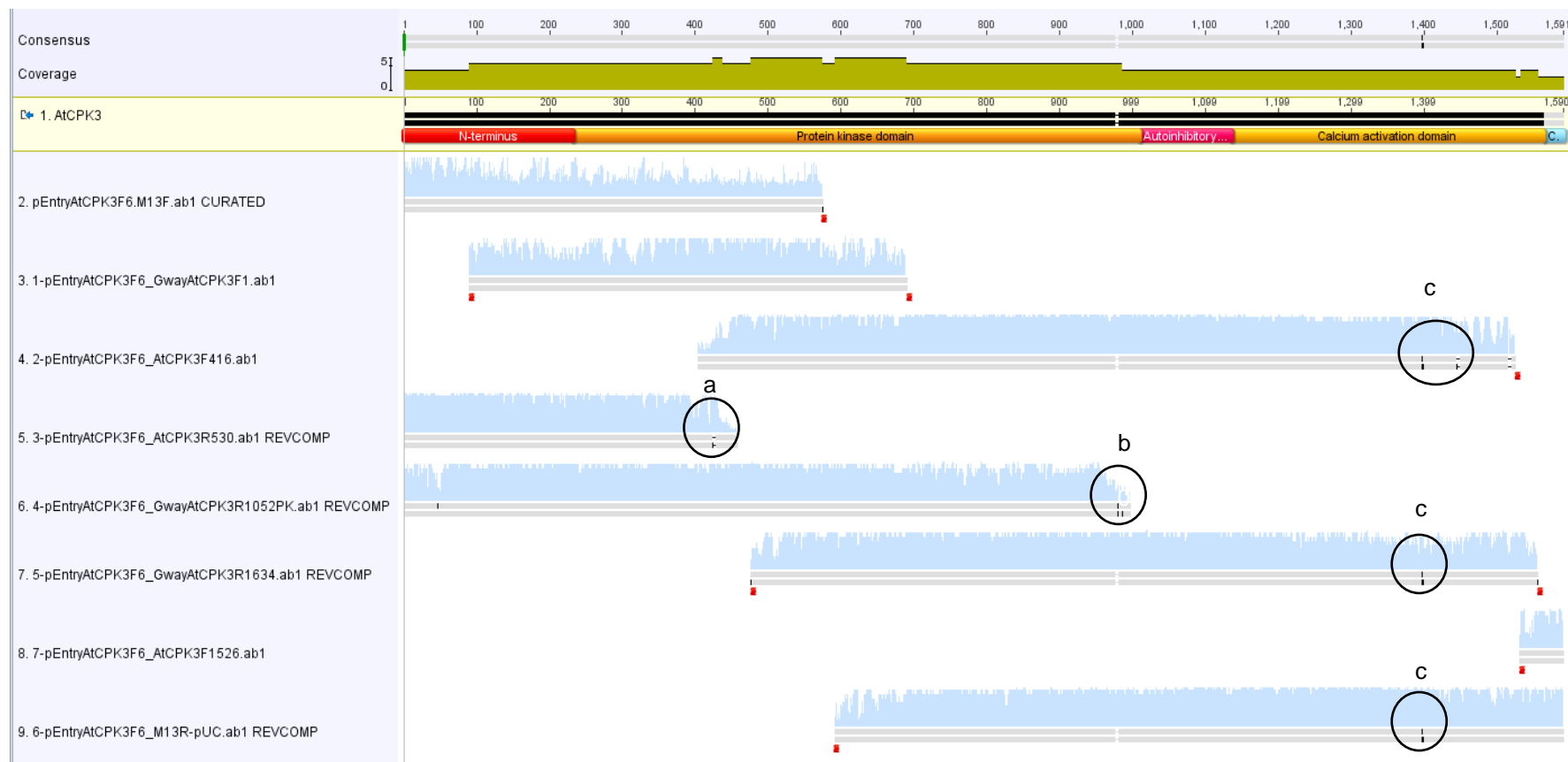
**Figure 4.32. Assembly of AtCPK3 entry clone number 6.** Sequenced clones were assembled to the published AtCPK3 sequence (1). Different primers were used to cover the length of AtCPK3: M13F forward primer (2), GatewayAtCPK3F1 forward primer (3), AtCPK3F416 forward primer (4), AtCPK3R530 reverse primer (5), AtCPK3R1052PK reverse primer (6), AtCPKR1634 reverse primer (7), AtCPKF1526 forward primer (8) and M13R reverse primer. Mismatches are encircled and highlighted in black (a, b and c). Mismatch a and b were disregarded as other sequences covering these positions did not show a mismatch. Mismatch at position 1398 (c) is shown in detail in Figure 4.33.
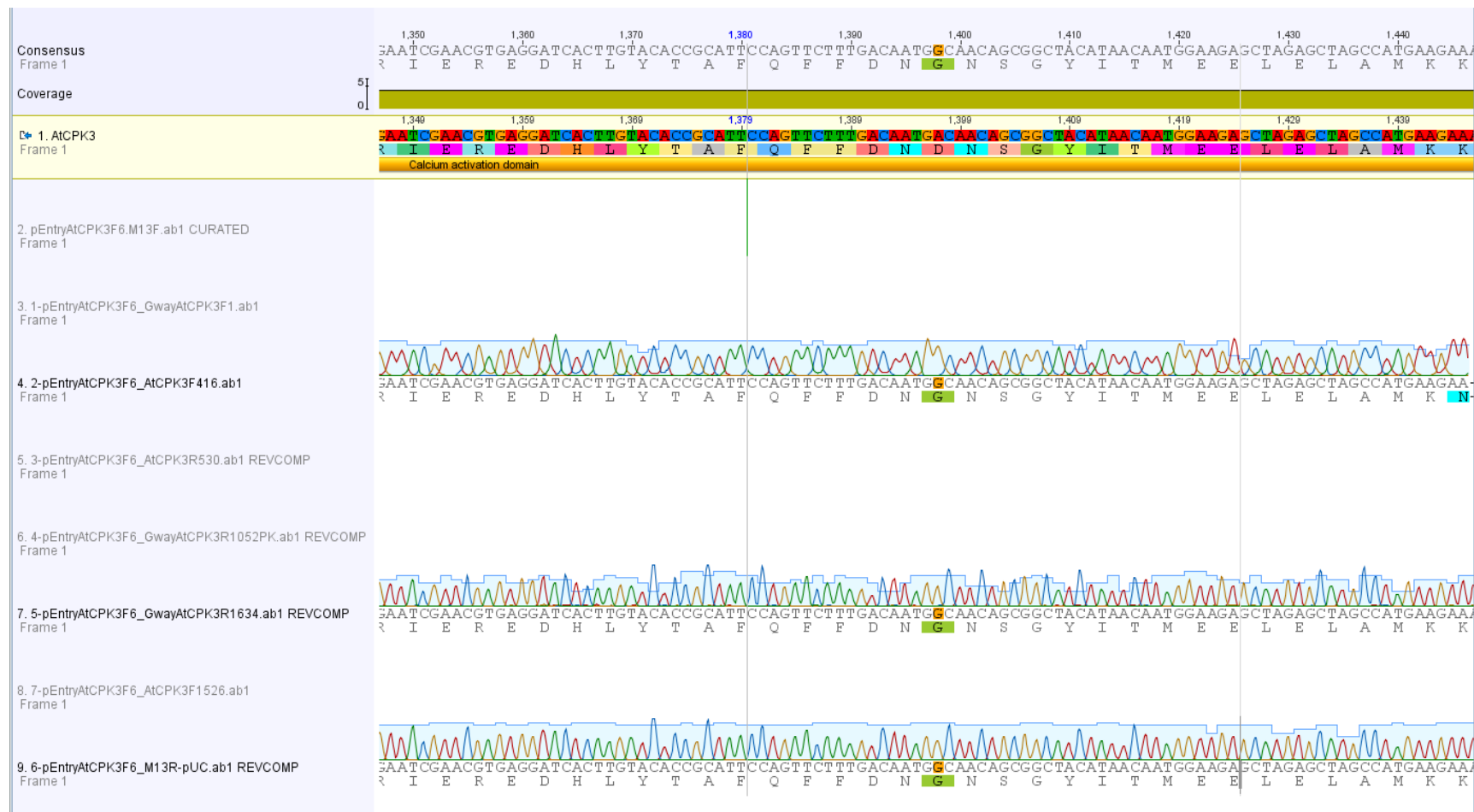
**Figure 4.33. Mismatches in the assembly of AtCPK3F6.** Sequenced clones were assembled to the published AtCPK3 sequence (1). Different primers used in sequencing were shown in Figure 4.34. Mismatch at position 1398 is confirmed by three sequencing results: from AtCPK3F416 forward primer (4), AtCPKR1634 reverse primer (7) and M13R reverse primer.

From the cloning experiment with 4:5 insert to vector molar ratio, further colony PCRs were carried out until a total of 95 colonies were screened. From this, seven positive colonies were identified: 54, 57, 59, 64, 66, 83, and 93 (as indicated by the red arrows in Figure 4.34).

Plasmids were isolated from the seven colonies and were sequenced. The resulting sequences were correctly identified as AtCPK3. However, all of the colonies had one to seven nucleotide differences from the published sequence (Appendix 27). Five of the colonies have changes in the resulting aa sequence while the other two did not. From these two, one colony (F66) was selected as the entry clone as it only had one nucleotide difference (position 477, from T to C), which resulted in no aa change (Figure 4.35).
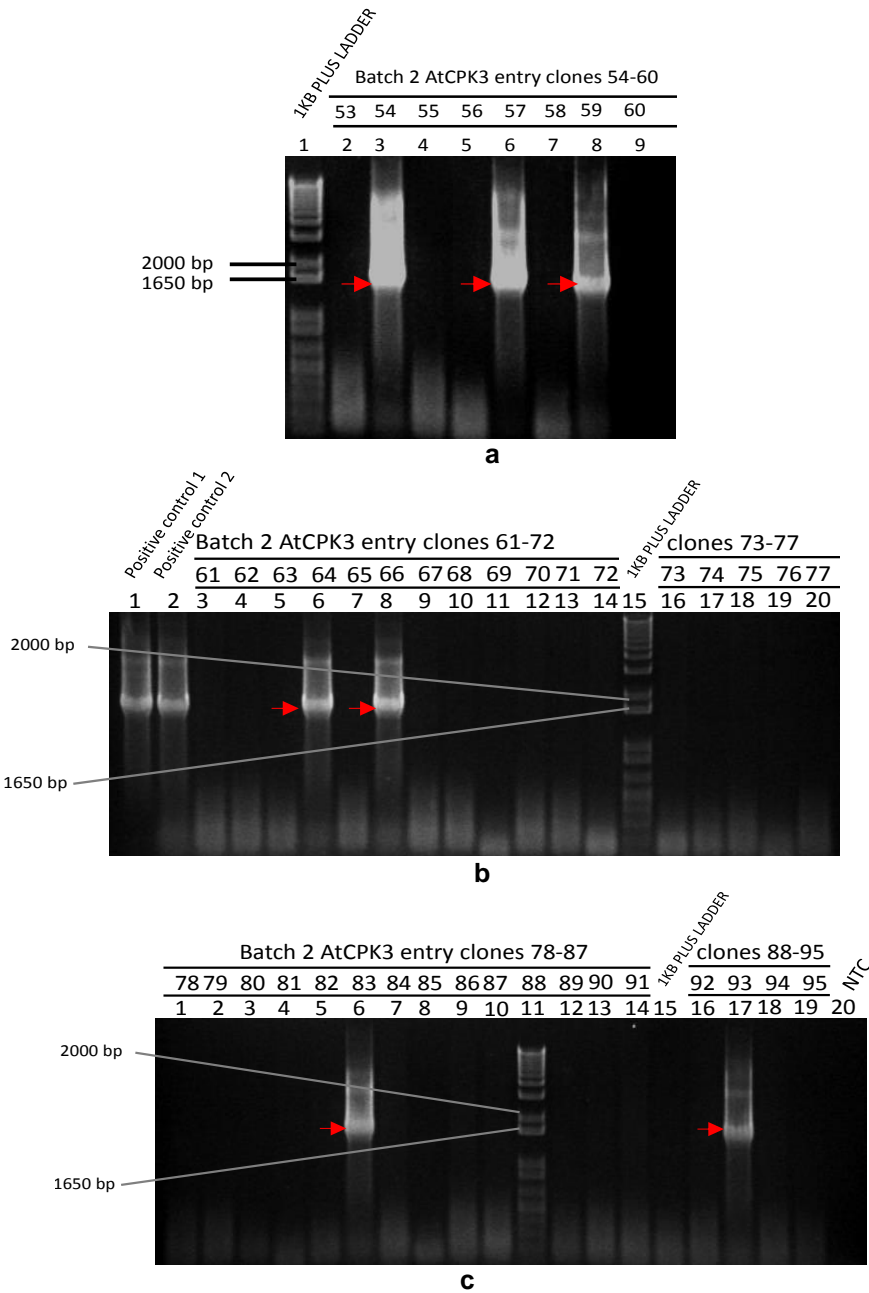


**Figure 4.34. Further colony PCR reactions up to 95 colonies (a to c).** Positive colonies indicated by the red arrow.
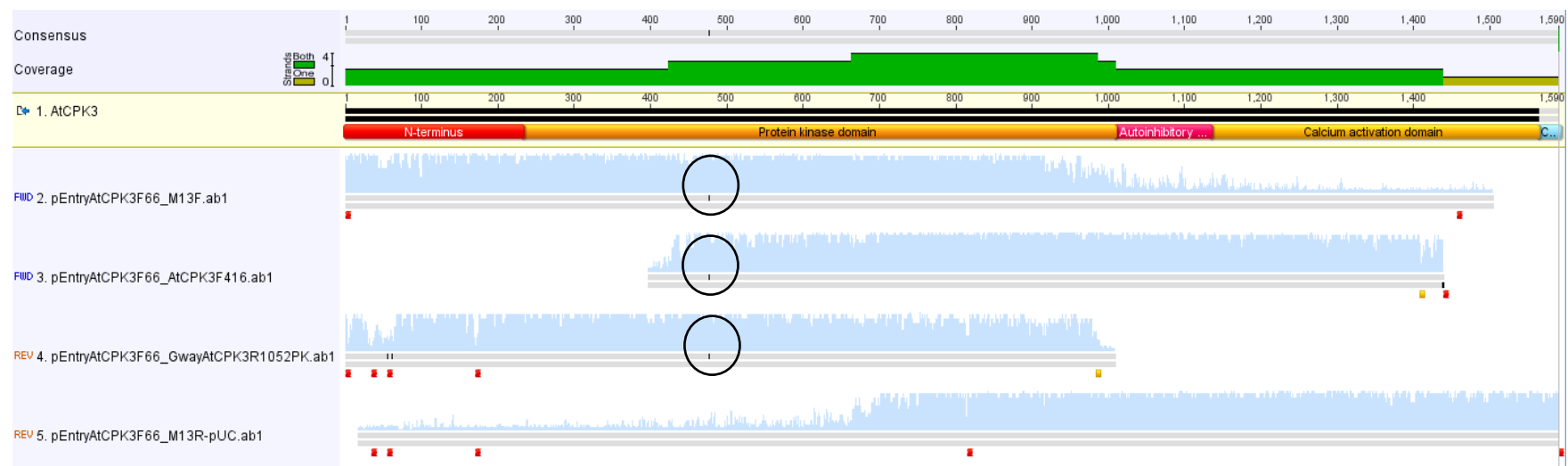
**Figure 4.35 Assembly of AtCPK3 entry clone 66.** Sequenced clones were assembled to the published AtCPK3 sequence (1). Different primers were used to cover the length of AtCPK3: M13F forward primer (2), AtCPK3F416 forward primer (3), and M13R reverse primer (4). Mismatches are encircled and highlighted in black. Mismatch at position 477 is encircled and highlighted in black.

The LR recombination reaction was performed to transfer the AtCPK3 coding sequence from the entry clone (pEntryAtCPK3F66) to the destination vector (pHEX2), creating the AtCPK3 overexpression clone. An LR reaction between the donor and destination vector with a 1:1 molar ratio resulted in $1.3 \times 10^4$ CFU/µg donor DNA. Five colonies were screened by colony PCR and all colonies contained an insert of the correct size (Figure 4.36). These constructs were named pHEX2AtCPK3 constructs 1 to 5.
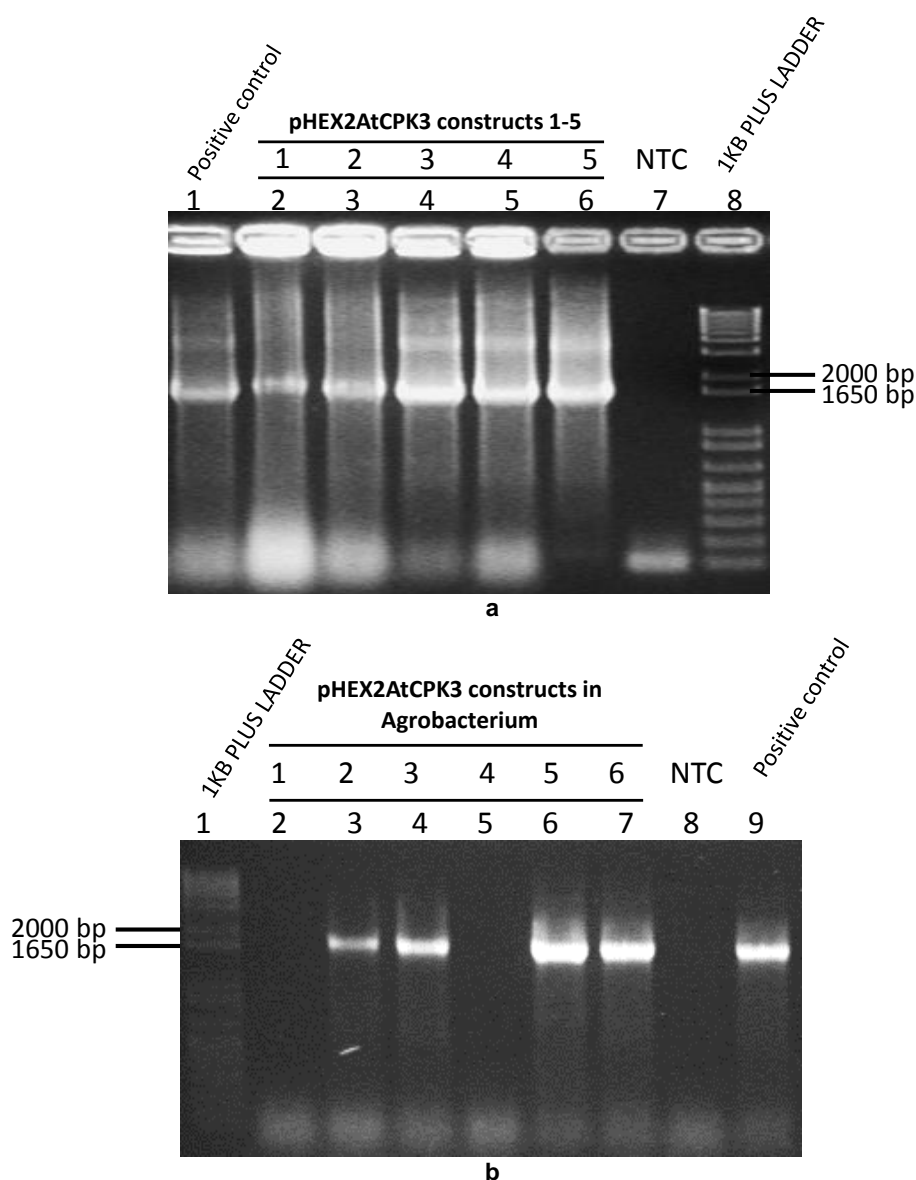


**Figure 4.36. Colony PCR of pHEX2AtCPK3 constructs in (a)** *E. coli* **and (b)** *A. tumefaciens* **after transformation. (a) pHEX2AtCPK3 in** *E. coli.* All pHEX2AtCPK3 clones showed the expected PCR product size. Lane 1: AtCPK3 entry clone as positive control; Lanes 2 to 6: pHEX2AtCPK3 transformants; Lane 7: NTC; Lane 8: 1 Kb plus DNA ladder (Invitrogen). **(b) pHEX2AtCPK3 in** *A. tumefaciens.* Four out of six pHEX2AtCPK3 clones showed the expected PCR product size. Lane 1: 1 Kb plus DNA ladder (Invitrogen); Lanes 2 to 7: pHEX2AtCPK3 transformants; Lane 8: NTC; Lane 9: AtCPK3 entry clone as positive control.

All five pHEX2-AtCPK3 constructs had the same sequence as the entry clone used. One construct was randomly selected (pHEX2-AtCPK3 construct 3, lane 3 in Figure 4.36) for plasmid isolation and transformed into Agrobacterium GV3101 as described in section 4.2.2.6b. This was done by electroporation and resulted in 3.75 x $10^6$ CFU/ug plasmid construct. Six colonies were randomly chosen and screened by colony PCR (Figure 4.36b). Four out of these six colonies appeared to be successful transformants. Two clones from were selected for plasmid extraction and re-cloning into *E. coli* for sequencing, to check that it has the correct sequence and that the orientation of the sequence was correct. Both constructs had the same AtCPK3 sequence as in the entry clone.

One of the clones was randomly selected and transformed into three pots of *A. thaliana* by floral dipping as described in section 4.2.2.6c. The cycle of transformation of *A. thaliana* and collection of first (T1) generation seeds and was done four times as the first three transformation attempts did not produce any successful T1 transformants. The fourth transformation resulted in nine transformants out of approximately 400 seeds sown. These lines were propagated until the third (T3) generation. One line (pHEX2AtCPK3.3) was further propagated until the fourth (T4) generation to ensure homozygosity of the transgene.

### 4.3.3.1b Verification of overexpressor and knockout lines for AtCPK3 expression

Due to the initial problem of low transformation efficiency with developing the AtCPK3 overexpressor plants, a search was done to look for potential external source of seeds of AtCPK3 overexpressor plants that had been developed elsewhere. One seed line was found, which was used in a previous study in Vienna (Mehlmer et al. 2010). This seed line, SAIL_120-H09, was obtained from NASC along with three T-DNA knockout lines of AtCPK3: SALK_106720C (*atcpk-1*), SALK_022862 (*atcpk-2*), and SALK_095134 (*atcpk-3*).

The overexpression of AtCPK3 in SAIL-120-H09 and pHEX2AtCPK3.3 and the knocking out of AtCPK3 in *atcpk-1*, *atcpk-2* and *atcpk-3* were verified by end-point RT-PCR. The end-point RT-PCR carried out as described in section 4.2.2 resulted in PCR products as shown in Figure 4.37. All reactions involved cDNA that was synthesised from 1 µg of RNA, therefore the amount of template used is assumed to be of equal amounts. Assuming equal amounts of template, RT-PCR products of pHEX2AtCPK3.3 plants appeared to be about three times brighter than the wild-

type, while the SAIL_120_H09 line was about five times more intense, which indicate higher expression of AtCPK3. All the T-DNA knockout lines did not show any PCR product. These results match the verification done by Mehlmer et al. (2010).
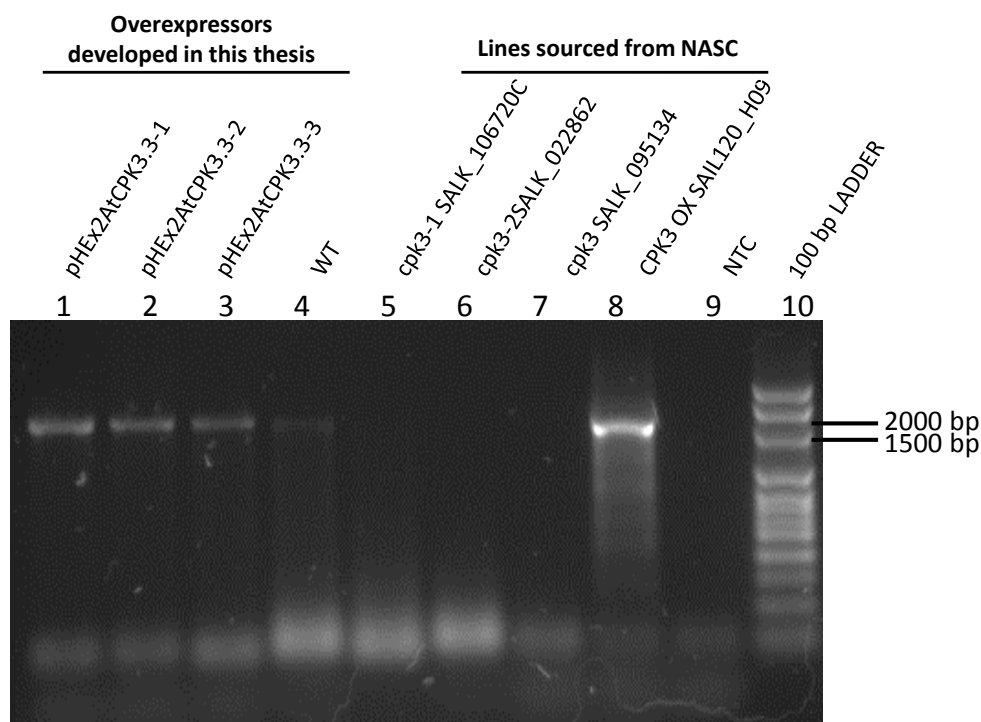


**Figure 4.37. RT-PCR results comparing AtCPK3 transcript accumulation in wildtype, overexpressor and knockout Arabidopsis plants.** Lanes 1 to 3: Overexpressors developed from pHEX2AtCPK3.3; Lane 4: Wt Arabidopsis; Lane 5 to 7: AtCPK3 knockouts SALK_106720C (*atcpk-1*), SALK_022862 (*atcpk-2*), and SALK_095134 (*atcpk-3*); Lane 8: AtCPK3 overexpressor SAIL-120-H09. Lane 9: NTC; Lane 10: 100 bp ladder (Solis BioDyne)

To further verify the knocking out or overexpression of AtCPK3 in the plants, an AtCPK3 antibody was designed as described in section 4.2.2.5. The AtCPK3 antibody was tested for its sensitivity and specificity to detect AtCPK3 using western blot analysis. However, no specific binding was achieved (Appendix 28).

### *4.3.3.1c Phenotype analysis of AtCPK3 WT, OX and KO lines*

*Phenotype measurements in Arabidopsis WT, OX and KO plants in response to drought*

Wild type Arabidopsis, two AtCPK3 OX lines (SAIL_120_H09 and pHEX2AtCPK3) and two AtCPK3 KO lines (*atcpk3-1* and *atcpk3-2)* were subjected to drought treatment for 14 days as described in section 4.2.2.8. Plant height (primary inflorescence), severity scores and dry weights of ten plants per treatment for each line were measured to compare their phenotypic responses to drought conditions.

Significant reductions in plant height at 7 and 14 d were observed in plants subjected to drought treatment compared to those under control conditions (Figure 4.38). This was observed in all plant lines as expected because drought was known to reduce plant growth rate and plant height. There was strong statistical evidence to support the reduction in plant height among drought treated plants compared to control, at both time points for all the plant lines (Tukey's P value ranges from 0.004 to 0.000).

When comparing drought treated plants at 7 and 14 d, SAIL_120_H09 plants appeared to be marginally higher than WT, whereas *atcpk3-1* and *atcpk3-2* plants appeared to be marginally shorter than wildtype. pHEX2AtCPK3 plants appeared to be slightly higher than *atcpk3-1* and *atcpk3-2*, and slightly higher than WT. There was good statistical evidence to support the difference in height between SAIL_120_H09 and WT at 14 d (Fisher's LSD P= 0.023) and but very weak statistical support for the difference in height between *atcpk3-1* or *atcpk3-2* and WT (Fisher's LSD P=0.102 and P= 0.106).

Severity scores at 7 d among SAIL_120_H09 and pHEX2AtCPK3 plants were markedly less than WT, while *atcpk3-1* and *atcpk3-2* plants did not differ from WT (Figure 4.39). There was good statistical evidence to support the difference of the overexpressing lines (Tukey's P =0.002 and 0.022, Fisher's LSD P= 0.000 and 0.002). At 14 d only SAIL_120_H09 showed significantly lower severity scores than WT (Tukey's P= 0.014, Fisher's LSD= 0.002).

Dry weights of drought treatment plants were significantly lower than the control plants, similar to plant height as expected (all Tukey's P= 0.000) (Figure 4.40). Only SAIL_120_H09 showed significantly higher dry weights than WT (Tukey's P = 0.010), although pHEX2AtCPK3 also showed slightly higher dry weights than WT. The KO plants *atcpk3-1* and *atcpk3-2* showed slightly lower dry weights than WT, but did not have statistical support.
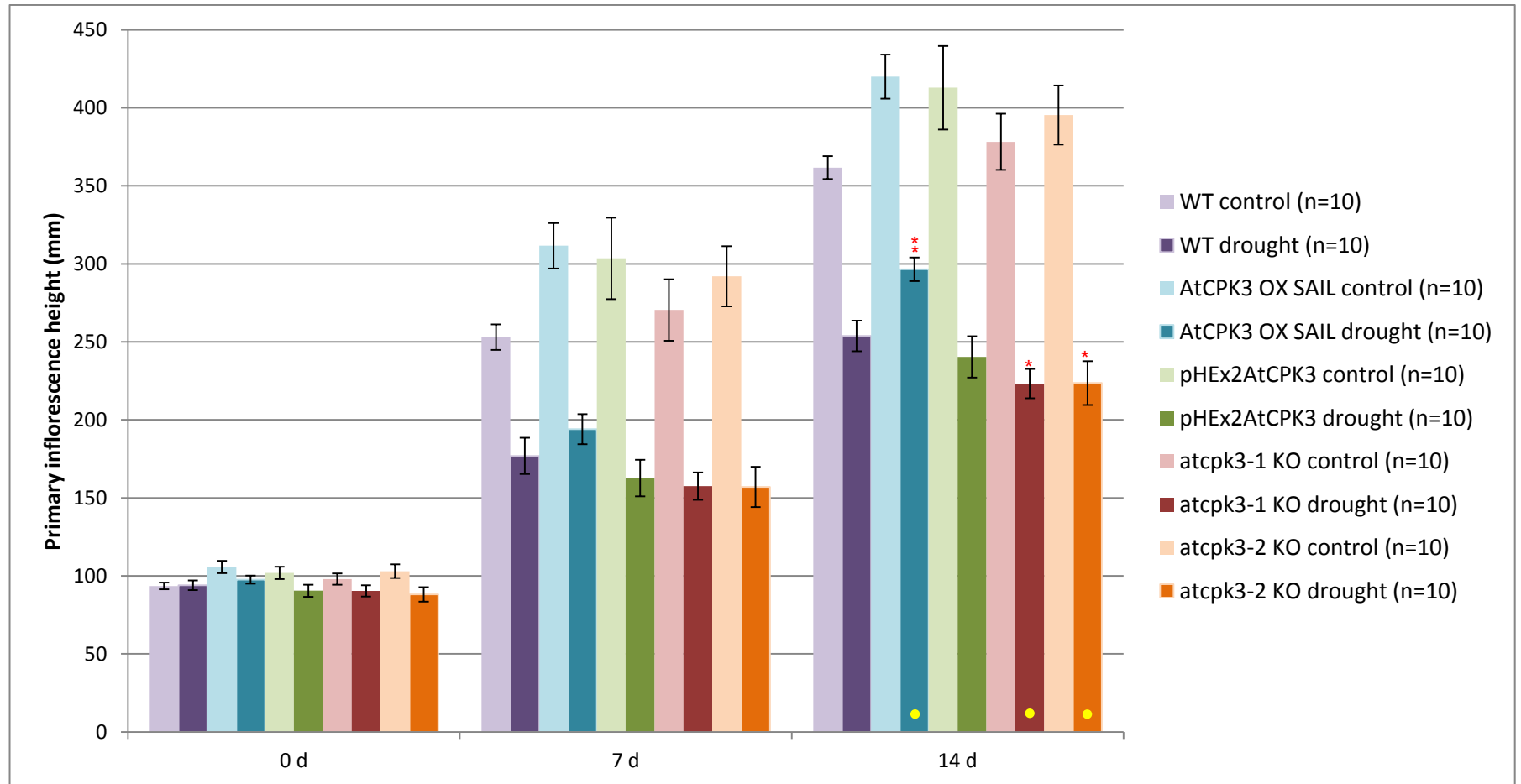
**Figure 4.38. Mean primary inflorescence height of Arabidopsis WT, OX and KO plants in response to drought.** Colour of bars match the Arabidopsis lines: purple, WT Arabidopsis; blue, AtCPK3 OX SAIL-120-H09; green, pHEX2AtCPK3.3; red, AtCPK3 KO SALK_106720C (*atcpk3-1*); and orange, AtCPK3 KO SALK_022862 (*atcpk3-2*). Control treatment has lighter shade while drought treatment has darker shade. Measurements done in millimetres. Line bars indicate SE of the mean. Yellow dots indicate a marked difference between control and treatment. Statistical support is indicated as: strong (***, P≤ 0.01), good (**, 0.01<P≤ 0.05) or weak (*, 0.05<P<~0.10).
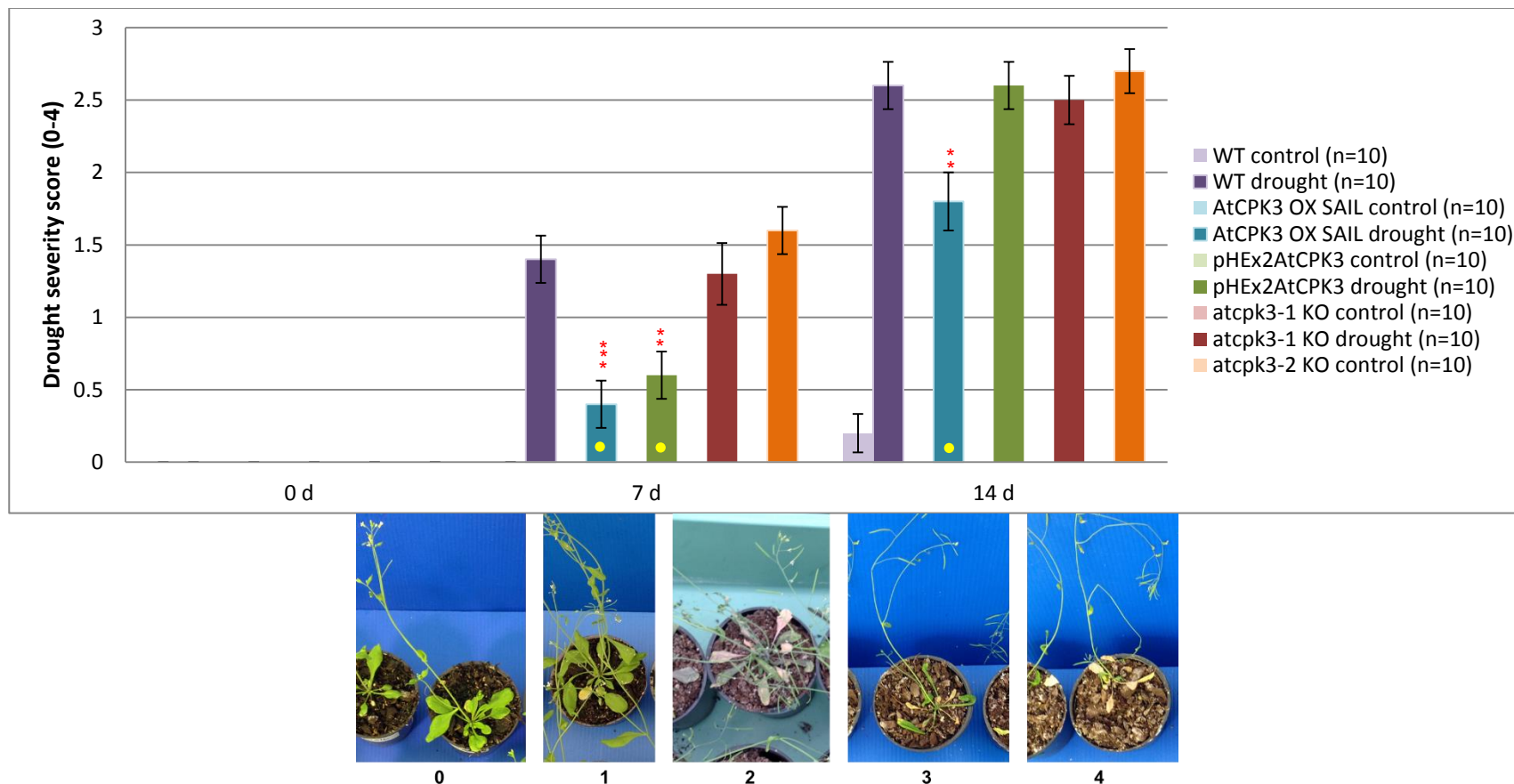
**Figure 4.39. Mean severity scores of Arabidopsis WT, OX and KO plants in response to drought.** Colour of bars match the Arabidopsis lines: purple, WT Arabidopsis; blue, AtCPK3 OX SAIL-120-H09; green, pHEX2AtCPK3.3; red, AtCPK3 KO SALK_106720C (*atcpk3-1*); and orange, AtCPK3 KO SALK_022862 (*atcpk3-2*). Control treatment has lighter shade while drought treatment has lighter shade. Severity scores dome in a scale of 0 to 4: 0, no symptoms; 1, a few leaves showing symptoms; 2, most of leaves showing symptoms; 3, all of leaves showing symptoms; 4, dead or dying. Line bars indicate SE of the mean. All plants at 0 dpi and all control treatment plants had a score of 0 except for one WT at 14 dpi. Yellow dots indicate a marked difference between control and treatment. Statistical support is indicated as: strong (***, $P \leq 0.01$), good (**, $0.01 < P \leq 0.05$) or weak (*, $0.05 < P < \sim 0.10$).
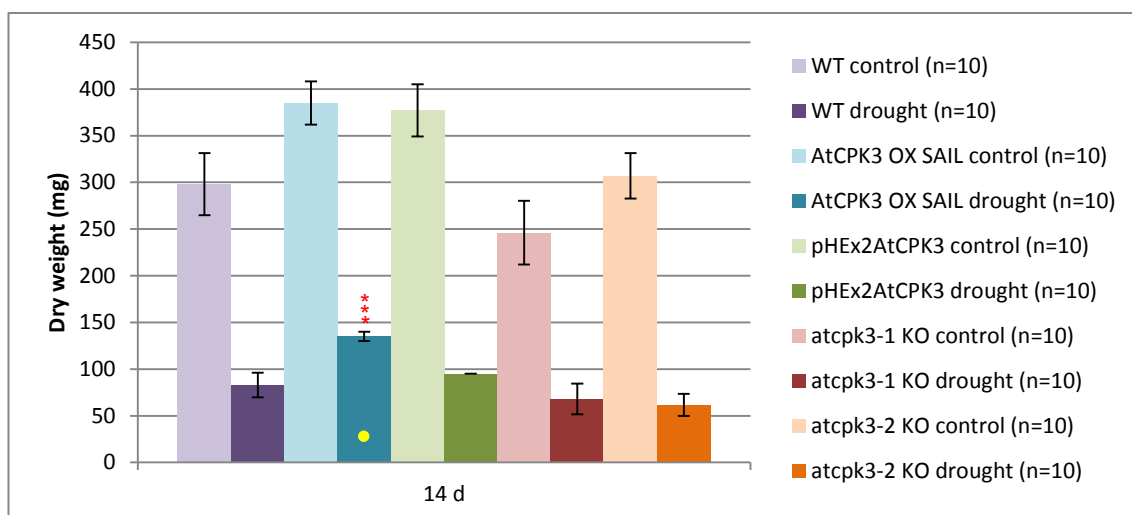
**Figure 4.40. Mean dry weights of Arabidopsis WT, OX and KO plants in response to drought.** Colour of bars match the Arabidopsis lines: purple, WT Arabidopsis; blue, AtCPK3 OX SAIL-120-H09; green, pHEX2AtCPK3.3; red, AtCPK3 KO SALK_106720C (*atcpk3-1*); and orange, AtCPK3 KO SALK_022862 (*atcpk3-2*). Control treatment has lighter shade while drought treatment has darker shade. Line bars indicate SE of the mean. Yellow dots indicate a marked difference between control and treatment. Statistical support is indicated as: strong (***, P≤ 0.01), good (**, 0.01<P≤ 0.05) or weak (*, 0.05<P<~0.10).

In summary, in response to drought the AtCPK3 OX lines appeared to have marginally greater plant heights and dry weights and lower severity scores than WT, although significant difference was mostly observed in the SAIL_120_H09 plants and not in the pHEX2AtCPK3 plants. The AtCPK3 KO lines *atcpk3-1* and *atcpk3-2* knockouts had marginally lower plant heights and dry weights and slightly higher symptom scores than WT, although significant difference was only observed in plant heights.

*Phenotype measurements in Arabidopsis WT, OX and KO in response to B. cinerea*

Only symptom scores were gathered in measuring the response of Arabidopsis plants to *B. cinerea* (Figure 4.41). No marked difference was observed between wild type and all the transgenic plants, except for SAIL_120_H09 overexpressors, which had marginally lower mean scores at 28 dpi. This was supported by weak statistical evidence (Fisher's LSD P= 0.086).
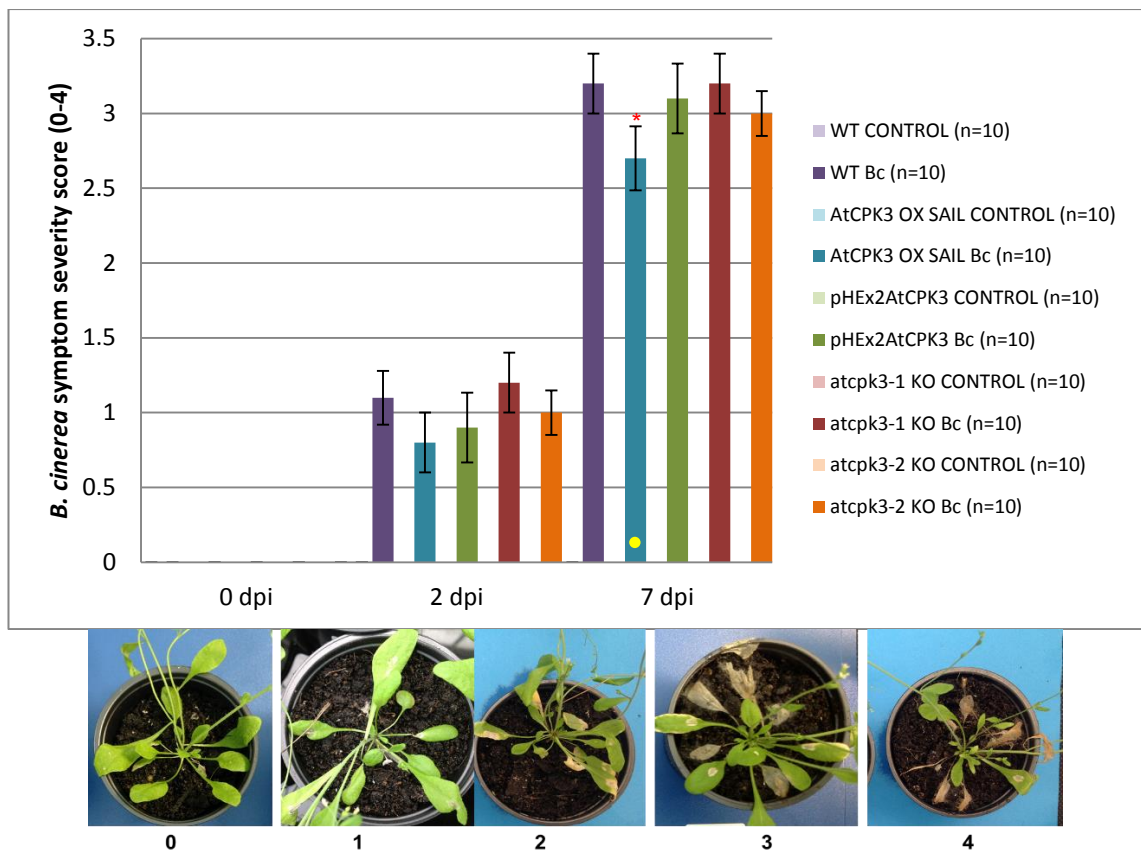


**Figure 4.41. Mean severity scores of Arabidopsis WT, OX and KO plants in response to *B. cinerea*.** Colour of bars match the Arabidopsis lines: purple, WT Arabidopsis; blue, AtCPK3 OX SAIL-120-H09; green, pHEX2AtCPK3.3; red, AtCPK3 KO SALK_106720C (atcpk-1); and orange, AtCPK3 KO SALK_022862 (atcpk-2). Control treatment has lighter shade while *B. cinerea* infection has darker shade. Severity scores dome in a scale of 0 to 4: 0, no symptoms; 1, a few leaves showing symptoms; 2, many leaves showing symptoms; 3, almost all of leaves showing symptoms, some dying; 4, most of the leaves dead or dying. Line bars indicate SE of the mean. All plants at 0 dpi and all control treatment plants had a score of 0. Yellow dots indicate a marked difference between control and treatment. Statistical support is indicated as: strong (***, P≤ 0.01), good (**, 0.01<P≤ 0.05) or weak (*, 0.05<P<~0.10).

*Phenotype measurements in Arabidopsis WT, OX and KO in response to TYMV*

Wild type Arabidopsis, two AtCPK3 OX lines (SAIL_120_H09 and pHEX2AtCPK3) and two AtCPK3 KO lines (*atcpk3-1* and *atcpk3-2)* were compared in terms of their phenotypic responses to TYMV. As described in section 4.2.2.8, inoculation with TYMV and with inoculation buffer (mock inoculation) were carried out to ten plants per treatment for each plant line. TYMV was used among the other viruses used in section 4.2.2.2a due to the availability of good inoculum material. However, with the first several attempts, no successful TYMV infection was achieved despite doing the same methodology and conditions as described in 4.2.2.2a. Plants inoculated with TYMV did not show any symptoms and no virus titre was detected using ELISA and RT-PCR. This may be due to the reduction in the infectivity of the virus inoculum material during storage. In the last attempt, only few plants were successfully infected with virus: seven wildtype plants, two SAIL-120_H09 plants, one pHEX2AtCPK3 plant, five atcpk3-1 plants and six atcpk3-2 plants. Only mild symptoms were observed generally. Plant height (primary inflorescence), severity scores and dry weights of these plants were then measured to compare their phenotypic responses to drought conditions.

Marked reduction in plant height at 14, 21 and 28 dpi were mostly observed in plants infected with TYMV compared to the mock inoculated plants, most supported by Fisher's LSD (P ≤0.050) (Figure 4.42). In comparing the different plant lines in response to TYMV-infection at 14 dpi, only SAIL_120_H09 plants (mean= 89.00 mm) appeared to be marginally taller than WT (68.71 mm), whereas only *atcpk3-2* plants (42.5 mm) appeared to be marginally shorter than WT. At this timepoint, the difference between SAIL_120_H09 plants and WT did not have statistical support while the lower plant height among *atcpk3-2* plants compared to WT had weak statistical support (Fisher's LSD P= 0.068). At 21 and 28 dpi, both SAIL_120_H09 (181.5 mm and 309.5 mm) and pHEX2AtCPK3 (183.00 mm and 300.00) plants were marginally taller than WT (156.29 mm and 271.00 mm) while the *atcpk3-2* plants (137.67 mm and 256.50 mm) were marginally shorter than WT. Only the marked difference between SAIL_120_H09 and WT at 28 dpi had good statistical support (Fisher's LSD P=0.009), while the others had no statistical support. However, these measurements were only taken from a very small number of plants that had successful TYMV infection.

With regards virus symptom severity scores, symptoms only started becoming observable in SAIL_120_H09 and pHEX2AtCPK3 at 21 and 28 dpi, and were less severe than

WT (Figure 4.43). *atcpk3-1* and *atcpk3-2* plants did not show a marked difference in severity from WT, although *atcpk3-2* appeared to be marginally more severe. Good statistical evidence to support the marked difference from WT was determined among SAIL_120_H09 and pHEX2AtCPK3 lines at 14 dpi (Tukey's P= 0.030 and 0.142; Fisher's LSD P= 0.004 and 0.024).

With regards number of siliques at 21 and 28 dpi, SAIL_120_H09 (mean= 102.00 and 172.00 siliques) and pHEX2AtCPK3 (125.00 and 210.00 siliques) had more siliques than WT (72.71 siliques) while *atcpk3-1* (52.00 and 113.30 siliques) and *atcpk3-2* (52.67 and 113.40 siliques) plants had fewer siliques. There was weak statistical evidence to support the marked difference of all plant lines from the WT at 21 dpi (Fisher's LSD P= 0.096, 0.027, 0.107, and 0.101, respectively) and among *atcpk3-1* and *atcpk3-2* lines at 28 dpi (Fisher's LSD P= 0.090 and 0.074).

For dry weights at 28 dpi, SAIL_120_H09 and pHEX2AtCPK3 were marginally heavier (180.0 and 210.0 mg) than WT (125.7 mg) while *atcpk3-1* and *atcpk3-2* plants did not show a marked difference (122.0 and 111.7 mg) from WT. However, there was no statistical evidence to support the marked differences in dry weights.

In summary, in response to TYMV infection, AtCPK3 OX lines appeared to be marginally taller and have more siliques, greater dry weights, and lower severity scores than WT. However, the statistical evidences for significant difference were not strong as there were only two plants for SAIL_120_H09 and one for pHEX2AtCPK3 that had successful TYMV infection. The AtCPK3 KO lines did not show marked differences with WT except for plant height (*atcpk3-2* only) and number of siliques.
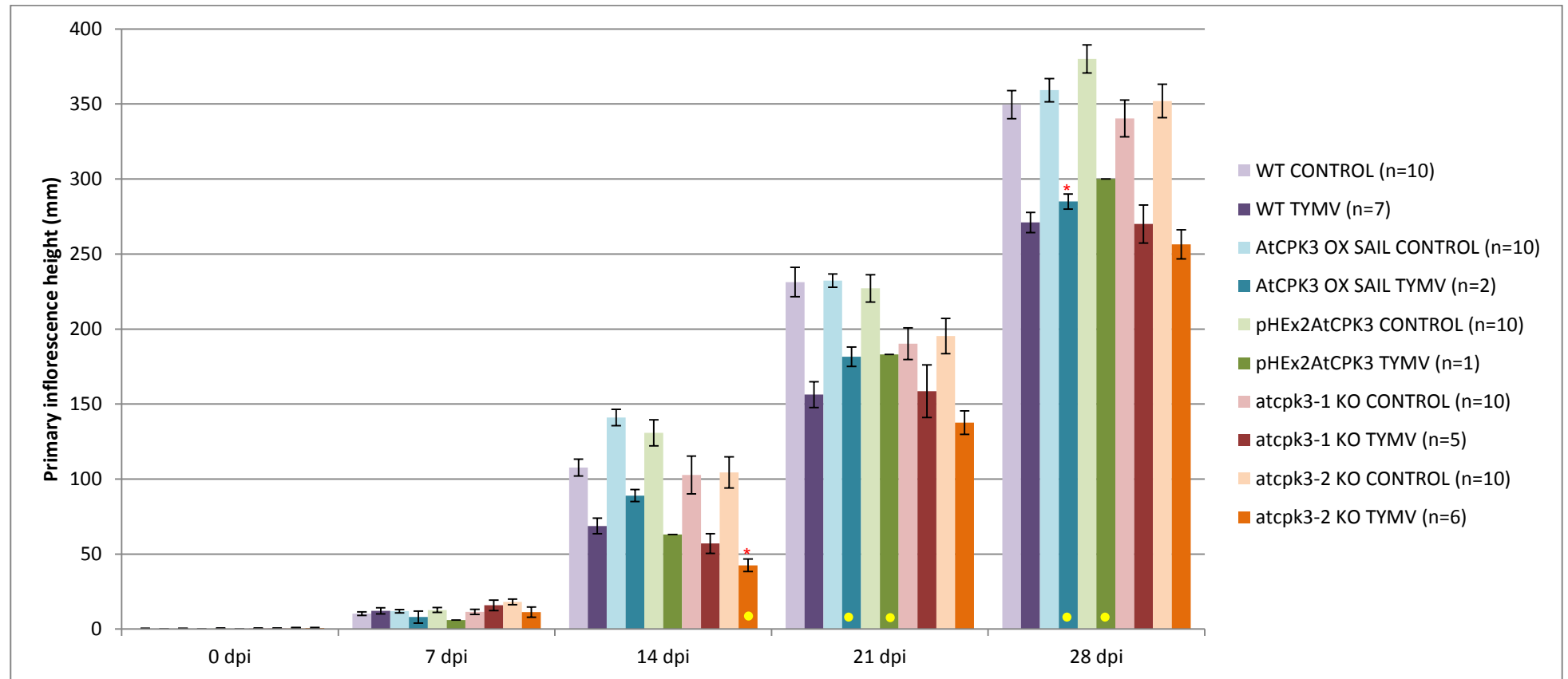
**Figure 4.42. Mean primary inflorescence height of Arabidopsis WT, OX and KO plants upon TYMV infection.** Colour of bars match the Arabidopsis lines: purple, WT Arabidopsis; blue, AtCPK3 OX SAIL-120-H09; green, pHEX2AtCPK3.3; red, AtCPK3 KO SALK_106720C (*atcpk-1*); and orange, AtCPK3 KO SALK_022862 (*atcpk-2*). Control treatment (mock inoculated) has lighter shade while TYMV infection has darker shade. Measurements done in millimetres. Line bars indicate SE of the mean. Yellow dots indicate a marked difference between the transgenic line and WT. Statistical support is indicated as: strong (***, $P \le 0.01$), good (**, $0.01 < P \le 0.05$) or weak (*, $0.05 < P < \sim 0.10$).
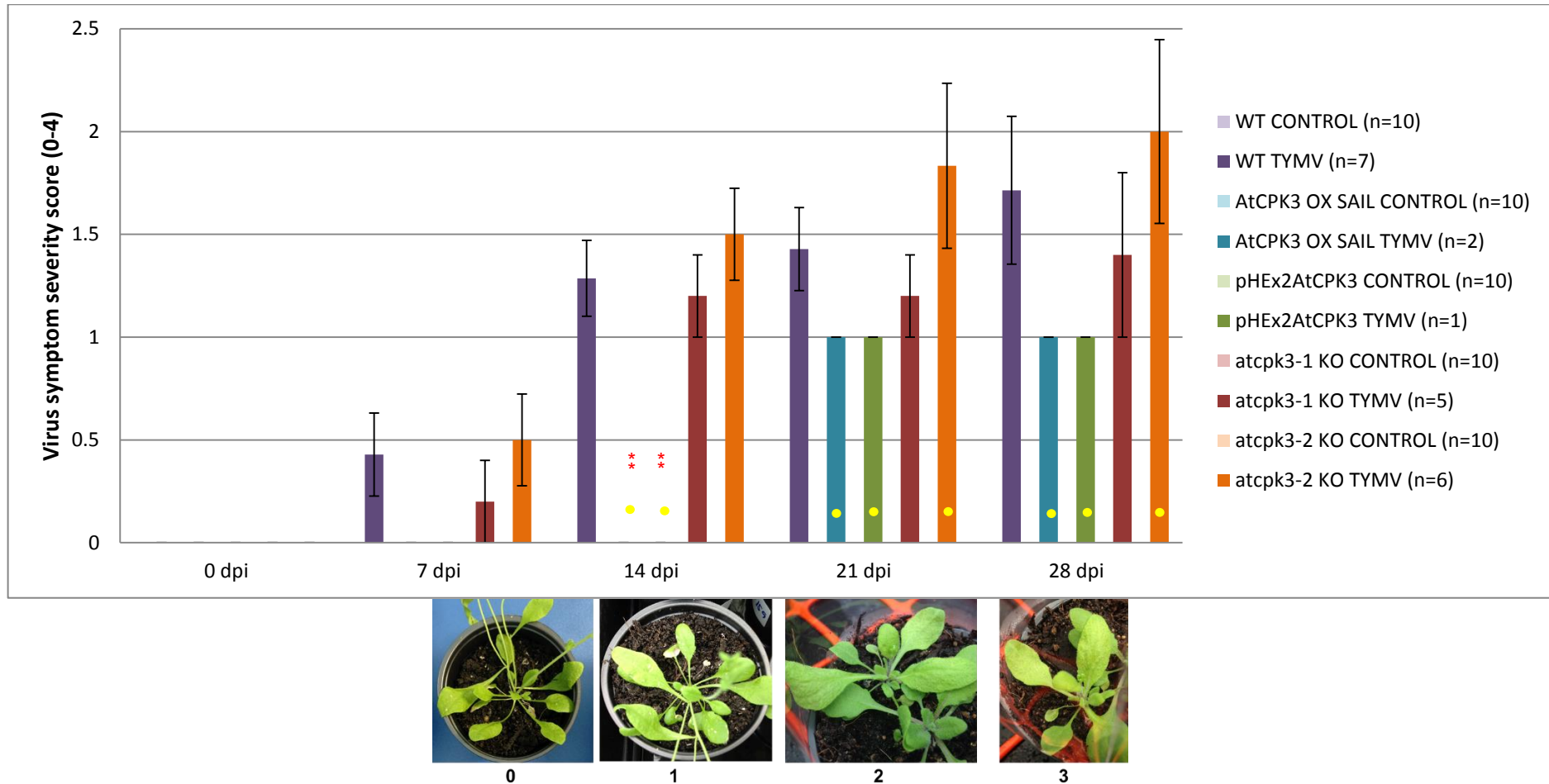
**Figure 4.43. Mean virus symptom scores of Arabidopsis WT, OX and KO plants upon TYMV infection.** Colour of bars match the Arabidopsis lines: purple, WT Arabidopsis; blue, AtCPK3 OX SAIL-120-H09; green, pHEX2AtCPK3.3; red, AtCPK3 KO SALK_106720C (atcpk-1); and orange, AtCPK3 KO SALK_022862 (atcpk-2). Control treatment (mock inoculated) has lighter shade while TYMV infection has darker shade. Severity scores done in a scale of 0 to 4: 0, no symptoms; 1, a few leaves showing symptoms; 2, most of leaves showing symptoms; 3, all of leaves showing symptoms; 4, dead or dying. Line bars indicate SE of the mean. Yellow dots indicate a marked difference between the transgenic line and WT. Statistical support is indicated as: strong (***, $P \le 0.01$), good (**, $0.01 < P \le 0.05$) or weak (*, $0.05 < P < \sim0.10$).

**Figure 4.44. Mean number of siliques of Arabidopsis WT, OX and KO plants upon TYMV infection.** Colour of bars match the Arabidopsis lines: purple, WT Arabidopsis; blue, AtCPK3 OX SAIL-120-H09; green, pHEX2AtCPK3.3; red, AtCPK3 KO SALK_106720C (*atcpk-1*); and orange, AtCPK3 KO SALK_022862 (*atcpk-2*). Control treatment (mock inoculated) has a lighter shade while TYMV infection has a darker shade. Line bars indicate SE of the mean. Yellow dots indicate a marked difference between the transgenic line and WT. Statistical support is indicated as: strong (***, P≤ 0.01), good (**, 0.01<P≤ 0.05) or weak (*,0.05<P<~0.10).
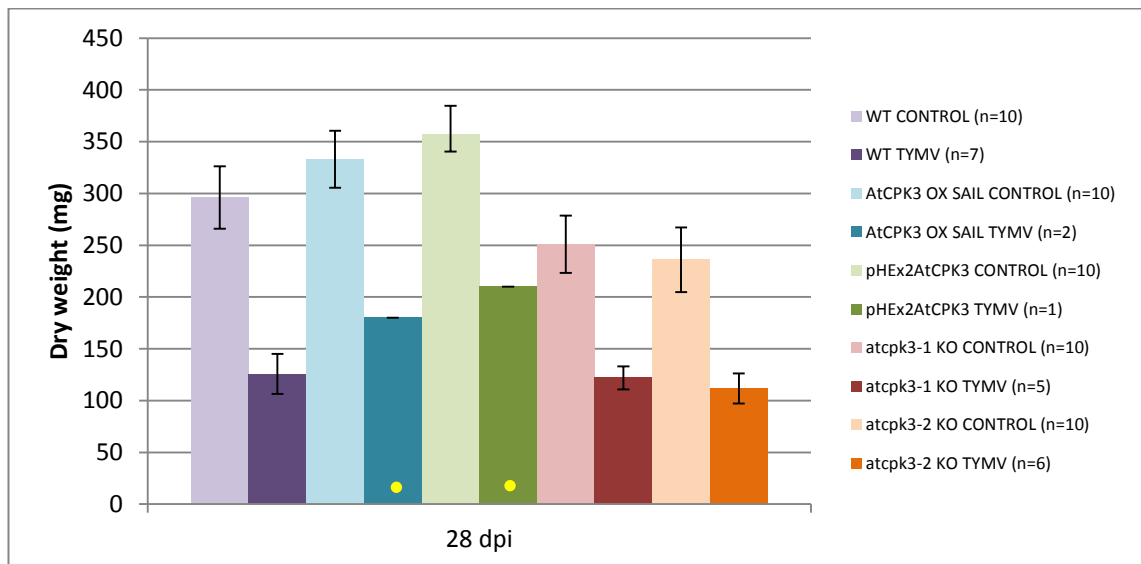


**Figure 4.45 Dry weight of Arabidopsis WT, OX and KO plants upon TYMV infection after 28 dpi.** Colour of bars match the Arabidopsis lines: purple, WT Arabidopsis; blue, AtCPK3 OX SAIL-120-H09; green, pHEX2AtCPK3.3; red, AtCPK3 KO SALK_106720C (*atcpk-1*); and orange, AtCPK3 KO SALK_022862 (*atcpk-2*). Control treatment (mock inoculated) has lighter shade while TYMV infection has darker shade. Line bars indicate SE of the mean. Yellow dots indicate a marked difference between the transgenic line and WT. Statistical support is indicated as: strong (***, P≤ 0.01), good (**, 0.01<P≤ 0.05) or weak (*, 0.05<P<~0.10).

### 4.3.3.2 A. chinensis overexpressors and knockout lines

#### 4.3.3.2a Development and verification of overexpressor and knockout lines for AcCPK16 expression

The AcCPK16 overexpressor and knockout lines were developed with the assistance of the Breeding and Genomics team at PFR, as described in section 4.2.2.7. Assuming equal amount of template, AcCPK16OX line E05 showed higher transcript accumulation of AcCPK16 than WT whereas all AcCPK16 KO lines E05, E10 and E11 showed absence of AcCPK16. Accumulation of AcCPK16 in OX line E06 showed marginally higher accumulation of AcCPK16 than WT.
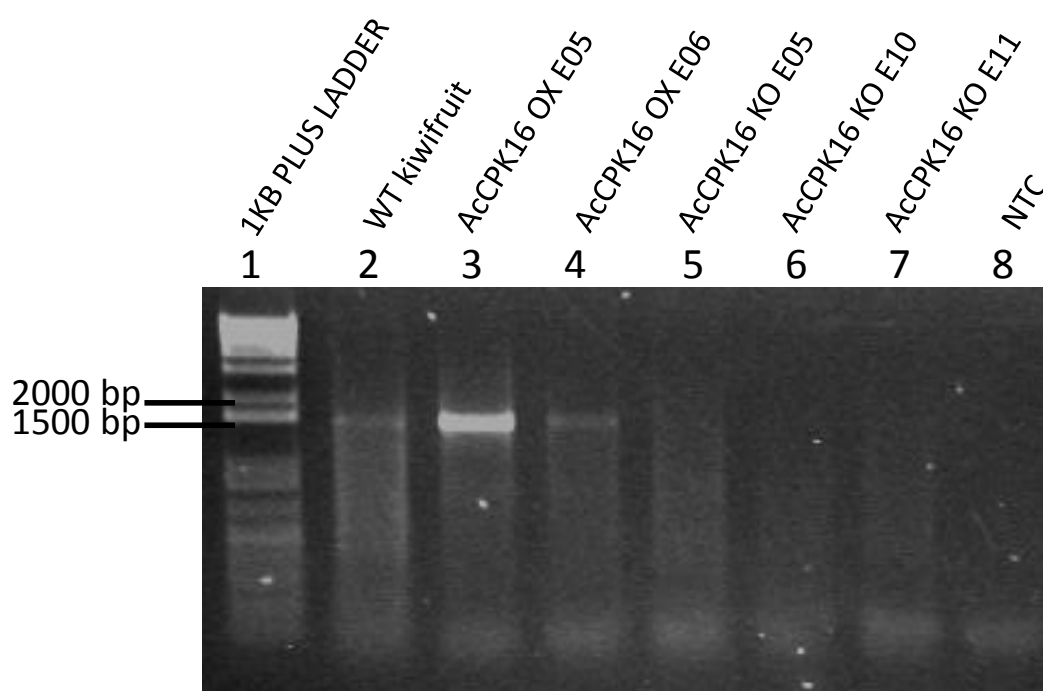


**Figure 4.47 Verification of AcCPK16 WT, OX and KO plants using RT-PCR.** Lane 1: 1 Kb plus DNA ladder (Invitrogen); Lane 2: WT kiwifruit; Lane 3: AcCPK16 OX E05; Lane 4: AcCPK16 OX E06; Lane 5: AcCPK16 KO E05; Lane 6: AcCPK16 KO E10; Lane 7: AcCPK16 KO E11; Lane 8: NTC.

***4.3.3.2b Phenotype Analysis of AcCPK16 WT, OX and KO lines***

*Phenotype measurements in kiwifruit WT, OX and KO in response to drought*

Wild type kiwifruit, three AcCPK16 OX lines (AcCPK16 OX 05, 06 and 07) and three AcCPK16 KO lines (AcCPK16 OX 05, 06 and 07) were subjected to drought treatment for 14 days as described in section 4.2.2.9. Plant height, drought severity scores and dry weights of available plants per treatment for each line (refer to Table 4.17 in section 4.2.2.9) were measured to compare their phenotypic responses to drought conditions.

Similar to Arabidopsis, marked reduction in plant height at 7 and 14 dpi were observed in kiwifruit plants subjected to drought treatment compared to those under control conditions (Figure 4.48). This was observed in all plant lines as expected, and with strong statistical support at 14 dpi (Tukey's P ranges from 0.005 to 0.000). Moreover, similar heights were observed in drought plants at 7 d and 14 d as plants appeared to stop growing within 7 to 14 d under drought conditions. At 7 and 14 d, all the three AcCPK16 OX lines appeared to be marginally higher than WT (60.00, 58.00, and 51.00 mm against 43.50 mm at 7d and 61.00, 60.00 and 51.00 mm against 43.50 mm at 14 d), whereas the AcCPK16 KO lines did not show significant difference from WT (41.00, 44.50 and 47.00 mm against 43.50 mm at 7 d and 43.50, 45.50 and 48.00 mm against 43.5 mm at 14 d). The marked difference in height compared to WT has good statistical support at 7 dpi for all three lines (Fisher's LSD P= 0.002, 0.004 and 0.101 respectively) and at 14 dpi for AcCPK16 OX E05 and E06 (Fisher's LSD P= 0.003 and 0.004 respectively).

Severity scores for drought at 7 d were lower than WT (mean score= 2.50) among all AcCPK16 OX lines (all 0.5) and two KO lines E05 (0.5) and E10 (1.0) (Figure 4.49). All these five lines showed good statistical support (all Fisher's LSD P=0.013, except E10 with 0.044). At 14 d, the severity scores of AcCPK16 OX lines (2.5, 2.5 and 3.0) were significantly lower than WT (4.0) (Fisher's LSD P= 0.009, 0.009 and 0.050 respectively), whereas AcCPK16 KO lines (3.5, 4.0 and 4.0) did not show significant difference.

Dry weights of drought treated plants at 14 dpi were all significantly lower than the control plants, particularly for WT and the AcCPK16 KO lines (Figure 4.50) (all Fisher's LSD P=0.000). Interestingly, there was no significant difference or very little difference between the control and drought treatment groups among AcCPK16 OX E05 and E06 lines (Tukeys P= 0.957 and 0.778,

Fisher's LSD P=0.165 and 0.071). AcCPK16 OX E05 and E06 also showed marginally higher dry weights (945.0 and 925.0 mg) than WT (780.0 mg), while AcCPK16 KO E10 plants showed marginally lower dry weights (705.0 mg) than WT. The marked difference among the OX lines were supported by good statistical evidence (Fisher's LSD P=0.017 and 0.033), while the KO lines were not significantly different (Fisher's LSD P=0.110).

In summary, in response to drought the AcCPK16 OX lines appeared to have marginally greater plant heights and dry weights and lower severity scores than WT. The AcCPK16 KO lines did not show marked differences with WT in terms of plant height, severity scores and dry weight, except for KO E05 and 10 at 7d with lower severity scores and KO E10 at 14 d with a very marginal difference in dry weight.
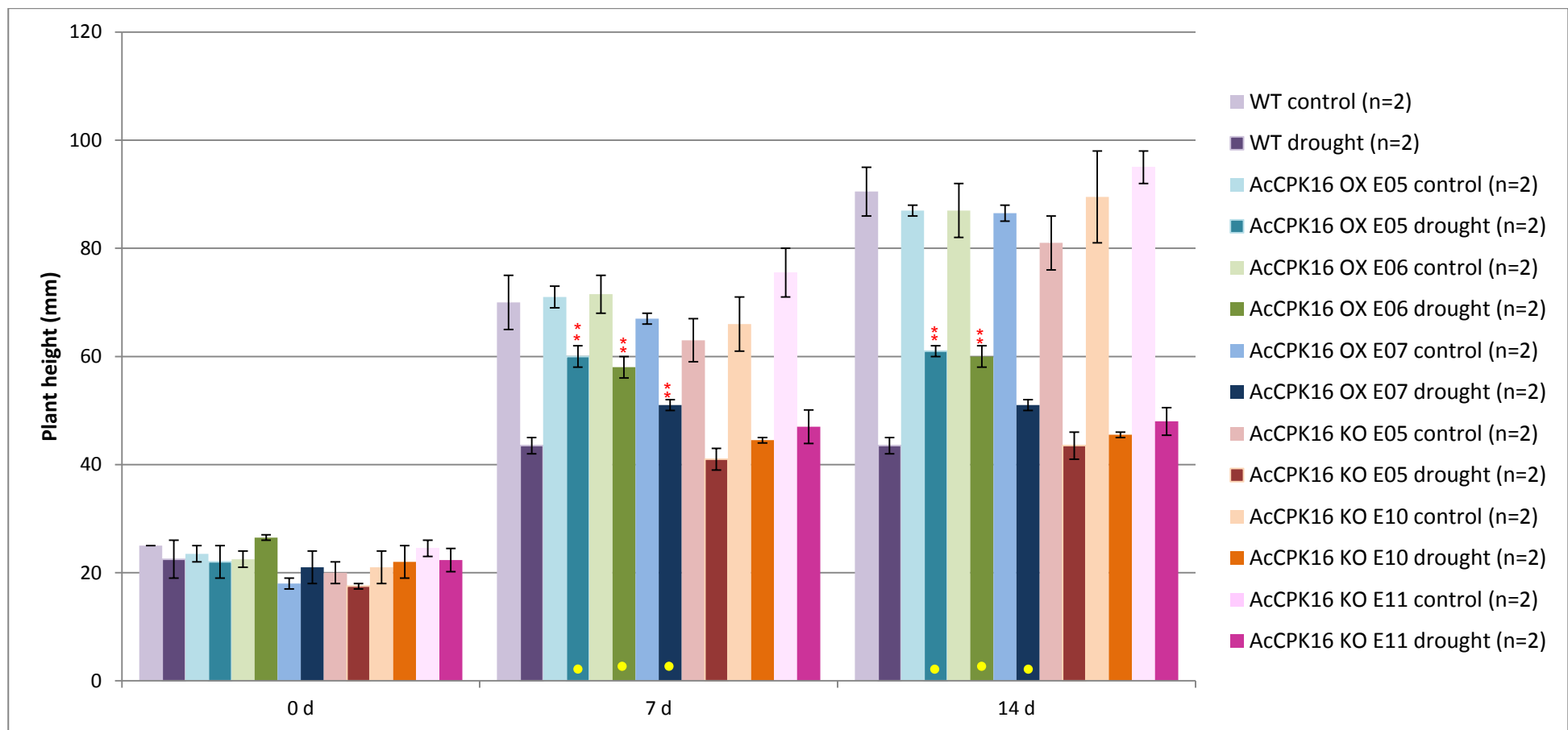
**Figure 4.48. Mean height of kiwifruit WT, OX and KO plants in response to drought.** Colour of bars match the kiwifruit lines: purple, WT kiwifruit; blue, AcCPK16 OX E05; green, AcCPK16 OX E06; dark blue, AcCPK16 OX E07; red, AcCPK16 KO E05; orange, AcCPK16 KO E10; and pink, AcCPK16 KO E11. Control treatment has lighter shade while drought treatment has darker shade. Measurements done in millimetres. Line bars indicate SE of the mean. Yellow dots indicate a marked difference between the transgenic line and WT. Statistical support is indicated as: strong (***, P≤ 0.01), good (**, 0.01<P≤ 0.05) or weak (*, 0.05<P<~0.10).

185
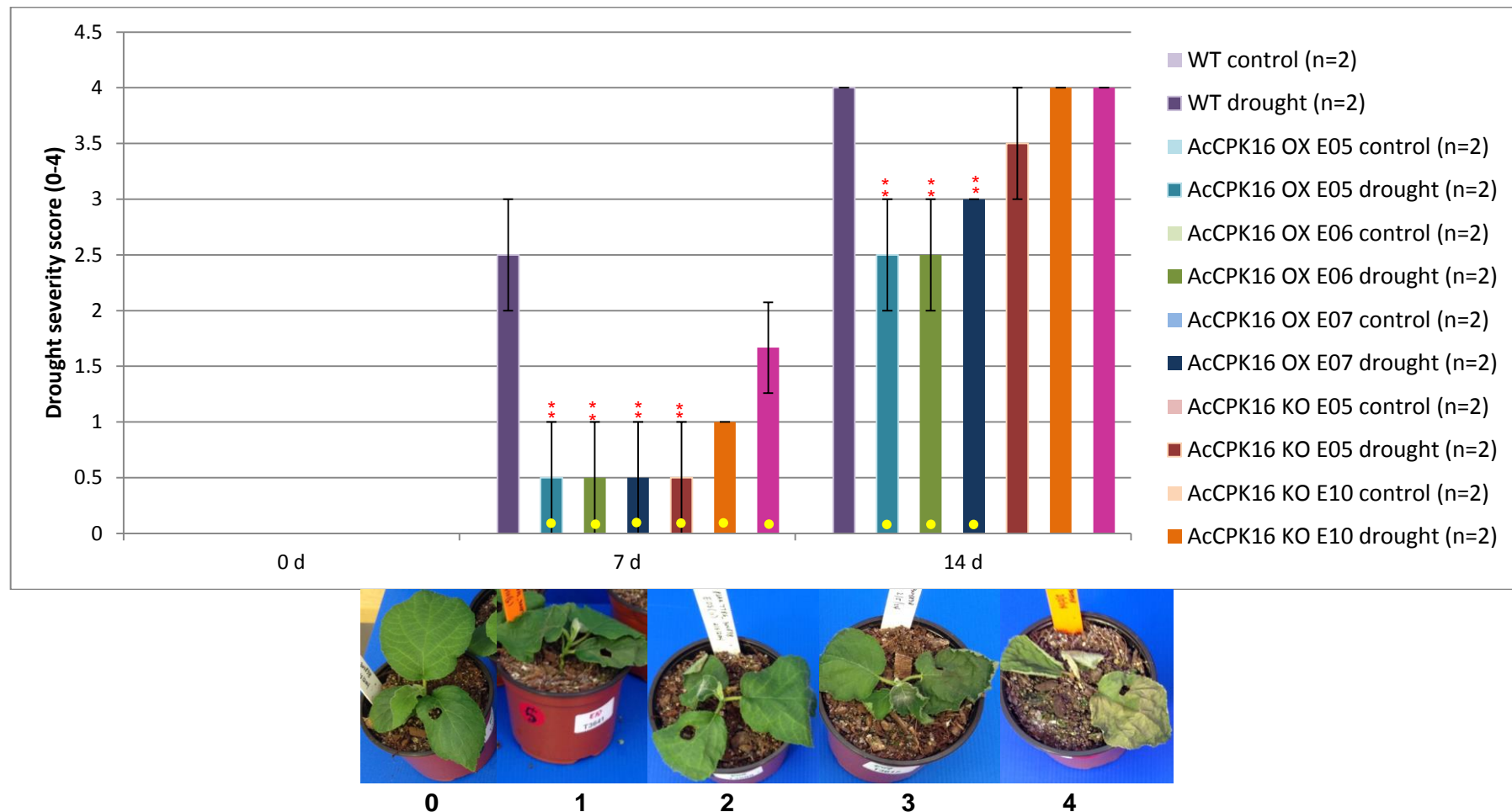
**Figure 4.49 Mean severity scores of kiwifruit WT, OX and KO plants in response to drought.** Colour of bars match the kiwifruit lines: purple, WT kiwifruit; blue, AcCPK16 OX E05; green, AcCPK16 OX E06; dark blue, AcCPK16 OX E07; red, AcCPK16 KO E05; orange, AcCPK16 KO E10; and pink, AcCPK16 KO E11. Control treatment has lighter shade while drought treatment has darker shade. Severity scores done in a scale of 0 to 4: 0, no symptoms/leaves turgid; 1, a few leaves showing symptoms/wilting; 2, most of leaves showing symptoms/wilting; 3, all of leaves showing symptoms/ wilting, some drying; 4, Plant totally dried/dead. Line bars indicate SE of the mean.Yellow dots indicate a marked difference between the transgenic line and WT. Statistical support is indicated as: strong (***, P≤ 0.01), good (**, 0.01<P≤ 0.05) or weak (*, 0.05<P<~0.10).
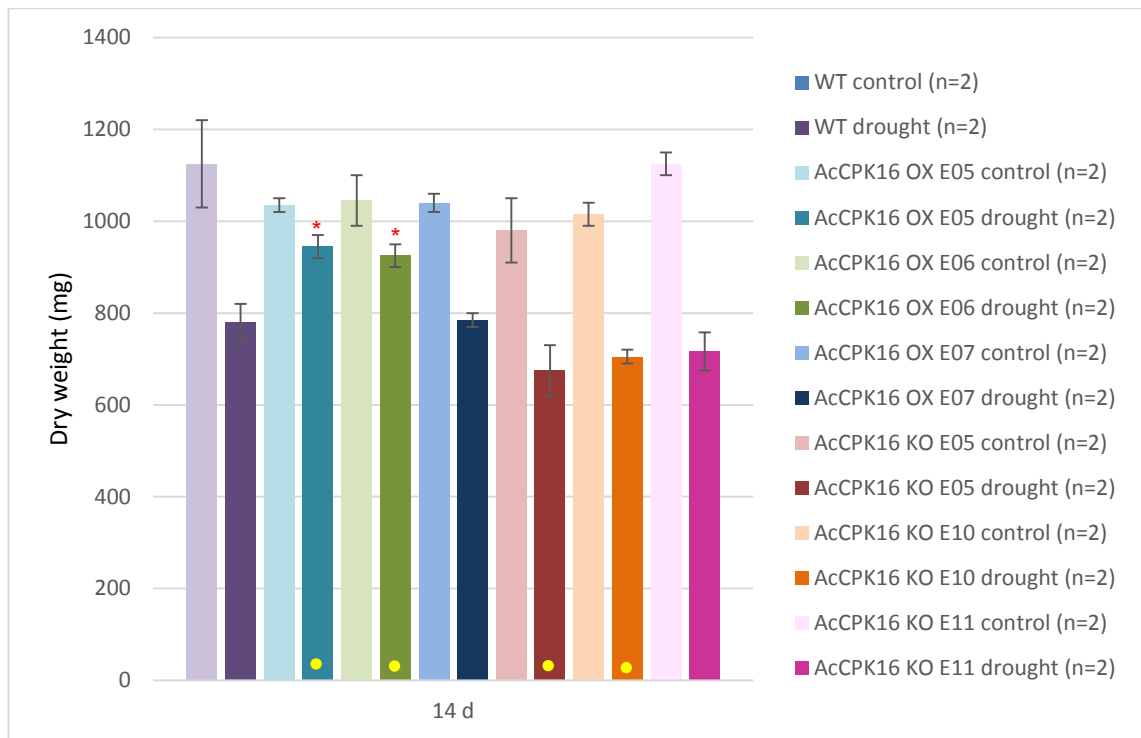
**Figure 4.50 Mean dry weights of kiwifruit WT, OX and KO plants in response to drought.** Colour of bars match the kiwifruit lines: purple, WT kiwifruit; blue, AcCPK16 OX E05; green, AcCPK16 OX E06; dark blue, AcCPK16 OX E07; red, AcCPK16 KO E05; orange, AcCPK16 KO E10; and pink, AcCPK16 KO E11. Control treatment has lighter shade while drought treatment has darker shade. Line bars indicate SE of the mean. Yellow dots indicate a marked difference between the transgenic line and WT. Statistical support is indicated as: strong (***, P≤ 0.01), good (**, 0.01<P≤ 0.05) or weak (*, 0.05<P<~0.10).

*Phenotype measurements in kiwifruit WT, OX and KO in response to B. cinerea*

Detached leaf cuttings of wild type kiwifruit, three AcCPK16 OX lines (AcCPK16 OX 05, 06 and 07) and three AcCPK16 KO lines (AcCPK16 OX 05, 06 and 07) were spot inoculated with *B. cinerea* as described in section 4.2.2.9. Measurements of susceptibility or fungal growth on the leaves were taken at 2 and 7 dpi (Figure 4.51 a). The area of growth (mm$^2$) was calculated by multiplying the biggest length and width of the irregularly shaped fungal lesion. At 2 dpi, no marked difference in fungal growth area among all lines was observed, except for KO E05 which was bigger than WT (Figure 4.51b). However, this difference did not have statistical support (Fisher's LSD P=0.192). At 7 dpi, AcCPK16 OX E05 leaves showed markedly greater fungal growth were compared to all WT and KO leaves (Figure 4.51c). This was supported by good statistical evidence (Fisher's LSD P= 0.019). No significant difference in fungal growth area was observed among WT and KO leaves. The percentage of fungal growth difference between 2 and 7 dpi was also calculated, which appeared to be higher in OX E05 and lower in KO E05 compared to WT (Figure 4.51d). There was good statistical evidence to support the marked difference of OX E05 to WT (Fisher's LSD P=0.053), but not KO E05. AcCPK16 KO E10 and E11 did not show any marked difference to WT.
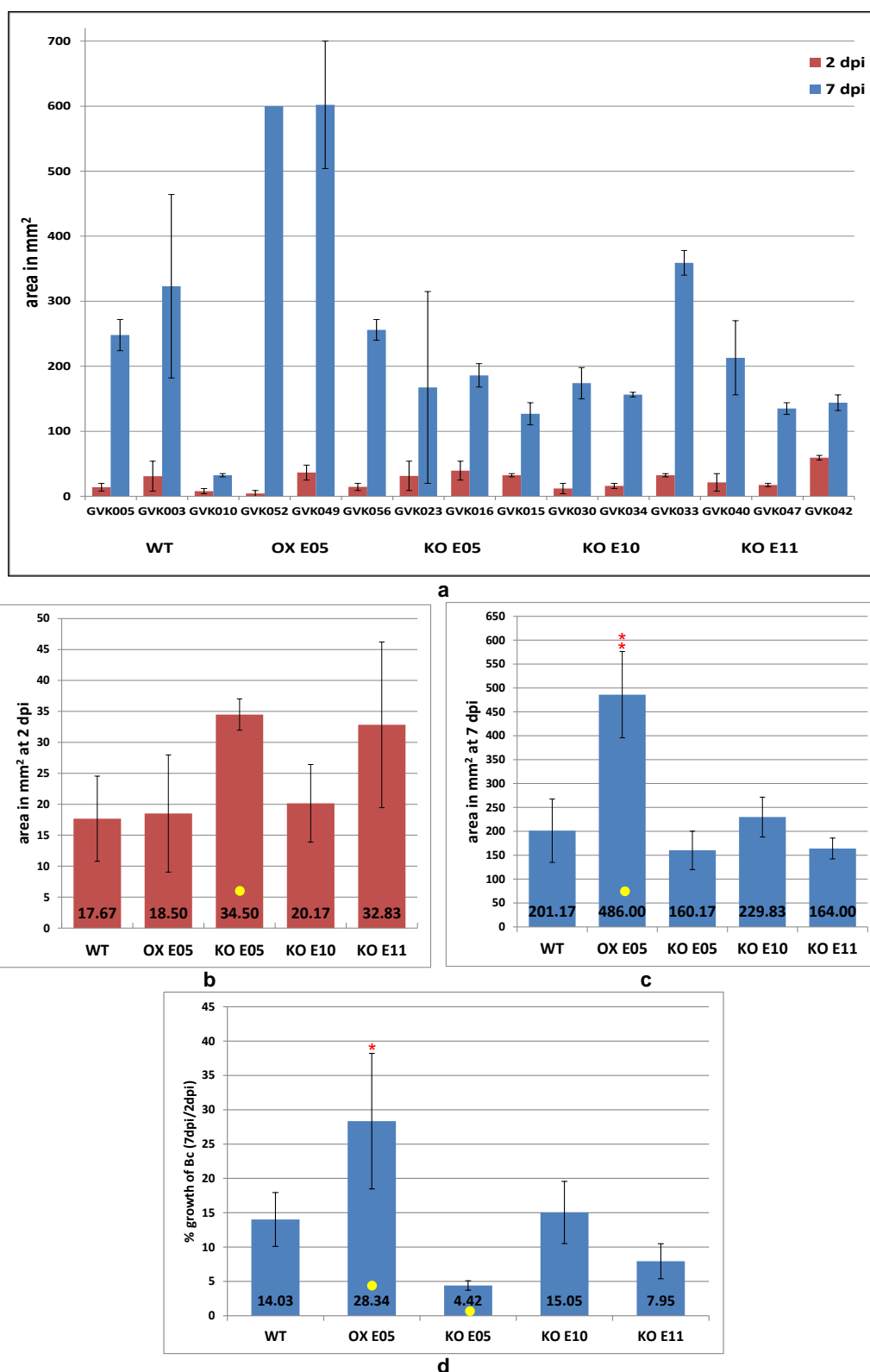
**Figure 4.51. Growth of *B. cinerea* in detached leaves of kiwifruit WT and transgenic plants.** **(a) Growth in all plants at 2 dpi (red) and 7 dpi (blue). (b) Average growth for each plant line at 2 dpi. (c) Average growth for each plant line at 7 dpi. (d) Percent growth of B. cinerea in detached leaves for each kiwifruit plant line between 2 and 7 dpi.** Growth of *B. cinerea* was measured as area in mm$^2$ (length x width) of the irregularly shaped infection zone on the leaves. Percent growth between 2 and 7 dpi was calculated by dividing the measurement at 7 dpi by the measurement at 2 dpi and multiplying by 100, for each detached leaf. Error bars indicate the SE of the mean. Yellow dots indicate a marked difference between the transgenic line and WT. Statistical support is indicated as: strong (***, P≤ 0.01), good (**, 0.01<P≤ 0.05) or weak (*, 0.05<P<~0.10).

188

**4.4 Discussion**

Four main points can be inferred from the results presented in this chapter. Firstly, it can be confirmed that Group IIb.1 CPKs are important in plant responses to environmental stress and pathogen infections, whereas Group IIb.2 CPKs are important in pollen development. For this reason, the subsequent analysis and experiments focused in Group IIb.1 CPKs. Secondly, the qPCR results measuring transcript accumulation of Group IIb.1 CPKs in Arabidopsis (AtCPK3) and rice (OsCPK1 and 15) in response to biotic and abiotic stress generally support the microarray data available from public expression databases, although there were a few differences. Thirdly, the AtCPK3 responses observed in Arabidopsis are similar to the responses of its corresponding orthologues in rice and kiwifruit (AcCPK16), but an example of potential subfunctionalisation was seen in the rice CPKs. Fourthly, the phenotype measurement experiments suggest the following: that AtCPK3 overexpression can possibly confer tolerance to drought, fungal and virus infection; that absence of AtCPK3 may or may not contribute to susceptibility to these stresses; and that AcCPK16 overexpression may also confer tolerance to drought but possible susceptibility to fungi. However, the qPCR and phenotype experiments need further validation analysis with increased sample size and optimised conditions. A summary of the findings about group IIb.1 CPKs in Arabidopsis, kiwifruit and rice in response to abiotic and biotic stresses are shown in Table 4.20 and 4.21.

**Table 4.20. Summary of results for abiotic stress**

| ABIOTIC STRESS | Group IIb.1 CPK | In silico approach | In planta approach - qPCR | In planta approach- knockouts and overexpressors |
|---|---|---|---|---|
| **DROUGHT** | **AtCPK3** | ↑ 2-fold in roots (Genevestigator)<br>↓ 1.5 fold in leaf and roots, 3-24 h (TAIR, Affymetrix chip) | ↓ 1.5-fold in leaf at 7 d and 14 d (plants in soil) | overexpressors were slightly more tolerant, knockouts were slightly more susceptible |
| | **OsCPK1** | ↓2 to 4-fold in leaf, panicle and root (Genevestigator) | ↑ 1.5-fold in leaf at 14 d | Not determined |
| | **OsCPK15** | ↑ 2-fold in some/ no change in most experiments (Genevestigator) | ↓2-fold in leaf at 14 d | Not determined |
| | **AcCPK16** | Not determined | ↓1.3 to 1.5 -fold in leaf at 7d and 14 d | overexpressors were slightly more tolerant, knockouts had no significant difference with wild type |
| **SALT** | **AtCPK3** | ↑ slight in shoots, 3-24 h (TAIR, Affymetrix chip)<br>↓ in roots, 3-24 h (TAIR, Affymetrix chip) | no significant change in leaf (plants in soil)<br>↑ slight in shoots at 15 min, 1 h and 4 h (grown on agar)<br>↑ slight in roots at 24h (grown on agar) | Not determined |
| | **OsCPK1** | ↓2-fold (Genevestigator) | ↓slight in leaf at 4 to 24 h, ↑ 1.5-fold at 14d | Not determined |
| | **OsCPK15** | no significant change (Genevestigator) | ↑ 1.5-fold in leaf at 7d and 14 d | Not determined |
| | **AcCPK16** | Not determined | ↓1.2 to 1.8 -fold in leaf at 30 min, 48h, 7 d and 14 d | Not determined |
| **MANNITOL/OSMOTIC STRESS** | **AtCPK3** | ↓ in shoots, 24 h (TAIR, Affymetrix chip)<br>↑ in roots, 1-24 h (TAIR, Affymetrix chip)<br>↓ late osmotic stress | ↓slight in shoots at 48 h (grown on agar)<br>↑ slight in shoots at 15 min (grown on agar)<br>↓slight in roots at 15 min and 1 h | Not determined |

**Table 4.21. Summary of results for biotic stress**

| BIOTIC STRESS | | In silico approach | In planta approach - qPCR | In planta approach- knockouts and overexpressors |
|---|---|---|---|---|
| BACTERIA | AtCPK3 | ↑ 1.5-fold in *P. syringae* and *X. campestris*<br><br>↓ in Pto DC3118 and Phytophthora | ↓ in Pto DC3000 at 2 dpi | Overexpressors were slightly more tolerant; knockouts had no difference with wild type |
| | OsCPK1 | ↑ in *A. tumefaciens*, ↑ in *X. oryzae* | no change in Pss | N/A |
| | OsCPK15 | ↑ in *A. tumefaciens*, no changes in *X. oryzae* | ↑ in Pss at 2, 6 and 10 dpi | N/A |
| | AcCPK16 | Not determined | Not determined | Not determined |
| FUNGI | AtCPK3 | ↓ in *A. brassicicola* at 6h | ↓ in *B. cinerea* at 2 and 6 dpi | Overexpressors were slightly more tolerant; knockouts had no difference with wild type |
| | OsCPK1 | ↓ in *M. grisea*; ↑ in *M. oryzae* | ↓ in *M. grisea* at 10 dpi | N/A |
| | OsCPK15 | ↓ in *M. grisea* except in 1 study; ↑ in M. oryzae | ↑ in *M. grisea* at 6 and 10 dpi | N/A |
| | AcCPK16 | Not determined | ↓ slight in *B. cinerea* | Overexpressors were slightly more susceptible; knockouts had no difference with wild type |
| VIRUS | AtCPK3 | ↑ 2-fold in CalCuV and TuMV | ↑ 1.5 to 4-fold in CaMV, TMV, TSWV, TuMV and TYMV at 14-35 dpi | overexpressors were slightly more tolerant and had more siliques; knockouts had no difference in symptoms with wildtype, but had less siliques |
| | OsCPK1 | Not determined | ↑ slight in CymMV at 28 dpi | Not determined |
| | OsCPK15 | Not determined | ↑ 1.5-fold in CymMV at 14 to 28 dpi | Not determined |
| | AcCPK16 | Not determined | ↑ 1.5-fold in CMV at 28 and 35 dpi | Not determined; results were inconclusive as virus symptoms only reached the adjacent systemic leaf, for all the wildtype and transgenic plants |

In the *in silico* approach, Group IIb.1 CPKs in Arabidopsis and rice (AtCPK3, OsCPK1 and OsCPK15) were shown to function in response to biotic and abiotic stresses, while Group IIb.2 CPKs (AtCPK17, AtCPK34, OsCPK2, OsCPK14, OsCPK25 and OsCPK26) were shown to function primarily in pollen development. This was inferred based on the genes' localisation throughout the plant anatomy, abundance during development and significant changes in transcript accumulation in response to stress and pathogen treatments. Group IIb.1 CPK transcripts were moderately abundant throughout the plant anatomy and all developmental stages, while Group IIb.2 CPK transcripts were only abundant in the stamen and were detectable from whole plant samples during floral development. As Group IIb.2 CPKs were only present in the flowers and not anywhere else in the plants, expression of these genes was not detected in experiments that explored transcript level changes in response to various stresses. Group IIb.1 CPKs on the other hand were detectable in leaves, shoots and roots and showed significant changes in transcript accumulation in response to various treatments. Group IIb.2 CPKs were also reported to be important in the development of pollen in Arabidopsis (Myers et al. 2009; Zhou et al. 2009), although no similar studies have yet been performed in rice. Conversely, Group IIb.1 CPKs were reported to function in abiotic and biotic stresses such as drought, high salinity and bacterial infections (Arimura and Sawasaki 2010; Boudsocq et al. 2010; Cousson 2011; Hubbard et al. 2012; Latz et al. 2013; Mehlmer et al. 2010; Munemasa et al. 2011).

Group IIb.1 CPKs that have been examined in biological experiments in Arabidopsis, rice and kiwifruit carried out in this study showed changes in transcript accumulation in response to abiotic stress and pathogen infections (Table 4.20 and 4. 21). In response to drought, AtCPK3 decreased in transcript accumulation, both based on the information from the experiment by Kilian et al. (Kilian et al. 2007) as shown in the Arabidopsis eFP Browser database (Winter et al. 2007) and on the qPCR analysis performed, but not with the two studies detected in Genevestigator. Decrease in transcript accumulation was also seen for OsCPK15 and AcCPK16 in the qPCR experiments performed, but not with OsCPK1. Interestingly, OsCPK1 showed a decrease in transcript accumulation in the experiments detected in Genevestigator, while OsCPK15 showed no significant change in most experiments; and in one case it had increased. The contrasting differences may be due to variations in the experimental set up, plant cultivar, the time point of isolation, the developmental stage of the plants and the tissue samples taken. The experiment by Killian et al (2007) involved leaf and root samples (wild type *col-0* ecotype) and the qPCR

experiment in this study utilised leaf samples (wild type *col-0* ecotype), while the experiments presented in Genevestigator used roots from wild type *col-0* and whole plant samples from an *srk2dei* mutant. There is therefore a need for confirmatory studies using higher numbers of biological replicates, different ecotypes and several tissue types sampled in a time series. Nonetheless, in general it can be inferred that AtCPK3, AcCPK16 and either OsCPK1/OsCPK15 were downregulated in response to drought conditions.

In response to high salinity, Group IIb.1 CPKs showed varying results. AtCPK3 showed a slight increase in shoots based on the TAIR database, which correlates with the qPCR analysis done on agar-grown plants. The Arabidopsis plants grown on soil however did not show significant change in response to salt. The results in the root samples were, however, conflicting. AtCPK3 decreased in roots in Killian et al. (2007), but slightly increased in the qPCR experiment done. This contrast in results could be due to the difference in the experimental set-ups. In their experiment, the plants were grown in specialised rafts that were initially placed on MS agar and then transferred in MS liquid media after 11 days, where 150 mM NaCl was added. In the study described in this chapter (section 4.3.2.2), the plants were grown on MS agar for two weeks and then 100 mM or 200mM NaCl was added to the medium. It is possible that the consistency of the medium used affected the response in roots, whereas the responses in the shoots were similar. OsCPK1 decreased slightly in leaf at 4 h to 24 h in the qPCR experiment, which correlates with the two-fold decrease in the experiment described in Genevestigator. However, OsCPK1 increased by 1.5-fold at 14 d. On the other hand OsCPK15 increased both in the qPCR experiment done (at 7 d and 14 d) and in the experiment described in Genevestigator. In kiwifruit, AcCPK16 continuously decreased throughout the time series of sampling between 30 min and 14 d. Based on these results, it can be inferred that in response to high salinity AtCPK3 and OsCPK15 both increased in shoot tissue while OsCPK1 and AcCPK16 decreased in these tissues.

In response to bacteria, Group IIb.1 CPKs seem to have unique responses depending on the bacterial pathogen. In the Genevestigator experiments AtCPK3 showed an increase in response to different *P. syringae* mutant strains and *X. campestris*, but decreased in response to *P. parasitica* and a mutant strain of *P. syringae* DC3118. In the qPCR experiment performed with a wild type *P. syringae* DC3000 strain, AtCPK3 decreased. In Genevestigator both OsCPK1 and OsCPK15 increased in response to *A. tumefaciens* but only OsCPK1 increased in response to *X.*

*oryzae.* In the qPCR experiment, OsCPK15 showed an increase in response to *P. syringae* while OsCPK1 did not show any change.

In response to fungi, Group IIb.2 CPKs mostly decreased in transcript accumulation. In the qPCR experiment done, AtCPK3, OsCPK1 and AcCPK16 mRNA decreased in leaves in response to fungi, whereas OsCPK15 increased. In the Genevestigator experiments, AtCPK3 decreased in response to *A. brassicicola* and OsCPK1 and OsCPK15 decreased in response to *M grisea.* Transcripts of both the rice genes however showed increased accumulation in response to *M. oryzae.*

In response to virus, all Group IIb.1 CPKs showed increased transcript accumulation, regardless of the type of virus or the plant species. AtCPK3 showed 1.5 to four-fold increase in response to the five plant viruses tested, OsCPK1 and OsCPK15 both increased slightly in response to CymMV and AcCPK16 increased in response to CMV. The increase in transcript accumulation was observed during the later stages of the infections, between 14 to 35 dpi.

Genes that are up- or down-regulated in response to stress, pathogen infection or other stimuli potentially play a role in cellular pathways related to these responses. Therefore, AtCPK3 and its orthologues, being kinases, potentially phosphorylate proteins involved in cellular signalling pathways of plant stress and pathogen responses. In previous studies as mentioned in Chapter 3, AtCPK3 was found to function in the regulation of guard cell channels and induction of stomatal closure along with AtCPK6 (Mori et al. 2006) and in MAPK-independent salt stress acclimatisation (Mehlmer et al. 2010). In recent studies, AtCPK3 was reported to phosphorylate 14-3-3 proteins *in vitro* in response to sphingolipid that leads to programmed cell death (Lachaud et al. 2013) and *in vivo* using MS analysis (Swatek et al. 2014). Moreover, AtCPK3 was found to phosphorylate a 14-3-3 binding motif in vacuolar two-pore K+ channel 1 (TPK1) involved in salt stress adaptation. The slight increase in drought tolerance seen in both AtCPK3 and AcCPK16 overexpressors in the phenotype analysis could be supported by these findings; however, further studies are required because of the limited number of plants available when this study was conducted. Furthermore, the AtCPK3 knockouts were only slightly susceptible and the AcCPK16 did not show any difference with wild type, which suggests that AtCPK3 and its orthologues have redundancy in function with other CPK genes, such as AtCPK6. The qPCR experiments showed that AtCPK3 and its orthologues generally decrease in transcript accumulation during drought

and osmotic pressure, which was not expected as they function in response to these stresses. Why and how they are downregulated in terms of transcript accumulation remains to be determined. It is possible that the AtCPK3 protein has become stabilised, or show increased function despite decreased transcript accumulation. Further studies are required to determine the exact mechanism of AtCPK3 transcription, translation and protein activity during drought conditions.

In response to biotic stress, Group IIb.1 CPKs transcript levels changed in response to bacteria, decreased in response to fungi and increased in response to virus. CPKs appeared to vary in response to different bacteria, which may indicate that these CPK responses are specific to certain bacterial effectors. The decrease in these CPKs in response to fungi may indicate that they are involved in cellular signalling pathways that are hindered by fungal infections. Increase in these CPKs in response to virus could indicate that they could be involved in RNA silencing pathways. Further investigation, however, is required to test these suppositions. For example, plants can be treated with different bacterial or fungal effectors and determine the changes in transcript accumulation and detect protein-protein interactions. Similar experiments can be done to determine if Group IIb.1 CPKs could have substrates that are known to be involved in plant viral defence or could be targets of viral suppressors of RNA silencing.

A potential case of subfunctionalisation was observed in Group IIb.1 CPKs in rice, between OsCPK1 and OsCPK15. While duplicate genes were common in monocots, probably as a result of multiple polyploidisation events, the specific function and response of these genes may still differ despite the high similarity in gene sequence. In the example seen in OsCPK1 and 15, there was a contradicting level in transcript accumulation in response to drought, salt and fungal infection, while there was a significant increase in OsCPK15 mRNA in response to virus but insignificant increase in OsCPK1 mRNA, and a small increase in OsCPK15 mRNA in response to bacteria but no change in OsCPK1 mRNA. Further investigation is required to verify this hypothesis, which may involve analysing single and double knockouts and overexpressors of these genes in response to different stress and pathogen infections.

Lastly, the phenotype experiments have demonstrated that AtCPK3 overexpressors showed some level of tolerance to drought, fungal and virus infection while AtCPK3 knockouts showed susceptibility to these stresses and AcCPK16 overexpressors showed some tolerance to

drought. These preliminary findings suggest that overexpression of Group IIb.1 CPKs may confer tolerance to these stresses and infections; however, further experiments using a higher number of samples and different kinds of treatments and infections must be performed.

The experiments performed in this thesis project were limited by the number of plants available for analysis, the treatments and pathogens available and permitted for laboratory handling, and also the number of samples that could be managed within the facility, the resources and the time frame. With regards the kiwifruit experiments, the number of plants were particularly limited because of the amount of seeds that were available and that germinated in each batch of experiments, as well as the amount of transgenic kiwifruit plants that were developed for use in the phenotype analysis. Moreover, only drought and fungal infections were confirmed in kiwifruit. The virus infection phenotype experiment in transgenic kiwifruit was inconclusive as the symptoms were present only in a few of the inoculated leaves and the adjacent leaves; the virus could not be detected and symptoms disappeared in the subsequent leaves. This may mean that the virus used in kiwifruit (CMV) was unable to move throughout the plants' anatomy and failed to persist in the plants. The case was different when virus infection was performed for measuring infections in wild-type plants alone, as the infection was successful in all five plants inoculated. The difference in virus symptomology may be due to the origin of the plant samples, as the wild-type plants used for measuring transcript accumulation were grown from seed, whereas the plants for phenotype experiments were grown from tissue culture. Moreover, inherent properties of the plant lines might also have rendered them slightly resistant to the virus. It is also highly possible that infectivity of the inoculum material might have changed through the duration of the experiments. Future research may involve the use of several viruses (CMV was used because of restrictions in the facility and availability of inoculum material), higher number of plants to be inoculated, and the use or propagation of wild-type/ vector-only lines that show susceptibility to the virus.

The findings in this thesis with regards Group IIb.1 CPKs were limited to transcript accumulation and phenotype measurements alone. Analysis of protein levels, determination of substrates, and analysis of biochemical properties and protein-protein interactions and networks are necessary to provide further understanding of the role of the most conserved CPKs. Determination of AtCPK3 protein levels and activity was attempted in this study, however, no

specificity was determined in the antibody designed. Future research must involve the factors mentioned above, focusing on the characteristics and interactions of the protein itself.

# Chapter Five

# What influences CPK functional specificity?

## 5.1 Introduction

There is a need to determine what factors influence CPK functional specificity because of the overlapping and redundant functions within subgroups and because of the distinct difference in function of the two subgroups within the most evolutionarily conserved group, Group IIb. Factors that potentially influence protein function include protein structure, gene regulatory elements that control gene expression, and the localisation of the expressed gene. Protein structure may influence the function of CPKs, both in their sensor and responder functions. Difference in structure within the PK domain, particularly around the active sites, can determine substrate specificity, binding intensity and phosphorylation activity. Similarly, differences in the CAD may influence the protein's sensitivity to $Ca^{2+}$ ions, especially within the EF hands. Difference in structure within the N-VD may also influence the protein's cellular localisation and membrane binding. Gene regulatory elements may determine functional specificity as they control whether a gene is transcribed and expressed into proteins. These elements may function in response to specific stimuli, which include abiotic and biotic stresses and developmental signals. Lastly, the tissue localisation of a gene's transcript or expressed protein may correlate with function, especially if a gene is exclusively expressed in particular tissues.

This chapter explores the different factors that may influence CPK functional specificity. In brief, this chapter presents the following: 1) protein motifs in the primary, secondary and tertiary levels that may influence CPK function; 2) gene regulatory regions that may correlate with function; and 3) tissue localisation differences that may correlate with function. The first two aims involved group IIb CPKs in the model plants Arabidopsis and rice, while the last aim only focused

on group IIb CPKs in Arabidopsis due to limitations in the timeframe of research and NZ restrictions in facilities for research involving rice.

## 5.2 Materials and Methods

### 5.2.1 Primary and secondary structure analysis

Multiple alignments were made in order to compare the protein sequence of Group IIb.1 CPKs (stress-responsive) against Group IIb.2 CPKs (developmental) in Arabidopsis and rice. In Arabidopsis, AtCPK3 was compared with AtCPK17 and 34, while in rice OsCPK1 and 15 were compared with OsCPK14, 2, 25 and 26. Secondary structures were predicted using the EMBOSS Protein analysis tool plugin within Geneious 8.0 (http://www.geneious.com, Kearse et al., 2012). Sequence alignments were analysed to determine patterns of amino acid sequence and secondary structure that are unique or distinctive to each subgroup; patterns were assigned as motifs and numbered accordingly. Structure analysis and motif assignments were done with Arabidopsis CPKs at first, and then applied to rice CPKs. The gross structure of the proteins were also compared.

### 5.2.2 Tertiary structure prediction and analysis

Tertiary structures of all Group IIb CPKs from Arabidopsis and rice were predicted using two tertiary structure prediction tools, Swiss-Model (Bordoli et al. 2009) and I-TASSER (Roy et al. 2010). These tools detected known protein structures from the Protein Data Bank that are similar to the CPK sequence entered. The Swiss-Model tool chose the structure with the highest similarity to the CPK sequence as the template for 3D structure prediction, while I-TASSER chose the top five structures and used them as templates by threading, utilising the Monte Carlo method. Statistical support for the structures constructed were noted for each of the two methods.

The structures constructed using the two methods were compared in terms of statistical support and coverage for the CPK sequences. Structures predicted using I-TASSER were used for the 3D structure comparison analysis. The 3D structures were viewed, aligned and analysed using the molecular structure viewer software called PyMol (PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC). Colours were assigned to a CPK sequence and/or a region within the CPK. Analyses were done for Arabidopsis and rice CPKs separately at first, and then combined.

### 5.2.3 Gene regulatory structure analysis

The location of AtCPK3, 17 and 34 within the genome were determined using the NCBI MapViewer database (Entrez MapViewer, NCBI) and Gene (Entrez Gene, NCBI). The sequences of non-coding upstream flanking regions were extracted from these databases and were analysed for the presence of transcription factor binding sites. These binding sites were detected using an online tool called MatInspector (Cartharius et al. 2005, Genomatix Software GmBH), which utilises a library of DNA sequences that have been annotated as transcription factor binding sites with correlating functions. Detected binding sites were determined for each of the three genes and AtCPK3 sites were compared and contrasted with AtCPK17 and 34 sites in relation to functional annotations. The same procedure was performed for their orthologues in rice.

### 5.2.4 Development of Arabidopsis plants that overexpress AtCPK34

Detailed method of developing of CPK overexpressing lines of Arabidopsis has been described in section 4.2.2.6. Briefly, the full AtCPK34 gene was PCR amplified from an Arabidopsis leaf cDNA using AtCPK34 gene specific primers that were flanked with Gateway attB sequence: forward primer GatewayAtCPK34F1- 5'GGGGGACAAGTTTGTACAAAAAAGCAGG CTATGGGAAATTGTTGCTCTCATGGAAGA3' and reverse primer Gateway AtCPK34R1572- 5'GGGGACCACTTTGTACAAGAAAGCTGGGTTCATTTGAATGATAGTTCACGCCGCTTCTTA GGATTA3'. The entry vector used was pDONR/Zeo (Gateway®) and the destination vector used was pHEX2 (obtained from S. Karunarietnam, PFR). The PCR reaction components and conditions were the same as described in section 4.2.2.6a but using the GatewayAtCPK34 forward and reverse primers and using healthy Arabidopsis flower cDNA. Bacterial transformations were done using OneShot® TOP10 chemically competent *E.coli* (Life Technologies) following the manufacturer's procedure. BP and LR reactions were carried out as described in section 4.2.2.6a, with 1:1 insert to vector molar ratio in both reaction steps. Plasmids from entry and destination clones were extracted using GenElute™ HP Plasmid Miniprep kit (Sigma-Aldrich, St. Louis, MO, USA). Plasmids were sent for full sequencing of inserts to Macrogen Inc. using M13 forward and reverse primers in order to confirm the sequence and its orientation in each step, as described in section 4.2.2.6a.

Transformation of the expression clones pHEX2_AtCPK34Full into *Agrobacterium tumefaciens* was done by electroporation as described in section 4.2.2.6a. Large-scale culture from a screened colony was used to transform three pots of Arabidopsis (about 20 seedlings per

pot) with young inflorescence by the floral dipping method. Seeds were allowed to develop for three weeks and were dried before collection. Successful transformants were selected by growing seeds in ½ MS agar with 100 mg/mL kanamycin. Plant selection cycles on kanamycin were performed until the fourth (T4) generation to ensure homozygosity.

### 5.2.5 Seed germination assay

The germination rate of wild type Arabidopsis, AtCPK3 knockouts, AtCPK3 overexpressors, AtCPK34 knockouts and AtCPK34 overexpressors under normal conditions and high salt conditions was determined. Seeds were surface sterilised as described in section 4.2.2.6d and were sown on Hoagland's Medium plates (Appendix 22) that contain 0 mM, 75 mM, 150 mM and 300 mM NaCl. Seeds were allowed to sprout and plantlets were counted at 7 d. Three replicate experiments with about 100-200 plantlets in each plate were performed.

### 5.2.6 Pollen germination assay

This assay was adapted from Myers et al. (Myers et al. 2009). Inflorescences from wild type Arabidopsis, AtCPK3 knockouts, AtCPK3 overexpressors, AtCPK34 knockouts and AtCPK34 overexpressors were collected and pollen-containing anthers were touched onto separate microscope slides that contain 30 µL of liquid pollen growth medium (Appendix 22). Slides were inverted and kept in a wet chamber for 6 h. Chambers used were comprised of covered petri dishes with wet filter paper inside to keep the system moist and Blutak to make raised surfaces. Slides were then reinverted and stained with lactophenol blue (Sousa et al. 2013) and observed under a compound microscope at 400x magnification. The number of germinating and non-germinating pollen were counted. Also, the number of pollen with noticeably long pollen tube were counted. All microscopes and equipment used were wiped with 70% ethanol after use to prevent escape of pollen outside PC2 containment.

### 5.2.7 Statistical analysis

Statistical analysis for seed and pollen germination assays were done as described in section 4.2.2.10.

**5.3 Results**

**5.3.1 Are there motifs in the primary and secondary structure that potentially influence CPK function?**

*5.3.1.1 Primary and secondary structure analysis of AtCPKs and OsCPKs*

*5.3.1.1a Separate analysis of AtCPKs and OsCPKs*

The primary and secondary structure analysis of CPKs was initially done separately and independently in Arabidopsis and rice. In Arabidopsis, the analysis showed sixteen motifs that differentiate the stress responsive CPK, AtCPK3 (Group IIb.1) from the developmental CPKs AtCPK17 and 34 (Group IIb.2) (Figure 5.1). On the other hand, rice CPKs showed fourteen motifs that differentiate Group IIb.1 members OsCPK1 and 15 from Group IIb.2 members OsCPK14, 2, 25 and 26 (Figure 5.2).

In both species, the identified motifs were found along the entire length of the CPKs (Figures 5.1 and 5.2). In the N-VD, two motifs were identified in Arabidopsis (motif At 1 and 2) while three motifs were identified in rice (motif Os 1, 2 and 3). Motifs At 1 and Os 1 both span the first 10-20 aa in the N-terminal end of the N-VD while At 2, Os 2 and Os 3 were near the PK domain. In the PK, nine motifs were identified in Arabidopsis (motif At 6 to 11) while six motifs were identified in rice (motif Os 4 to 9). A number of these motifs were located around active sites in the PK which may be important in substrate specificity. These were motifs At 3 to 8 and motifs Os 4 to 8. Motifs At 11 and Os 9 were adjacent to the AJ. Within the AJ, one motif was identified in Arabidopsis (motif At 12) while two motifs were identified in rice (motif Os 10 and 11). Motifs At 12 and Os 11 were adjacent to the beginning of the CAD, while motif Os 10 was in the middle part of the AJ. In the CAD, four motifs were identified in Arabidopsis (motif At 13 to 16) while three motifs were identified in rice. All of these motifs were located around the calcium binding sites except for motif Os 14, which was located in the last few aa of the CPKs. These locations may be important in CPK-specific calcium binding activity.
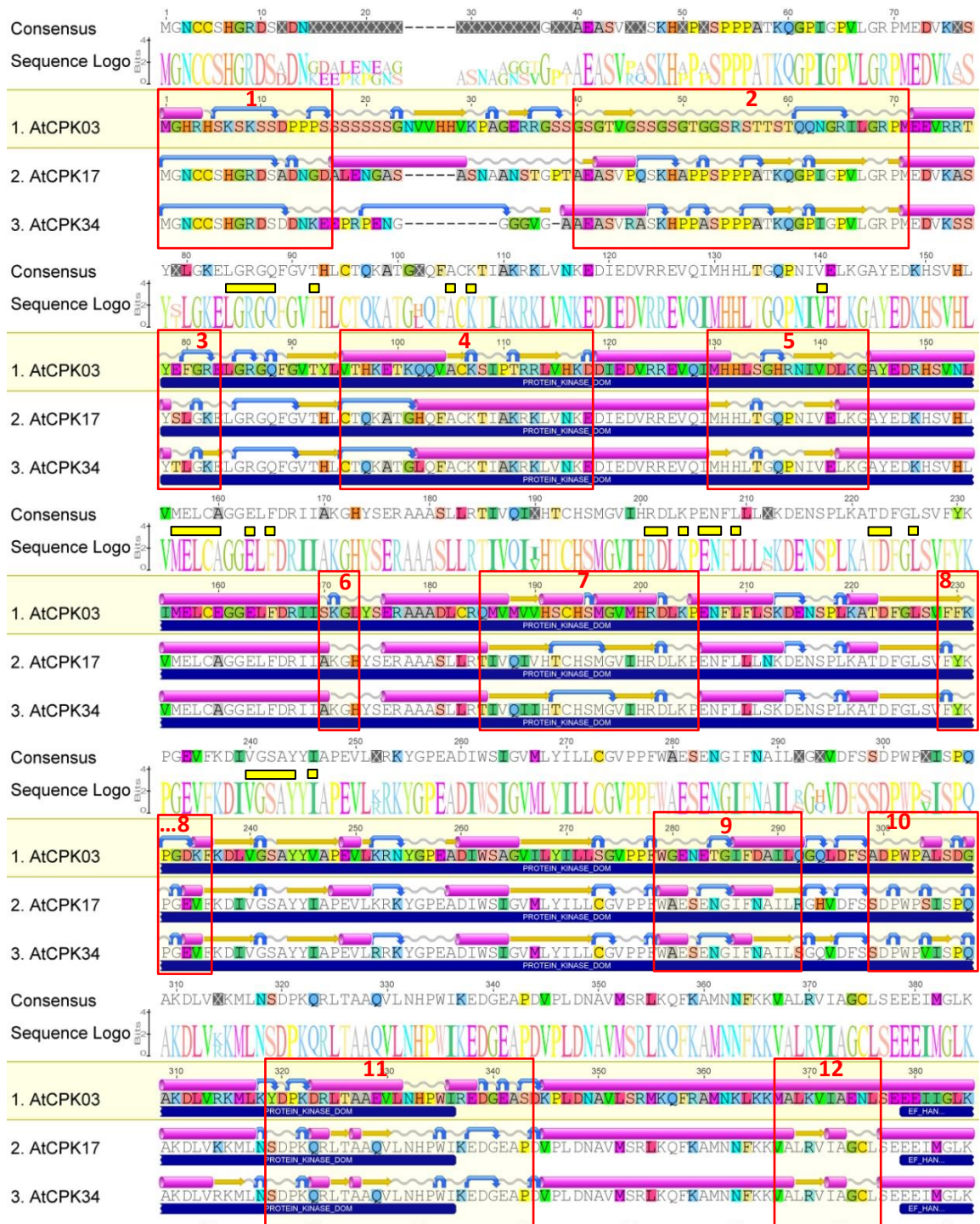
**Figure 5.1 Primary and secondary structure alignments comparing AtCPK3 against 17 and 34.** Unique motifs highlighted in numbered boxes. Loops shown as grey lines, turns as blue U-turn arrows, α-helices shown as pink cylinders, and β-strands as yellow block arrows. Phosphorylation active sites are indicated as yellow boxes and calcium binding sites as teal boxes. The PK domains and EF hands in CAD are indicated by navy blue boxes. Secondary structures generated using Geneious 6.0 (Continued on next page)

**Figure 5.1 Primary and secondary structure alignments comparing AtCPK3 against 17 and 34.** Unique motifs highlighted in numbered boxes. Loops shown as grey lines, turns as blue U-turn arrows, α-helices shown as pink cylinders, and β-strands as yellow block arrows. Phosphorylation active sites are indicated as yellow boxes and calcium binding sites as teal boxes. The PK domains and EF hands in CAD are indicated by navy blue boxes. Secondary structures generated using Geneious 6.0 (Continued from previous page)

**Figure 5.2 Primary structure and secondary structure alignments comparing OsCPK1 and 15 (AtCPK3 orthologues in rice) against 2, 14, 25 and 26.** Unique motifs highlighted in numbered boxes. Loops shown as grey lines, turns as blue U-turn arrows, α-helices shown as pink cylinders, and β-strands as yellow block arrows. Phosphorylation active sites are indicated as yellow boxes and calcium binding sites as teal boxes. Secondary structures generated using Geneious 6.0 (Continued on next page)

**Figure 5.2 Primary structure and secondary structure alignments comparing OsCPK1 and 15 (AtCPK3 orthologues in rice) against 2, 14, 25 and 26.** Unique motifs highlighted in numbered boxes. Loops shown as grey lines, turns as blue U-turn arrows, α-helices shown as pink cylinders, and β-strands as yellow block arrows. Phosphorylation active sites are indicated as yellow boxes and calcium binding sites as teal boxes. Secondary structures generated using Geneious 6.0 (Continued from previous page)

### 5.3.1.1b Combined analysis of AtCPKs and OsCPKs

After the independent analyses, a combined motif analysis of the At and Os CPKs was also performed (Figure 5.3). It is notable that many of the subgroup-specific motifs identified in At and Os matched with each other in terms of aa position and high sequence similarity within subgroups (Group IIb.1 or IIb.2). This includes At motif 1 and Os motif 1 (N-VD), At motif 3 and Os motif 4 (beginning of PK), At motif 4 and Os motif 5 (PK), At motif 5 and Os motif 6 (PK), At motif 7 and Os motif 7 (PK), At motif 8 and Os motif 8 (PK), At motif 11 and OsCPK9 (end of PK), At motif 12 and Os motif 11 (AJ), and At motif 15 and Os motif 14 (CT); which were shown in the previous section.

A combined alignment of Group IIb AtCPKs and OsCPKs was done as shown in Figure 5.3. Motifs that differ between Group IIb.1 and IIb.2 AtCPKs and OsCPKs were identified and named as AtOs 1 to 14. This consists of the nine sets of motifs listed above, together with another five motifs which were only identified in the combined alignment. Similar to the separate analyses, these motifs were found along the entire length of the CPKs. Two motifs were identified in the NV-D (AtOs motif 1 and 2), seven motifs in PK (AtOs motif 3 to 9), one in AJ, (AtOs motif 10), three in CAD (AtOs motif 11 to 13) and one in CT (AtOs 14). As these motifs were identified between a monocot and a dicot species, they are potentially important in functional specificity that is conserved among monocots and dicots. Further analysis of these motifs and their predicted tertiary structure was then performed as described in the succeeding section (section 5.3.1.2).

**Figure 5.3. Primary structure and secondary structure alignments comparing AtCPK3, OsCPK1 and 15 (Group IIb.1) against AtCPK17, AtCPK34, OsCPK2, OsCPK14, OsCPK25 and OsCPK26 (Group IIb.2).** Unique motifs highlighted in numbered boxes. Loops shown as grey lines, turns as blue U-turn arrows, α-helices shown as pink cylinders, and β-strands as yellow block arrows. Phosphorylation active sites are indicated as yellow boxes and calcium binding sites as teal boxes. Secondary structures generated using Geneious 6.0 (Continued on next page).

**Figure 5.3 Primary structure and secondary structure alignments comparing AtCPK3, OsCPK1 and 15 (Group IIb.1) against AtCPK17, AtCPK34, OsCPK2, OsCPK14, OsCPK25 and OsCPK26 (Group IIB.2).** Unique motifs highlighted in numbered boxes. Loops shown as grey lines, turns as blue U-turn arrows, α-helices shown as pink cylinders, and β-strands as yellow block arrows. Phosphorylation active sites are indicated as yellow boxes and calcium binding sites as teal boxes. Secondary structures generated using Geneious 6.0 (Continued from previous page and on next page).

**Figure 5.3 Primary structure and secondary structure alignments comparing AtCPK3, OsCPK1 and 15 (Group IIb.1) against AtCPK17, AtCPK34, OsCPK2, OsCPK14, OsCPK25 and OsCPK26 (Group IIB.2.** Unique motifs highlighted in numbered boxes. Loops shown as grey lines, turns as blue U-turn arrows, α-helices shown as pink cylinders, and β-strands as yellow block arrows. Phosphorylation active sites are indicated as yellow boxes and calcium binding sites as teal boxes. Secondary structures generated using Geneious 6.0 (Continued from previous page and on next page).

**Figure 5.3 Primary structure and secondary structure alignments comparing AtCPK3, OsCPK1 and 15 (Group IIb.1) against AtCPK17, AtCPK34, OsCPK2, OsCPK14, OsCPK25 and OsCPK26 (Group IIB.2).** Unique motifs highlighted in numbered boxes. Loops shown as grey lines, turns as blue U-turn arrows, α-helices shown as pink cylinders, and β-strands as yellow block arrows. Phosphorylation active sites are indicated as yellow boxes and calcium binding sites as teal boxes. Secondary structures generated using Geneious 6.0 (Continued from previous page).

### 5.3.1.2 Secondary and tertiary structure analysis of AtCPKs and OsCPKs

Since most Group IIb.1 CPKs are stress-responsive and Group IIb.2 CPKs are developmental, the identified motifs that differentiate the subgroups may influence tertiary structure that render a CPK as stress responsive or developmental. However, further investigation is required in order to support this hypothesis, such as tertiary structure analysis and gene functional analysis. Tertiary structure is addressed in the rest of this section but is limited to predicted CPK structures. Gene functional analysis will be addressed in the succeeding sections through comparisons in predicted gene regulatory regions (section 5.3.2) and comparisons in biological functions (section 5.3.3) such as pollen germination (development) rate and seed germination rate in high salinity (stress) among gene overexpression and knock out mutants.

The predicted tertiary structures of Group IIb CPKs in Arabidopsis and rice were analysed and compared in terms of their general structure and of the combined At and Os motifs identified in the primary and secondary structure analyses (section 5.3.1.1b). Two tertiary structure prediction methods were utilised, as mentioned in section 5.2.2. Swiss-Model used the sequence with highest similarity to the CPK as the template, whereas I-TASSER used the top five highly similar sequences to the target and used the combination of these structures as template by threading (Markov Model). The resulting predicted structures constructed using Swiss-Model and I-TASSER did not have very high statistical support, although values were acceptable for the I-TASSER structures (Appendix 29). The statistical support is influenced by the limitations in the availability of structures to be used as template. Only protist CPK structures were available as full sequence, which were expected to have a greater degree of sequence variation from plant CPKs. Available plant CPK structures from Arabidopsis (PDB ID 2AAO) and soybean (PDB ID 1S6J and 1S6I) were only limited to CAD and some to AJ. Moreover, since the known tertiary structures come from protist CPKs that have relatively short N-VD and CT, most of the Swiss-Model predicted structures only cover the PK, AJ and CAD domains. The I-TASSER structures cover the entire protein as the analysis also include *de novo* protein structure prediction in the absence of known structure within a region. Because of this, analysis and comparisons of the structures were only performed for I-TASSER predicted structures and not the Swiss Model structures.

In terms of their general structure, Group IIb.1 CPKs showed regions that are distinct from Group IIb.2, particularly in the N-VD, PK and CT regions (Figure 5.4). In the angle shown in Figure 5.4, the N-VD (red shade) in Group IIb.2 CPKs appeared to span the middle part of the molecule as an S-shaped structure, from the right hemisphere across to the left hemisphere, except for OsCPK02 where a small part in the far left was not covered. In contrast, among Group IIb.1 CPKs, the N-VD did not cover the entire left hemisphere, particularly with AtCPK03 where the left hemisphere was not reached by this domain. The PK domain appeared as a cup-shaped structure (orange) in the upper hemisphere of the molecule, with the active sites (yellow) in the centre of the cup-shaped opening. This opening appeared to be more pointed or v-shaped among Group IIb.1 CPKs compared to Group IIb.2 CPKs, which have more curved or irregularly shaped openings. Lastly, the CT region (purple) among Group IIb.1 CPKs appeared as a pointed protrusion at the base of the protein, while among Group IIb.2 CPKs this appeared globular or irregularly shaped. The predicted tertiary structures for Arabidopsis and rice in twelve different angles at 360 degrees rotation are shown in Appendix 30. The structures can also be viewed in all angles using the Pymol graphics system file (Appendix 31).
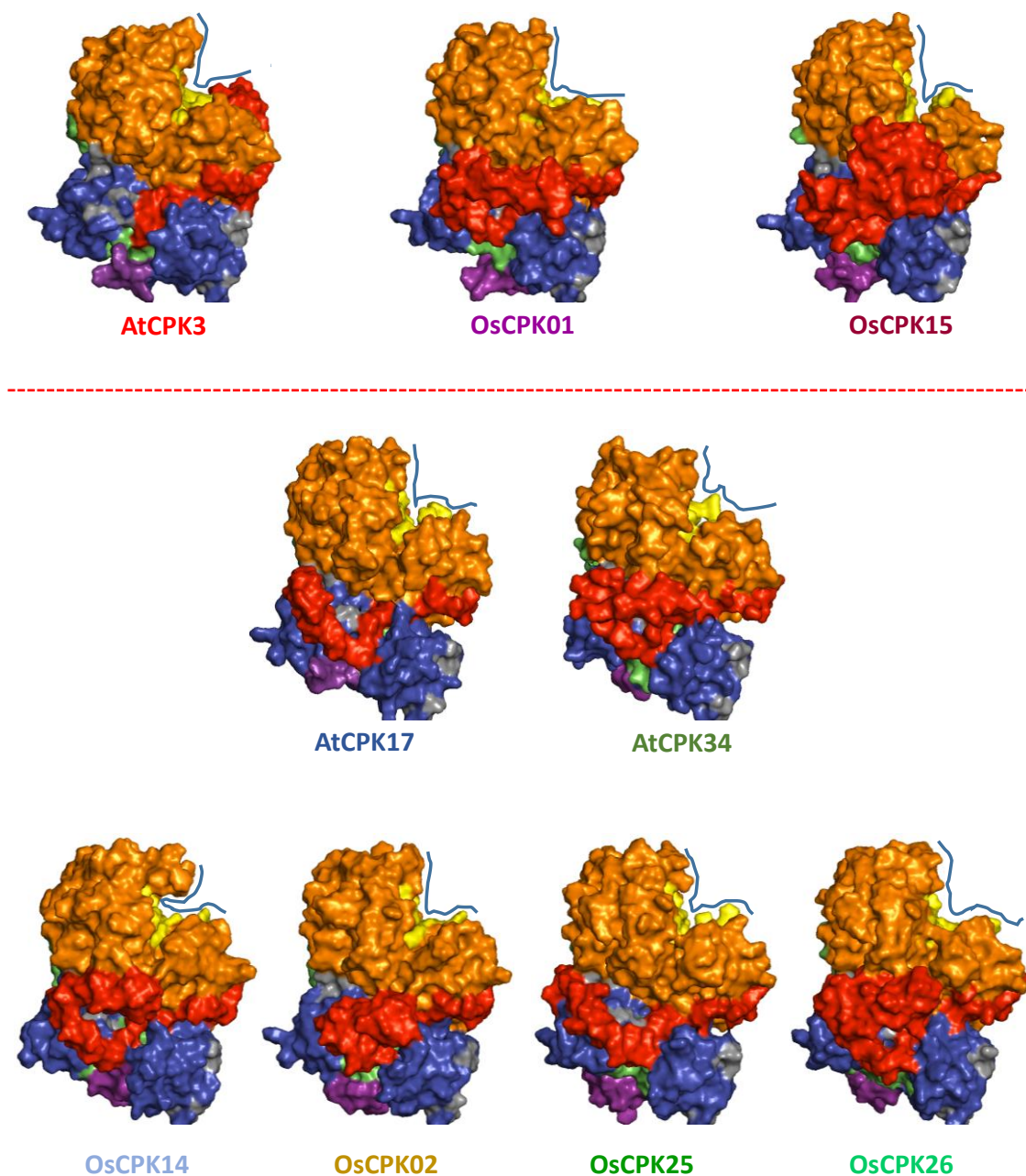
**Figure 5.4 Gross tertiary structures of Group IIb CPKs from Arabidopsis and rice.** CPK structures coloured based on domains: N-VD (red), PK (orange), AJ (green), CAD (blue) and CT (purple). Blue scribble line indicates the opening to the active sites. Phosphorylation active sites are coloured yellow while calcium binding sites in light grey shade. Group IIb.1 (top) and IIb.2 (bottom) CPKs are separated by the red broken line. Tertiary structures were predicted using I-TASSER (Roy et al. 2010) and figures were generated using Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.).

The tertiary structures were then analysed in terms of the motifs identified in the primary and secondary structure analysis. Distinction between Group IIb.1 and IIb.2 CPKs were also observed, but only in some. There were aa positions within the motifs that show subgroup specificity at the primary structure that matched with the secondary and tertiary levels, but there were some that did not exactly match and showed specificity in the surrounding regions. In certain cases, subgroup specificity was only observed when comparing group IIB CPKs within a species (At or Os CPKs only) and is lost in the combined analysis (all CPKs in At and Os). Each motif is described below.

Motif AtOs 1 (Figure 5.5, shaded dark grey) spans the N-terminal end of the CPK, from aa position 1 to 14. Within this motif, subgroup specificity in the primary structure was observed at aa positions 4, 5, 7, 13 and 14, although subgroup specificity in the secondary structure was observed only at aa positions 1 to 5. In Group IIb.1 CPKs, this position showed a helix-loop or turn-helix-loop structure, while in Group IIb.2 CPKs, this position is part of a long turn, or a series of turns separated by loops or β-sheets. At the tertiary level, no definite subgroup specificity was observed. The entire motif appeared as an irregular projection in between the big domains PK and the CAD in AtCPK3, OsCPK01, AtCPK17 and AtCPK34, which are members of either subgroup. In OsCPK15, 14, 2, 25 and 26, this motif also appears in between the two big domains but are embedded within the structure instead of projecting outward. However, if compared only within species, AtCPKs show subgroup specificity within this motif. In AtCPK3, this motif is located in the upper half of the protein, close to the substrate-binding regions of the PK domain; while in AtCPK 17 and 34, this motif is in the middle part (between PK and CAD). The gross structure appears similar between AtCPK17 and 34, but this motif is in the right side of AtCPK17 and in the left side of AtCPK34. No subgroup specificity in the tertiary level was observed in rice.

| Motif AtOs 1 | Structure level | amino acid position within the motif | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
| Group IIb.1 | | MG[HNA]R[HTA]S[KR][SH][KH][SR][SAQ][DAS]P[PED] | | | | | | | | | | | | | |
| | | M | G | HNA | R | HTA | S | KR | SH- | KH | SR | SAQ | DAS | P | PED |
| At 1 | primary | M | G | H | R | H | S | K | S | K | S | S | D | P | P |
| | secondary | helix | | helix | | loop | turn | | | | | | | loop | |
| Os 1 | primary | M | G | NA | R | TA | S | H- | H | R | AQ | AS | P | ED | |
| | secondary | turn/helix | | helix/loop | | | loop | turn | | turn/loop | | | | turn | |
| Group IIb.2 | | MG[NQ]CC[SPT][HGP][GS][RGS][DSK][SA]X[DPA]X | | | | | | | | | | | | | |
| | | M | G | NQ | C | C | SPT | HGP | GS | RGS | DK | SA | ADV | DPA | NADG |
| At 1 | primary | M | G | N | C | C | S | H | G | R | D | S | AD | D | N |
| | secondary | turn | | | | | | | | | | | loop/turn | loop/turn | loop |
| Os 1 | primary | M | G | NQ | C | C | PT | PG | GS | GS | K- | A- | V- | A- | G- |
| | secondary | turn | | turn/sheet | turn/loop/sheet | | turn/loop | turn | loop/turn | | turn/loop/sheet | | | | |
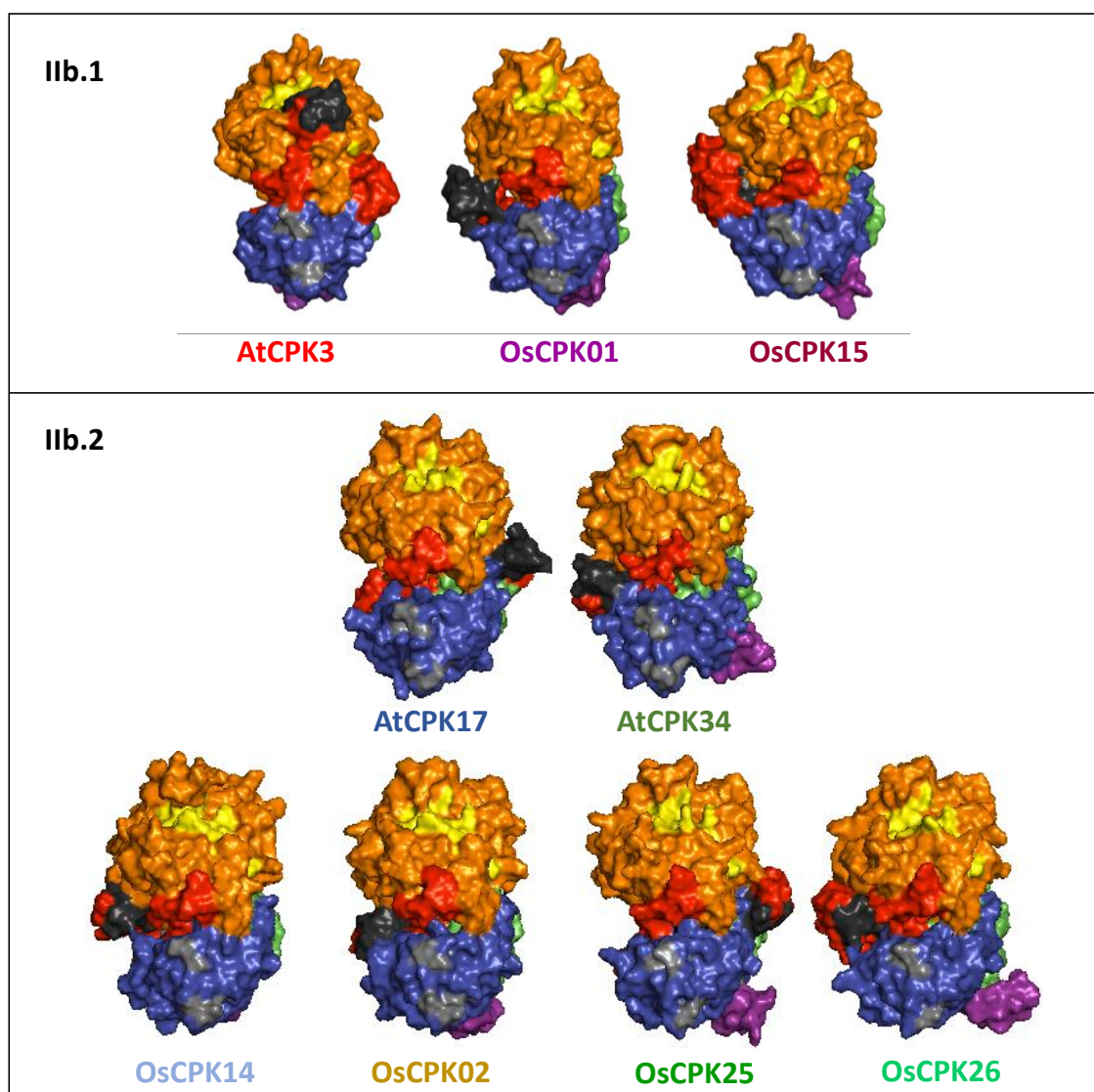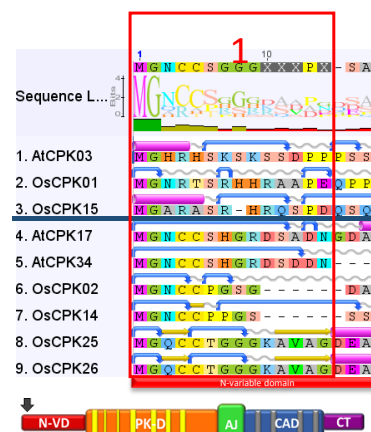




**Figure 5.5. Analysis of motif AtOs 1.** CPK structures coloured based on domains: N-VD (red), PK (orange), AJ (green), CAD (blue) and CT (purple). Phosphorylation active sites are coloured yellow while calcium binding sites in light grey shade. Tertiary structures were predicted using I-TASSER (Roy et al. 2010) and figures were generated using Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.). The motif specified is in dark grey shade.

Motif AtOs 2 (Figure 5.6, in dark grey shade) is comprised of about 20 to 30 aa located in the N-VD, approximately 12 aa upstream of the PK domain. Within this motif, subgroup specificity in the primary structure was observed at aa positions 8, 11, 12, 15, 17 and 19 to 22. Subgroup specificity in the secondary structure was observed at aa positions 1, 7 and 11 to 14. In Group IIb.1 CPKs, this motif showed a loop or a turn at position 1, a loop at position 7, and a loop or a loop-sheet at position 11 to 14. In contrast, in Group IIb.2 CPKs this position showed a turn, helix or sheet at position 1, a turn or helix at position 7, and a turn-sheet, turn-loop or helix-turn-loop at position 11 to 14. At the tertiary level, potential subgroup specificity was observed, although difference in structure was not highly defined. Motifs in both subgroups appeared at the lower hemisphere of the CPK, surrounding the CAD. In Group IIb.2 CPKs, the entire motif appears as a u-shaped structure while in Group IIb.1 CPKs the structure did not have a prominent shape. This was observed for both Arabidopsis and rice.

Motif AtOs 3 (Figure 5.7, in dark grey shade) is about 10 aa in length, located at the beginning of the PK domain. Within this motif, subgroup specificity in the primary structure was observed at aa positions 3 and 5, although subgroup specificity in the secondary structure was observed in adjacent aa positions 1 to 2 and 8 to 10. In Group IIb.1 CPKs, this motif showed a loop at position 1 to 2 and a turn-loop at position 8 to 10. On the other hand in Group IIb.2 CPKs, this motif showed a helix-loop, a sheet or a loop at positions 1 to 2 and a turn at position 10. At the tertiary level, some subgroup specificity was observed. The entire motif spans the lining of the cup-shaped active site of the PK domain. While the entire motif is better seen from the top view of the protein, there is no observable difference between the two subgroups at this angle. Potential subgroup difference is only manifested in a lateral angle as in Figure 5.7. In Group IIb.1 CPKs, the tip of this motif showed a small projection or some degree of projection towards the left; which is not evident among Group IIb CPKs.

| Motif AtOs 2 | Structure level | amino acid position within the motif | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
| **Group IIb.1** | | [SP][GR]X[VQ]X[SP]X[GP][SP][GPR]X(0-7)XX[TA]TA]XXXXXGR | | | | | | | | | | | | | | | | | | | | | |
| | | SP | GR | TPS | VQ | GQP | SP | STP | GP | SP | GPR | X(0-7) | RPQ | SAQ | TA | TA | SPA | TDE | QAD | QAG | NMV | G | R |
| **At 2** | primary | S | G | T | V | G | S | S | G | S | G | X(4) | R | S | T | T | S | T | Q | Q | N | G | R |
| | secondary | loop | | sheet | | loop | | | | | | | | | | | | | turn | | | | |
| **Os 2-3** | primary | P | R | PS | Q | QP | P | TP | P | P | P | X(0-7) | PQ | AQ | A | A | PA | DE | AD | AG | MV | G | R |
| | secondary | loop/turn | turn | turn/loop | | loop | | | loop/turn | loop | loop | loop/sheet | | helix | | | helix/turn | | | helix/sheet | | sheet | |
| **Group IIb.2** | | [AX][SA][KGT]X[APS][PA][PSA][SA][PA][PS][PAS][ASX][KTX]K[P-][QPA][GA]PIG[PE] | | | | | | | | | | | | | | | | | | | | | |
| | | QAKP | SA | KGT | HAEI | APS | PA | PSA | SA | PA | PS | - (7) | PAS | APSQ | TANK | K | P- | QPA | AG | P | I | G | PE |
| **At 2** | primary | AQ | S | K | H | A P | P | AP | S | P | P | - | P | A | T | K | - | Q | G | P | I | G | P |
| | secondary | turn/helix | turn | turn/loop | loop | loop/turn | turn | | loop | | | | turn | | sheet/turn | sheet | | loop | turn | sheet | | | |
| **Os 2-3** | primary | APK | A | TG | EIA | AS | AP | AS | AS | A | PS | - | AS | SPQ | KAN | K | P | PA | AG | P | I | G | PE |
| | secondary | sheet/helix | sheet/turn/helix | | sheet/loop/helix | | loop/helix | turn/helix | loop/helix | | | - | turn/loop/helix | | turn | | | loop | | sheet/loop | | sheet | |





**Figure 5.6. Analysis of motif AtOs 2.** CPK structures coloured based on domains: N-VD (red), PK (orange), AJ (green), CAD (blue) and CT (purple). Blue scribble line indicates the active site. Phosphorylation active sites are coloured yellow while calcium binding sites in light grey shade. Tertiary structures were predicted using I-TASSER (Roy et al. 2010) and figures were generated using Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.). The motif specified is in dark grey shade.

| Motif AtOs 3 | Structure level | amino acid position within the motif | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| Group IIb.1 | | Y[ET]FGRELGRG | | | | | | | | | |
| | | Y | ET | F | G | R | E | L | G | R | G |
| At 3 | primary | Y | E | F | G | R | E | L | G | R | G |
| | secondary | loop | | turn | | | loop | | turn | | loop |
| Os 4 | primary | Y | T | F | G | R | E | L | G | R | G |
| | secondary | loop | | sheet | turn | loop | | | turn | | loop |
| Group IIb.2 | | Y[ST][LIV]GKELGRG | | | | | | | | | |
| | | Y | ST | LIV | G | K | E | L | G | R | G |
| At 3 | primary | Y | ST | L | G | K | E | L | G | R | G |
| | secondary | helix | loop | | turn | loop/sheet | loop | | turn | | |
| Os 4 | primary | Y | ST | IV | G | K | E | L | G | R | G |
| | secondary | sheet/loop | | | | | loop | | turn | | |
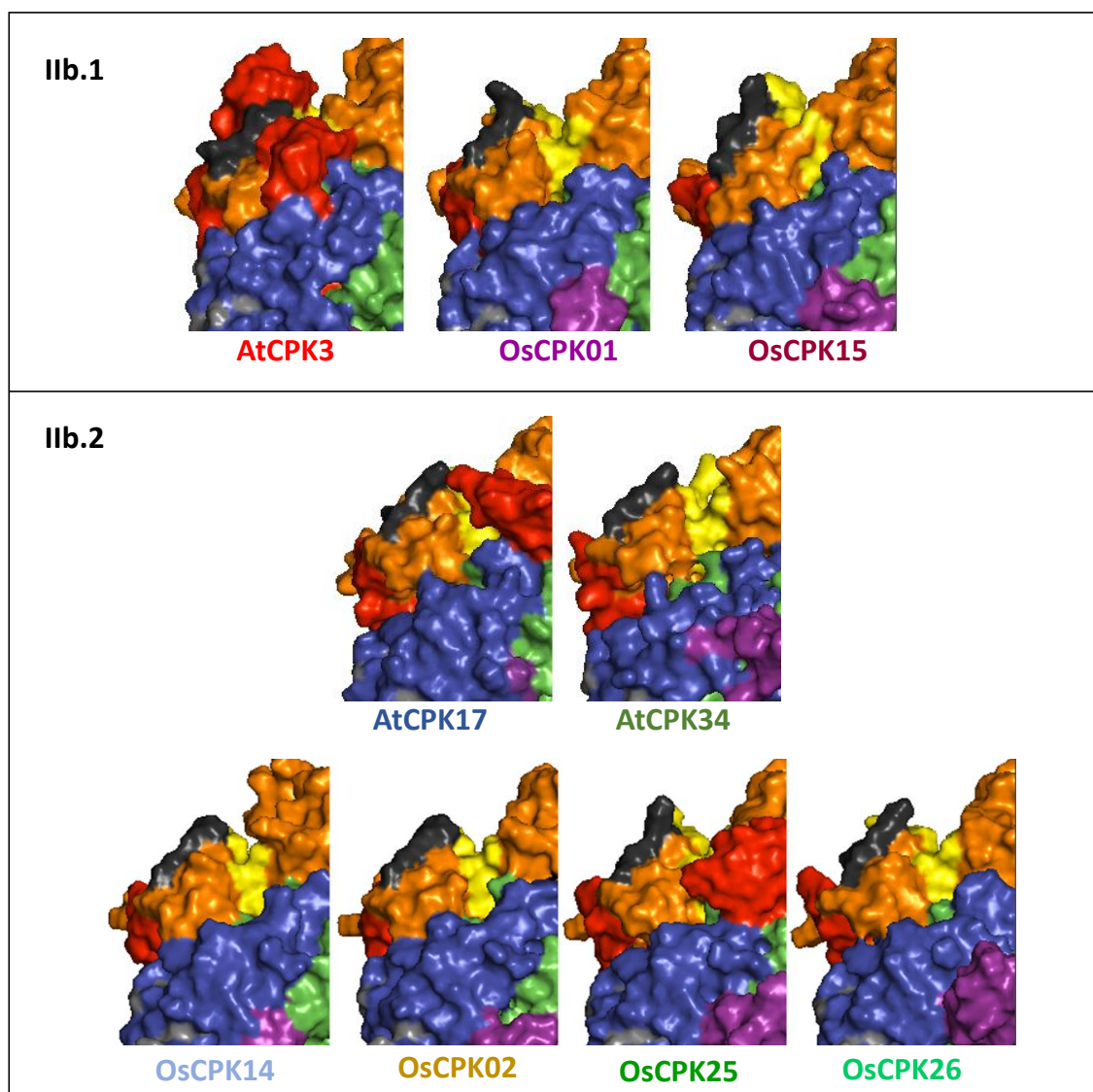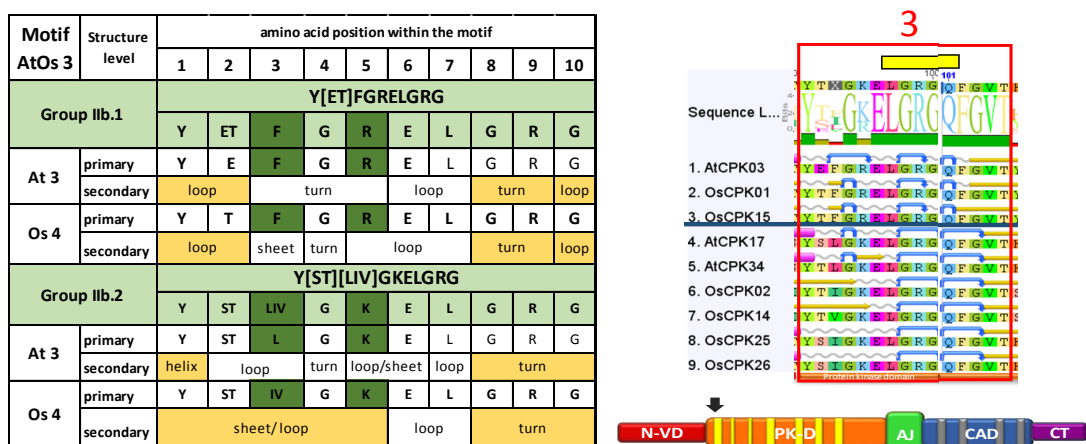


**Figure 5.7. Analysis of motif AtOs 3.** CPK structures coloured based on domains: N-VD (red), PK (orange), AJ (green), CAD (blue) and CT (purple). Blue scribble line indicates the active site. Phosphorylation active sites are coloured yellow while calcium binding sites in light grey shade. Tertiary structures were predicted using I-TASSER (Roy et al. 2010) and figures were generated using Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.). The motif specified is in dark grey shade.

Motif AtOs 4 (Figure 5.8, in dark grey shade) is a 26 aa long motif, near the beginning of the PK domain and spans three active sites, at a positions 4, 13 and 15 within the motif. Subgroup specificity in the primary structure was observed only at aa positions 1, 3, 16, 19, 24 and 26, while subgroup specificity in the secondary structure was observed almost at the entire length of the motif, from aa position 3 to 26. Both subgroups begin with a short sheet, but in Group IIb.1 CPKs, this position is followed by helices, sheet, turns and loops, while in Group IIb.2 CPKs, this position is followed by an entire long helix, or a turn followed by a long helix. Subgroup specificity was more obvious in Arabidopsis. AtCPK3 has a long helix followed by alternating sheet, turn and loop, whereas AtCPK17 and 34 only have a long turn and a long helix. On the other hand, subgroup specificity was not very obvious in rice. At the tertiary level, very little subgroup specificity was observed. The entire motif appeared as an irregular projection spanning one side of the PK domain. At the angle shown, in Group IIb.1 CPKs this motif is located more towards the left side of the protein, while in Group IIb.2 CPKs this motif is more towards the right side. In Arabidopsis, AtCPK3 stands out from AtCPK17 and 34 due to the presence of a portion of N-VD in the middle of this motif.

Motif AtOs 5 (Figure 5.9, in dark grey shade) is in the middle of the PK domain, 20 aa long and has one active site at position 11 (valine). Although relatively long, subgroup specificity in the primary structure was only observed at three aa positions within the motif: 7, 8, and 20. At the secondary structure level, subgroup specificity is shown in aa positions 1 to 2, 5, and 13 to 15. In Group IIb.1 CPKs, this motif showed a helix in aa positions 1 to 2, part of a loop in position 5, and a loop or sheet-loop-turn at position 13 to 15. On the other hand in Group IIb.2 CPKs, this motif showed a sheet at positions 1 to 2, a turn in position 5 and a helix or sheet-helix at positions 13 to 15. At the tertiary level, no definite subgroup specificity was observed. The entire motif appeared as a diagonal projection in between the big domains PK and the CAD, mostly located towards the interior of the protein and covered by the N-VD.

| Motif AtOs 4 | Structure level | amino acid position within the motif | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 |
| Group IIb.1 | | YL[VA]THK[EAP]T[KG][QKR][OR][VFY]ACKSI[PA][TA]R[RK]L[VA][HR][KRP]D | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | Y | L | VA | T | H | K | EA/P | T | KG | QK/R | OR | VF/Y | A | C | K | S | I | PA | TA | R | RK | L | VA | HR | KR/P | D |
| At 4 | primary | Y | L | V | T | H | K | E | T | K | Q | Q | V | A | C | K | S | I | P | T | R | R | L | V | H | K | D |
| | secondary | sheet | | | | helix | | | | | | | | sheet | | turn | | loop | | | turn | | sheet | | | loop | turn |
| Os 5 | primary | Y | L | VA | T | H | K | AP | T | G | KR | R | FY | A | C | K | S | I | A | TA | R | K | L | A | R | P | D |
| | secondary | sheet | | sheet/helix/t | helix/lo op | | | rn | | tur | | helix/turn | | | helix/sh eet | | | helix | | | | | helix/ sheet | helix/ loop | loop/ turn | loop/ turn | helix/ turn |
| Group IIb.2 | | [HS]LCT[HQ]KATGX[QKR][FL]ACKTIAKRKL[VSA][NT]KE | | | | | | | | | | | | | | | | | | | | | | | | | |
| | | HS | L | C | T | HQ | K | A | T | G | HQ/EL | QK/R | FL | A | C | K | T | I | A | K | R | K | L | VS/A | NT | K | E |
| At 4 | primary | H | L | C | T | Q | K | A | T | G | HL | Q | F | A | C | K | T | I | A | K | R | K | L | V | N | K | E |
| | secondary | sheet | | turn | | | | | | | helix | | | | | | | | | | | | | | | | |
| Os 5 | primary | HS | L | C | T | H | K | A | T | G | QE | KR | FL | A | C | K | T | I | AS | K | R | K | L | AS | TN | K | E |
| | secondary | sheet | | turn/helix | | | | | | | helix | | | | | | | | | | | | helix/loop | | | | helix |





IIb.1

AtCPK3    OsCPK01    OsCPK15

IIb.2

AtCPK17    AtCPK34

OsCPK14    OsCPK02    OsCPK25    OsCPK26

**Figure 5.8. Analysis of motif AtOs 4.** CPK structures coloured based on domains: N-VD (red), PK (orange), AJ (green), CAD (blue) and CT (purple). Blue scribble line indicates the active site. Phosphorylation active sites are coloured yellow while calcium binding sites in light grey shade. Tertiary structures were predicted using I-TASSER (Roy et al. 2010) and figures were generated using Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.). The motif specified is in dark grey shade.

| Motif AtOs 5 | Structure level | amino acid position within the motif | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 |
| Group IIb.1 | | MHHL[ST]GHRGAYEDR | | | | | | | | | | | | | | | | | | | |
| | | M | H | H | L | ST | G | H | R | N | I | V | DE | L | KR | G | A | Y | E | D | R |
| At 5 | primary | M | H | H | L | S | G | H | R | N | I | V | D | L | K | G | A | Y | E | D | R |
| | secondary | helix | | loop | | | turn | | loop | sheet | | | | | loop | | | helix | | | |
| Os 6 | primary | M | H | H | L | T | G | H | R | N | I | V | E | L | R | G | A | Y | E | D | R |
| | secondary | helix | | loop | | | | | loop | | sheet | | | | loop | turn | | helix | | | |
| Group IIb.2 | | M[HY]HL[TAS]GQPGAYEDK | | | | | | | | | | | | | | | | | | | |
| | | M | HY | H | L | TAS | G | Q | P | N | I | V | E | L | K | G | A | Y | E | D | K |
| At 5 | primary | M | H | H | L | T | G | Q | P | N | I | V | E | L | K | G | A | Y | E | D | K |
| | secondary | sheet | | loop | turn | | loop | | loop | sheet | | | | helix | | | | | | | |
| Os 6 | primary | M | HY | H | L | AS | G | Q | P | NG | IV | V | DE | L | KR | G | A | Y | E | D | K |
| | secondary | sheet | | sheet /loop | loo p | turn | loop | | loop/ turn | sheet | | | | helix/ sheet | | helix | | | | | |

Sequence L...

1. AtCPK03
2. OsCPK01
3. OsCPK15
4. AtCPK17
5. AtCPK34
6. OsCPK02
7. OsCPK14
8. OsCPK25
9. OsCPK26

N-VD   PK D   AJ   CAD   CT

**IIb.1**

AtCPK3    OsCPK01    OsCPK15

**IIb.2**

AtCPK17    AtCPK34

OsCPK14    OsCPK02    OsCPK25    OsCPK26
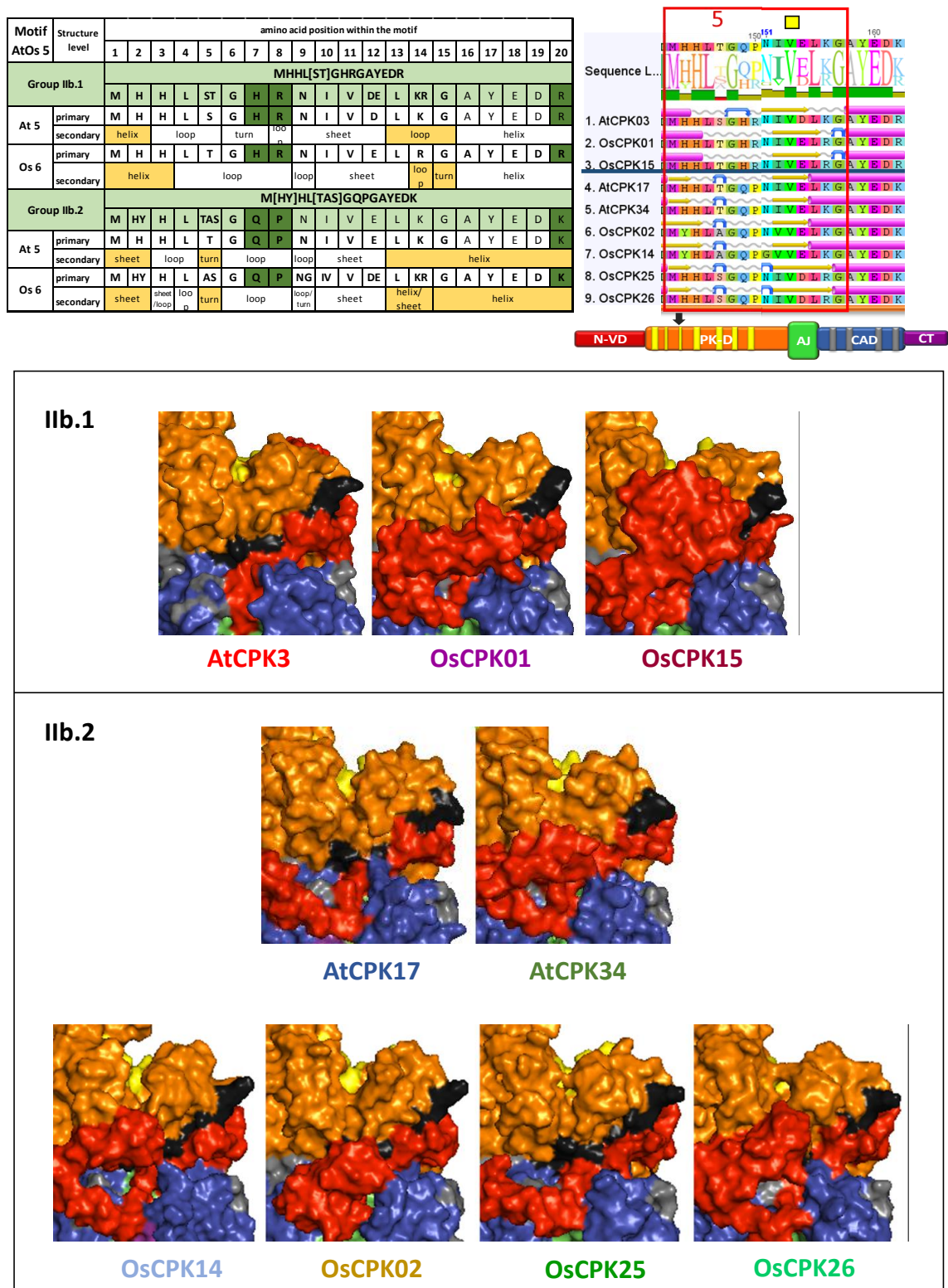
**Figure 5.9. Analysis of motif AtOs 5.** CPK structures coloured based on domains: N-VD (red), PK (orange), AJ (green), CAD (blue) and CT (purple). Blue scribble line indicates the active site. Phosphorylation active sites are coloured yellow while calcium binding sites in light grey shade. Tertiary structures were predicted using I-TASSER (Roy et al. 2010) and figures were generated using Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.). The motif specified is in dark grey shade.

Motif AtOs 6 (Figure 5.10, in dark grey shade) is also located in the middle of the PK domain, is comprised of 19 aa and spans part of an active site at position 19 (arginine). Within this motif, subgroup specificity in the primary structure was observed at aa positions 1, 3, 6, 7, 10 and 17, although subgroup specificity in the secondary structure was observed only at aa positions 10 to 19. The secondary structure in these positions showed good subgroup specificity. In Group IIb.1 CPKs, this position showed a helix-turn-helix or a helix-turn-sheet-helix structure, whereas in Group IIb.2 CPKs, this position starts with a long turn, followed by a long β-sheet. However, despite good specificity in secondary structure, no definite subgroup specificity was observed in the tertiary structure, as most of the motif was located in the interior of the middle part of the protein. Interestingly, although the motif is in the middle of the PK domain in the protein's primary structure, a portion of the motif is adjacent to an EF-hand in the protein's tertiary structure.

Motif AtOs 7 (Figure 5.11, in dark grey shade) is located just before the last active site in the PK domain. Subgroup specificity in the primary structure was observed only at aa positions 8 and 12, but subgroup specificity in the secondary structure was observed only at aa positions 6 to 9. There was only a low level of subgroup specificity in the secondary structure of this motif. Both Group IIb.1 and IIb.2 CPKs showed a turn-helix or a turn-loop-helix structure in these aa positions, but the helix was more upstream in Group IIb.2 CPKs. The helix starts at position 7 in Group IIb.2 whereas the helix starts at position 8 among Group IIb.1 CPKs. Similarly, at the tertiary level no definite subgroup specificity was observed. This motif appeared as an irregular region near the active site in the PK domain, but did not show any prominent shape difference between the two subgroups. In Arabidopsis, there was a small difference as AtCPK3 showed a small projection towards the periphery of the protein (encircled in red), but was not found in the rice CPKs OsCPK1 and 15.

| Motif AtOs 6 | Structure level | amino acid position within the motif | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| Group IIb.1 | | LCR[QE][MI]V[MAS]VVHSCHSMGV[MF]HRD | | | | | | | | | | | | | | | | | | |
| | | C | R | QE | MI | V | MAS | V | V | H | S | C | H | S | M | G | V | MF | H | R |
| At 7 | primary | C | R | Q | M | V | M | V | V | H | S | C | H | S | M | G | V | M | H | R |
| | secondary | helix | | | | | sheet | | | | helix | | | turn | | | helix | | | |
| Os 7 | primary | C | R | E | I | V | AS | V | V | H | S | C | H | S | M | G | V | FM | H | R |
| | secondary | helix | | | | | helix/ sheet | | | helix | helix/ turn | | turn | | | helix/turn | | helix | | |
| Group IIb.2 | | LLR[TA]IV[QEG]I[VI]HTCHSMGVIHRD | | | | | | | | | | | | | | | | | | |
| | | L | R | TA | I | V | QEG | I | VI | H | T | C | H | S | M | G | V | I | H | R |
| At 7 | primary | L | R | T | I | V | Q | I | VI | H | T | C | H | S | M | G | V | I | H | R |
| | secondary | helix | | | sheet | | | | | | turn | | | | sheet /turn | | sheet | | | |
| Os 7 | primary | L | R | TA | I | V | EG | I | VI | H | T | C | H | S | ML | G | V | I | H | R |
| | secondary | helix | | | helix/ sheet | | sheet | | | helix/ sheet | turn | | | | turn/ sheet | | sheet | | | |

Sequence L..

1. AtCPK03  AADLCRQMVVHSCHSMGVMHR
2. OsCPK01  AAALCREIVVVHSCHSMGVFHR
3. OsCPK15  AAALCREIVVVHSCHSMGVMHR
4. AtCPK17  AASLLRTIVIVHTCHSMGVIHR
5. AtCPK34  AASLLRTIVIIHTCHSMGVIHR
6. OsCPK02  AASLLRTIVIIHTCHSLGVIHR
7. OsCPK14  ASSLLRTIVIIHTCHSMGVIHR
8. OsCPK25  AAALLRAIVIVHTCHSMGVIHR
9. OsCPK26  AAALLRAIVIVHTCHSMGVIHR

N-VD  PK-D  AJ  CAD  CT

**IIb.1**

AtCPK3    OsCPK01    OsCPK15

**IIb.2**

AtCPK17    AtCPK34

OsCPK14    OsCPK02    OsCPK25    OsCPK26

**Figure 5.10. Analysis of motif AtOs 6.** CPK structures coloured based on domains: N-VD (red), PK (orange), AJ (green), CAD (blue) and CT (purple). Blue scribble line indicates the active site. Phosphorylation active sites are coloured yellow while calcium binding sites in light grey shade. Tertiary structures were predicted using I-TASSER (Roy et al. 2010) and figures were generated using Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.). The motif specified is in dark grey shade.

| Motif AtOs 7 | Structure level | amino acid position within the motif | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| Group IIb.1 | | VFFKPG[DE][KHQ]F[KR] | | | | | | | | | | | |
| | | V | F | F | K | P | G | [DE] | KHQ | F | KR | D | L |
| At 8 | primary | V | F | F | K | P | G | D | K | F | K | D | L |
| | secondary | sheet | | loop | | turn | | | helix | | sheet | | |
| Os 8 | primary | V | F | F | K | P | G | E | HQ | F | KR | D | L |
| | secondary | loop | turn | | loop | | turn | helix /loop | helix | | helix/ sheet | sheet | |
| Group IIb.2 | | VF[YF]K[PQE]GEVF[KR] | | | | | | | | | | | |
| | | V | F | YF | K | PQE | G | E | V | F | KR | D | I |
| At 8 | primary | V | F | Y | K | P | G | E | V | F | K | D | I |
| | secondary | sheet | turn | | loop | | turn | | helix | | sheet | | |
| Os 8 | primary | V | F | F | K | EQ | G | E | V | F | RK | D | I |
| | secondary | | helix | | | | helix /loop | | helix | | helix/sheet | sheet | |



**IIb.1**

AtCPK3    OsCPK01    OsCPK15

**IIb.2**

AtCPK17    AtCPK34

OsCPK14    OsCPK02    OsCPK25    OsCPK26
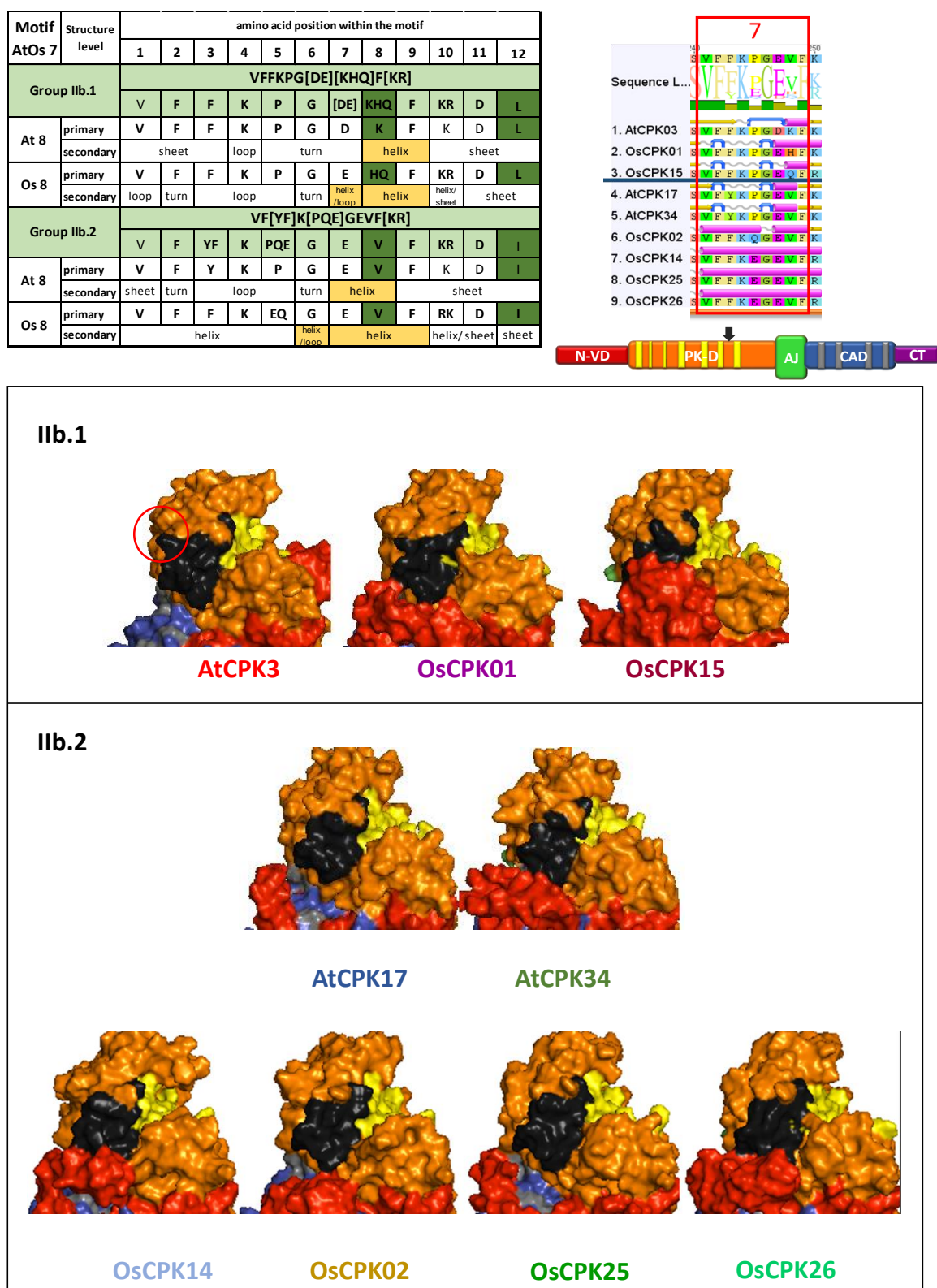
**Figure 5.11. Analysis of motif AtOs 7.** CPK structures coloured based on domains: N-VD (red), PK (orange), AJ (green), CAD (blue) and CT (purple). Blue scribble line indicates the active site. Phosphorylation active sites are coloured yellow while calcium binding sites in light grey shade. Tertiary structures were predicted using I-TASSER (Roy et al. 2010) and figures were generated using Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.). The motif specified is in dark grey shade.

Motif AtOs 8 (Figure 5.12, in dark grey shade) is located several aa after the last active site in the PK domain. Within this motif, subgroup specificity in the primary structure was only observed at aa positions 1 and 5 but the subgroup specificity in the secondary structure was observed along the entire length of the motif. In Group IIb.1 CPKs, it showed a turn followed by a long helix and ended with a turn, in both Arabidopsis and rice CPKs. In contrast, in Group IIb.2 CPKs, this motif showed a turn-helix-sheet-turn, turn-loop turn, At the tertiary level, no definite subgroup specificity was observed. The entire motif is located at the exterior of the protein towards the top of the PK domain, but doesn't directly span the active site.

Motif AtOs 9 (Figure 5.13, in dark grey shade) spans end part of the PK domain and includes the first three aa of the AJ domain. Subgroup specificity in the primary structure was observed at the ends of the motif, at aa positions 1, 2 and 20. However, subgroup specificity in the secondary structure was observed in the middle part of the motif, from position 6 to 14. In Group IIb.1 CPKs, this position comprised of a long helix (except for OsCPK01 which had a loop at position 14). In contrast, Group IIb.2 CPKs varied in their secondary structure. AtCPK17 and 34 showed a helix-sheet-helix-sheet structure while OsCPK02 and 14 showed a long sheet bounded by a short helix, turn or loop. On the other hand, OsCPK25 and 26 showed a long helix followed by a short sheet and then a loop. At the tertiary level, the motif appears as a bow-shaped structure in a slanted position, away from the active site of the PK domain but close to a calcium binding site of the CAD. The two subgroups differ in the general shape of this motif, as the ends of the structure were sharper and pointed among Group IIb.2 CPKs. Moreover, it is noticeable that in Group IIb.1 CPKs, the calcium-binding sites in the EF hand adjacent to this motif is more exposed and well-defined, compared to the Group IIb.2 CPKs.

| Motif AtOs 8 | Structure level | amino acid position within the motif | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| **Group IIb.1** | | TD | G | I | F | D | A | IV | L | QR | G | QH |
| At 9 | primary | T | G | I | F | D | A | I | L | Q | G | Q |
| | secondary | turn | helix | | | | | | | | turn | |
| Os* | primary | D | G | I | F | D | A | V | L | QR | G | H |
| | secondary | turn | helix | | | | | | | | turn | turn/loop |
| **Group IIb.2** | | NH | GA | I | F | NTA | AS | I | L | RS | G | QH |
| At 9 | primary | N | G | I | F | N | A | I | L | RS | G | QH |
| | secondary | turn | helix | | helix/sheet | | sheet | | | sheet/loop | turn | turn/sheet |
| Os* | primary | NH | GA | I | F | NTA | AS | I | L | R | G | QH |
| | secondary | turn/helix | turn/loop/helix | loop/helix | | loop/sheet/helix | loop/helix | | | turn/helix | turn | sheet/loop/turn |



**Figure 5.12. Analysis of motif AtOs 8.** CPK structures coloured based on domains: N-VD (red), PK (orange), AJ (green), CAD (blue) and CT (purple). Blue scribble line indicates the active site. Phosphorylation active sites are coloured yellow while calcium binding sites in light grey shade. Tertiary structures were predicted using I-TASSER (Roy et al. 2010) and figures were generated using Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.). The motif specified is in dark grey shade.

| Motif AtOs 9 | Structure level | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | amino acid position within the motif | | | | | | | | | | | | | | | |
| Group IIb.1 | | [KR][YQ]DPK[DE]RLT[AS]AE[VI]LNHPWIRED | | | | | | | | | | | | | | | | | | | | | |
| | | KR | YQ | D | P | K | DE | R | L | T | AS | A | E | VI | L | N | H | P | W | I | R | E | D |
| At 11 | primary | K | Y | D | P | K | D | R | L | T | A | A | E | V | L | N | H | P | W | I | R | E | D |
| | secondary | turn | loop | turn | turn | | helix | | | | | | | | | loop | | | | helix | | | turn |
| Os 9 | primary | R | G | D | P | K | E | R | L | T | AS | A | E | I | L | N | H | P | W | I | R | E | D |
| | secondary | helix | helix/loop | loop | | turn | helix | | | | | | | | helix/loop | loop | | turn | | loop | | | turn |
| Group IIb.2 | | [NH][SI][DN]PK[QEK]R[LI][TS][AS][AYF][QED]VLNHPWIKED | | | | | | | | | | | | | | | | | | | | | |
| | | NH | SI | DN | P | K | QEK | R | LI | TS | AS | AYF | QED | V | L | N | H | P | W | I | K | E | D |
| At 11 | primary | N | S | D | P | K | Q | R | LI | TS | AS | AYF | QED | V | L | N | H | P | W | I | K | E | D |
| | secondary | turn | | loop | | turn | helix | | sheet | | helix | | sheet | | | loop | | | turn | | loop | | turn |
| Os 9 | primary | NH | SI | DN | P | K | KRE | R | LI | ST | A | YF | QED | V | L | N | H | P | W | I | K | E | D |
| | secondary | turn/loop | | loop | | turn/loop | turn | helix/turn | | sheet /helix | | | sheet | turn/loop | | | loop | | | turn | helix/loop | loop/sheet | turn |
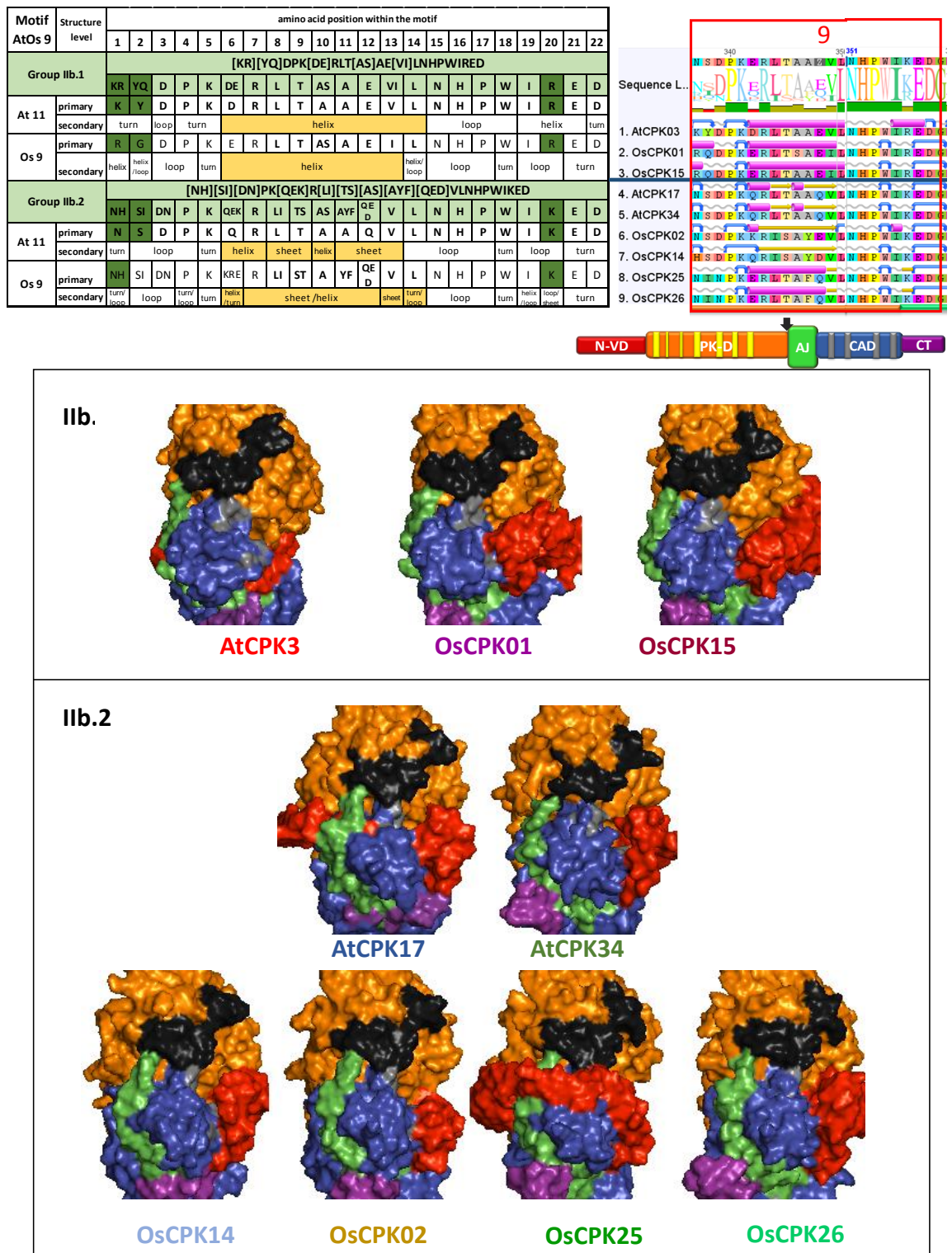


**Figure 5.13. Analysis of motif AtOs 9.** CPK structures coloured based on domains: N-VD (red), PK (orange), AJ (green), CAD (blue) and CT (purple). Blue scribble line indicates the active site. Phosphorylation active sites are coloured yellow while calcium binding sites in light grey shade. Tertiary structures were predicted using I-TASSER (Roy et al. 2010) and figures were generated using Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.). The motif specified is in dark grey shade.

Motif AtOs 10 (Figure 5.14, in dark grey shade) spans the end of the AJ and the first part of the CAD domain. Subgroup specificity in the primary structure of this motif was observed at aa positions 1, 2, 8, 12 and 13, and the subgroup specificity in the secondary structure was observed throughout the entire length of the motif. In Group IIb.1 CPKs, the whole motif is part of a long helix, while in Group IIb.2 CPKs, there was a loop at positions 12 to 14, within the long helix. Moreover, for AtCPK17 and 34, there was a sheet in positions 7 to 9. At the tertiary level, this motif appeared interior of the protein, with only a small region exposed at the surface of the protein and near to a calcium binding site. Despite the striking difference in secondary structures, no definite subgroup specificity was observed at the tertiary level of this motif.

Motif AtOs 11 (Figure 5.15, in dark grey shade) is located in between the first two EF hands in the CAD domain. Within this motif, subgroup specificity in the primary structure was observed at aa positions 1, 4, 6, 10 and 16. Subgroup specificity in the secondary structure was observed towards the ends of the motif; at aa positions 1 to 5 and 11 to 16. At aa positions 1 to 5, Group IIb.1 CPKs showed a turn-loop-turn or helix-turn structure, whereas Group IIb.2 CPKs mostly showed a long helix (except for OsCPK02 which showed a helix-turn-helix-loop-turn). Towards the end of the motif, both subgroup showed a long helix, although the beginning of the helix is more upstream among Group IIb.1. The helix started at positions 11 or 12 in Group IIb.1 CPKs and at positions 13 or 14 in Group IIb.2. At the tertiary level, this motif appears at the bottom part of the protein, adjacent to the CT domain. Some short finger-like projections were present within this motif. Subgroup specificity was observed at the tertiary level as the short projections were more prominent among Group IIb.2 CPKs than Group IIb.1 CPKs

| Motif AtOs 10 | Structure level | amino acid position within the motif | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| Group IIb.1 | | KLKK[MV]ALKV[VI]AENLS | | | | | | | | | | | | | | |
| | | K | L | K | K | M/V | A | L | K | V | VI | A | E | N | L | S |
| At 12 | primary | K | L | K | K | M | A | L | K | V | I | A | E | N | L | S |
| | secondary | helix | | | | | | | | | | | | | | |
| Os 11 | primary | K | L | K | K | V | A | L | K | V | V | A | E | N | L | S |
| | secondary | helix | | | | | | | | | | | | | | |
| Group IIb.2 | | [NQ]FKK[VA]ALR[VI]IAGCLS | | | | | | | | | | | | | | |
| | | NQ | F | K | K | VA | A | L | R | VI | I | A | G | C | L | S |
| At 12 | primary | N | F | K | K | V | A | L | R | V | I | A | G | C | L | S |
| | secondary | helix | | | | | sheet | | helix | | loop | | | helix | | |
| Os 11 | primary | Q | F | K | K | A | A | L | R | VI | I | A | G | C | L | S |
| | secondary | helix | | | | | | | | | | | loop | | sheet | |



**Figure 5.14. Analysis of motif AtOs 10.** CPK structures coloured based on domains: N-VD (red), PK (orange), AJ (green), CAD (blue) and CT (purple). Blue scribble line indicates the active site. Phosphorylation active sites are coloured yellow while calcium binding sites in light grey shade. Tertiary structures were predicted using I-TASSER (Roy et al. 2010) and figures were generated using Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.). The motif specified is in dark grey shade.

| Motif AtOs 11 | Structure level | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | amino acid position within the motif | | | | | | | | | | | | | | | |
| **Group IIb.1** | | [TSA]GLPKLG[ST]KISE[AS]E[IL]R | | | | | | | | | | | | | | | |
| | | TSA | G | L | P | K | L | G | ST | K | I | S | E | AS | E | IL | R |
| **At 14** | primary | T | G | L | P | K | L | G | S | K | I | S | E | A | E | I | R |
| | secondary | turn | helix | loop | | turn | loop | | | | | helix | | | | | |
| **Os \*** | primary | SA | G | L | P | K | L | G | T | K | I | S | E | S | E | IL | R |
| | secondary | turn/helix | loop/helix | | turn | loop | | | | | helix | | | | | | |
| **Group IIb.2** | | [QKN]GL[AS]K[QK]GT[KR][LF][ST][ED][YAN]E[VI][QE] | | | | | | | | | | | | | | | |
| | | QKN | G | L | AS | K | QK | G | T | KR | LF | ST | ED | YAN | E | VI | QE |
| **At 14** | primary | Q | G | L | A | K | Q | G | T | R | L | S | E | Y | E | V | Q |
| | secondary | helix | | | loop | | | | | | | helix | | | | | |
| **Os \*** | primary | KN | G | L | AS | K | QK | G | T | K | LF | TS | ED | AN | E | VI | QE |
| | secondary | helix | turn/helix | helix | loop/helix | turn/helix | turn/helix | loop/turn | loop | | loop/helix | | | helix | | | |
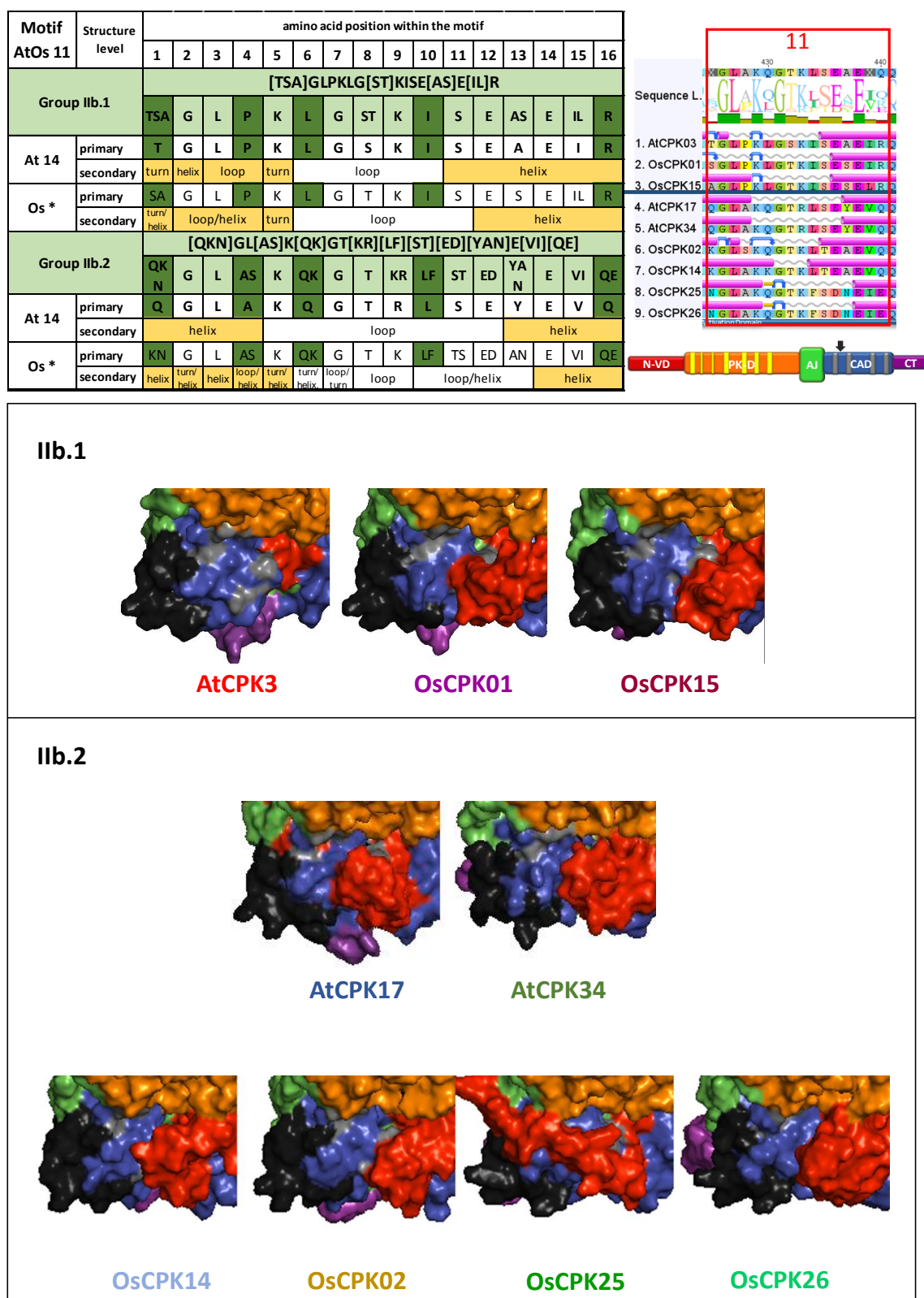


**Figure 5.15. Analysis of motif AtOs 11.** CPK structures coloured based on domains: N-VD (red), PK (orange), AJ (green), CAD (blue) and CT (purple). Blue scribble line indicates the active site. Phosphorylation active sites are coloured yellow while calcium binding sites in light grey shade. Tertiary structures were predicted using I-TASSER (Roy et al. 2010) and figures were generated using Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.). The motif specified is in dark grey shade.

Motif AtOs 12 (Figure 5.16, in dark grey shade) is located towards the beginning of the third EF hand in the CAD, with one calcium binding site at position 15. Within this motif, subgroup specificity in the primary structure was only observed at aa positions 2 and 5. Interestingly, the subgroup specificity in the secondary structure was observed at the other end of the motif, at aa positions 14 to 16. Both subgroups were comprised of a long helix, with sheets or turns at the end. The main difference between the two subgroups at the secondary structure level was the aa position where the turn begins. In Group IIb.1 CPKs, the turn begins at aa position 15 or 16, whereas in Group IIb.2 it was at aa position 14. This may or may not be related to the subgroup specificity at the tertiary level, which seemed to be well-defined. At the tertiary structure, this motif is more exposed at the surface of the protein among Group IIb.2 CPKs as compared with Group IIb.1 CPKs which have some portions in the interior of the protein.

Motif AtOs 13 (Figure 5.17, in dark grey shade) is located in between the last two EF hands, with one calcium-binding site at position 18. Within this motif, subgroup specificity in the primary structure was observed at aa positions 1 to 3, 5, 7 to 9, 15 and 19, although subgroup specificity in the secondary structure was observed along the entire length of the motif. In Group IIb.1 CPKs, this motif was comprised of a long helix, but in Group IIb.2 CPKs, there was a sheet at positions 12 to 16. Moreover, in Group IIb.2 AtCPKs, a turn-loop structure was present at positions 3 to 7. At the tertiary level, this motif appeared at the bottom of the protein, also adjacent to the CT domain. Some degree of subgroup specificity was also observed at the tertiary level, as the motif structure appeared to be wider among Group IIb.2 CPKs than Group IIb.1

| Motif AtOs 12 | Structure level | amino acid position within the motif | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
| **Group IIb.1** | | [IL]E[RK]ED | | | | | | | | | | | | | | | |
| | | IL | E | RK | E | D | H | LI | YL | KT | A | F | EQ | YF | F | D | KN |
| At * | primary | I | E | R | E | D | H | L | Y | T | A | F | Q | F | F | D | N |
| | secondary | helix | | | | | | | | | | sheet | | | turn | | |
| Os 12 | primary | L | E | K | E | D | H | I | YL | K | A | F | E | Y | F | D | K |
| | secondary | helix | | | | | | | | | | | | | | sheet/turn | |
| **Group IIb.2** | | [LM]D[RK]EE | | | | | | | | | | | | | | | |
| | | LM | D | RK | E | E | H | L | Y | ST | A | F | Q | HY | F | D | K |
| At * | primary | L | D | R | E | E | H | L | Y | S | A | F | Q | H | F | D | K |
| | secondary | loop | helix | | | | | | | | | | | | turn | | |
| Os 12 | primary | M | D | R | E | E | H | L | Y | T | A | F | Q | Y | F | D | K |
| | secondary | helix | | | | | | | | | | sheet | | turn | | | |



Sequence L...

1. AtCPK03
2. OsCPK01
3. OsCPK15
4. AtCPK17
5. AtCPK34
6. OsCPK02
7. OsCPK14
8. OsCPK25
9. OsCPK26

Calcium Activation Domain

N-VD  PK-D  AJ  CAD  CT



**IIb.1**

AtCPK3    OsCPK01    OsCPK15

**IIb.2**

AtCPK17    AtCPK34

OsCPK14    OsCPK02    OsCPK25    OsCPK26
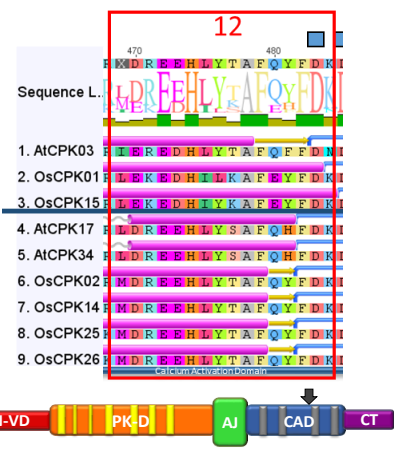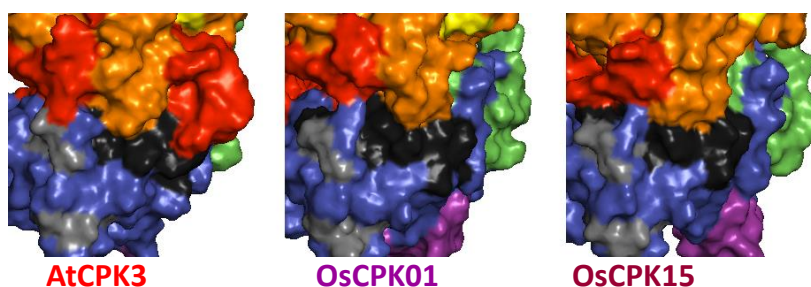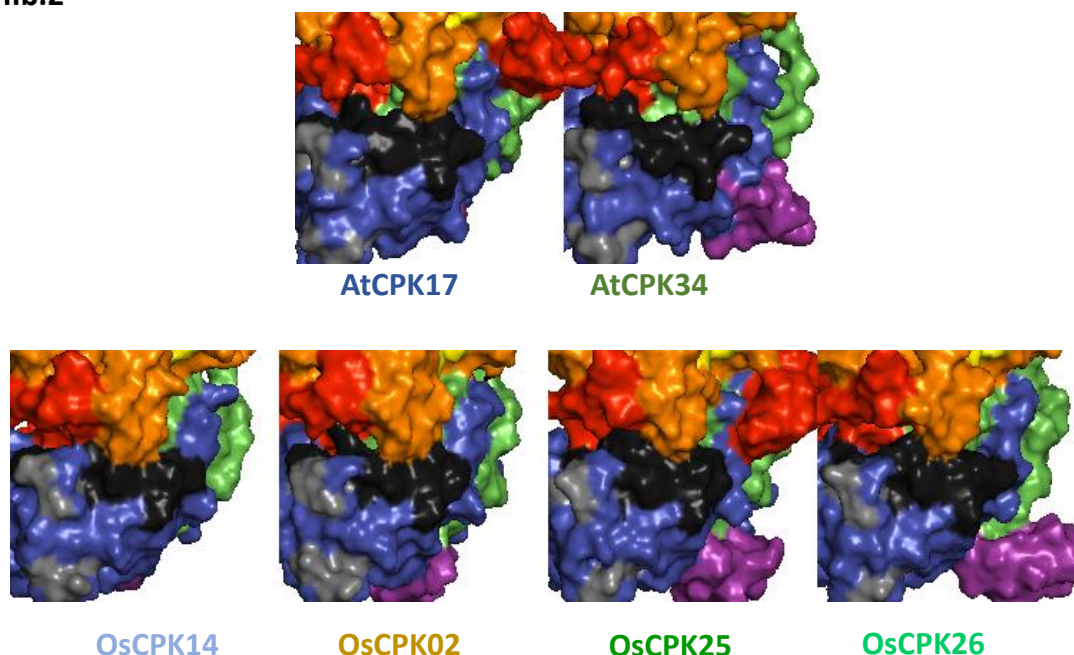
**Figure 5.16. Analysis of motif AtOs 12.** CPK structures coloured based on domains: N-VD (red), PK (orange), AJ (green), CAD (blue) and CT (purple). Blue scribble line indicates the active site. Phosphorylation active sites are coloured yellow while calcium binding sites in light grey shade. Tertiary structures were predicted using I-TASSER (Roy et al. 2010) and figures were generated using Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.). The motif specified is in dark grey shade.

233

| Motif AtOs 13 | Structure level | amino acid position within the motif | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 |
| Group IIb.1 | | KY[ND]MGD[DE][KA][ST]IKEIIAEVDT | | | | | | | | | | | | | | | | | | |
| | | K | Y | ND | M | G | D | DE | KA | ST | I | K | E | I | I | A | E | V | D | T |
| At 15 | primary | K | Y | N | M | G | D | D | K | S | I | K | E | I | I | A | E | V | D | T |
| | secondary | helix | | | | | | | | | | | | | | | | | | |
| Os 13 | primary | K | Y | D | M | G | D | DE | KA | T | I | K | E | I | I | A | E | V | D | T |
| | secondary | helix | | | | | | | | | | | | | | | | | | |
| Group IIb.2 | | E[FKQ]G[ML][NX]D[GA][RN][DE]IK[ED][IV]I[ST][ED][VA]D[GAS] | | | | | | | | | | | | | | | | | | |
| | | E | FK Q | G | ML | NY LM | D | GA | RN | DE | I | K | ED | IV | I | ST | ED | VA | D | GA S |
| At 1 | primary | E | F | G | M | N | D | G | R | D | I | K | E | I | I | S | E | V | D | G |
| | secondary | helix | | turn | | loop | | | helix | | | | sheet | | | | | turn | | |
| Os 1 | primary | E | KQ | G | L | YL M | D | GA | RN | ED | I | K | D | IV | I | ST | ED | VA | D | AS |
| | secondary | helix | | | | | | | | | helix/ sheet | | sheet | | | helix /loo | helix /turn | turn/ loop | | |



**Figure 5.17. Analysis of motif AtOs 13.** CPK structures coloured based on domains: N-VD (red), PK (orange), AJ (green), CAD (blue) and CT (purple). Blue scribble line indicates the active site. Phosphorylation active sites are coloured yellow while calcium binding sites in light grey shade. Tertiary structures were predicted using I-TASSER (Roy et al. 2010) and figures were generated using Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.). The motif specified is in dark grey shade.

234

Lastly, Motif AtOs 14 (Figure 5.18, in dark grey shade) spans the end part of the CT domain. Subgroup specificity in the primary and secondary structure was observed across the whole length of the protein. Group IIb.2 CPKs were two to three aa longer than Group IIb.1 CPKs in this motif. At the secondary structure, subgroup specificity was only observed within species. In Arabidopsis, all the Group IIb CPKs showed a sheet-loop-turn-helix-turn structure, but AtCPK17 and 34 (IIb.2) both showed longer helix than AtCPK3 (IIb.1). In rice, OsCPK1 and 15 (IIb.1) had different secondary structures but both had many turns and very little or no helix. In contrast, OsCPK2, 14, 25 and 26 generally showed a helix/loop-turn-helix-sheet structure. At the tertiary level, good subgroup specificity was also observed. The motif appeared at the bottom of protein, and is more projected/isolated from the main body of the protein among Group IIb.1 CPKs and more attached to the protein among Group IIb.2 CPKs.

| Motif AtOs 14 | Structure level | amino acid position within the motif | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | * | * | * | 9 | 10 | 11 |
| **Group IIb.1** | | E[LI][VA]PNR(3)M[F-] | | | | | | | | | | |
| | | E | LI | VA | P | N | R(3) | M | F- | | | |
| At 16 | primary | E | L | V | P | N | R(3) | M | | | | |
| | secondary | loop | sheet | | loop | turn | helix-turn | turn | | | | |
| Os 14 | primary | E | I | AV | P | N | R(3) | M | F | | | |
| | secondary | turn | loop/sheet | | turn | helix/turn- | sheet/turn | sheet | | | | |
| **Group IIb.2** | | [DEA][PAT][NI]PK(2-3)R(2)[ED][LV][SV][LFI][K-] | | | | | | | | | | |
| | | DEA | PAT | NI | P | K(2-3) | R(2) | ED | LV | SV | LFI | K- |
| At 16 | primary | D | P | I | P | K(2) | R(2) | E | L | S | F | K |
| | secondary | loop | | | sheet/loop | turn-helix | helix | | | | | turn |
| Os 14 | primary | EA | AT | NI | P | K(2-3) | R(2) | D | VL | V | LI | |
| | secondary | loop/helix | | | turn/helix | turn-helix | helix-sheet | sheet | | | | |
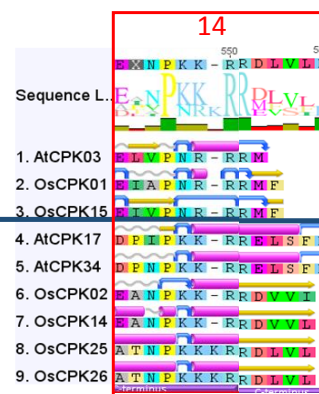


**IIb.1**
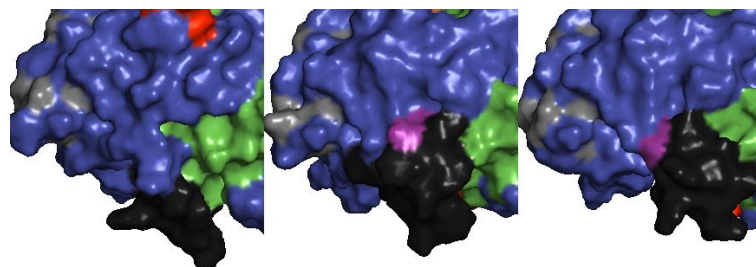
AtCPK3   OsCPK01   OsCPK15

**IIb.2**
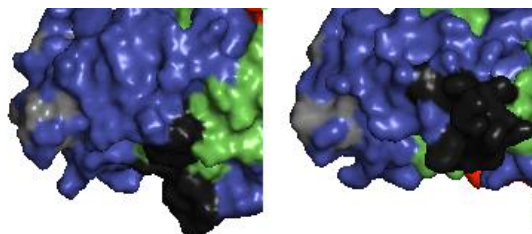
AtCPK17   AtCPK34
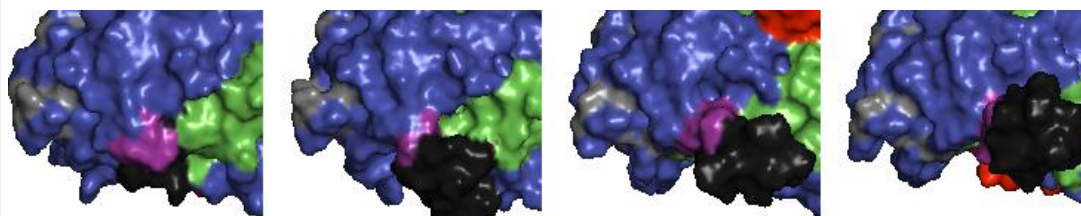
OsCPK14   OsCPK02   OsCPK25   OsCPK26

**Figure 5.18. Analysis of motif AtOs 14.** CPK structures coloured based on domains: N-VD (red), PK (orange), AJ (green), CAD (blue) and CT (purple). Blue scribble line indicates the active site. Phosphorylation active sites are coloured yellow while calcium binding sites in light grey shade. Tertiary structures were predicted using I-TASSER (Roy et al. 2010) and figures were generated using Pymol (The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.). The motif specified is in dark grey shade.

### 5.3.2 Do the gene regulatory regions of CPKs contribute to functional specificity?

The location of AtCPK3, 17 and 34 within the Arabidopsis genome was determined using MapViewer (Entrez MapViewer, NCBI) and Gene (Entrez Gene, NCBI) as shown in Figure 5.19 and 5.20. The upstream flanking regions were extracted, which encompassed the chromosomal region between the transcription start site and the preceding gene (Table 5.1). The AtCPK3 gene is located on chromosome 4 between positions 12324758 and 12327459, while its upstream region adjacent to the neighbouring gene was at position 12327693 to 12327460 (Table 5.1 and Figure 5.20). AtCPK17 and 34 are located on chromosome 5 located at positions 3937024 to 3939596 and 6521716 to 6524119, respectively, and the upstream regions were at positions 3935130 to 3937024 and 6526717 to 6524120 respectively (Table 5.1 and Figure 5.20). The direction of transcription for AtCPK3 and 34 are both upwards, while the direction of transcription for AtCPK17 is downward.



**Figure 5.19 Chromosome location of AtCPK3, 17 and 34**

**Table 5.1 Upstream flanking sequences of AtCPK3, 17 and 34**

| Gene | Upstream flanking sequence | Size (bp) | Locus |
|---|---|---|---|
| AtCPK3 | TTTCCTTTTTCAATCTCAAATTGCAAAAGATTAACATTGAATTAGCCGTGTGATTGATTGGAGAAGTGGAAACAACTAATATGCAGAAGATTTCCA CATCCACTAATTTATCTTCCAATGGGACAAATACTTTTGTTTATACGCGGTTCAACTACGTACCAATAAAAACACGCCACGTGGTCATCGTGTCT GATCTCAGACATTAAAAAGCCATAAGAATCGGAGATACAATCT | 233 bp | Chr4 12327693 to 12327460 |
| AtCPK17 | AACATAATCTTGTCTTCAAGAGTTGTTACCCTTTAAACATATCACTACAAATCGTGTGCTAAACCATTTAAAAGACAAATCAAAACAAAATAAGAAG AAAATAGCCTGCAACAATCAAAGTTACTTCATCAAAATTGTTAAATAGTAAAAGACTAAGCAAATCCTACAAACCACTGAGTTTTCTTAAACCATTT GATTCTACTTGTGGCAGGTCTCGAAGACACCACAAGTTGAAGGAAGACTAATAATATTTAGCTAATAAAGGTCATCTTTAACTAAACAGATTGAAT CACTATTGAATTCCGTTACCAAATTCTTCTACCGGTGCTTTTAGTGTAACCCCCAGACCAAATATCATCATCTCCAATAGAATCTACTCTTTCCCATT TGTTACTCTCAACTCTCAAGATTCATCTTAAAGATGTAAAACAAATGTTTCTCTTCAACTAGCTTTTCTAATAAACTCAAGATGATCAAAACCTCTTC TGTTTTCTGAATCACTATTCTCCTCTTCCAAATGTACTCCCACTCTTTTGAAACGTACTCAAACGGATGATCCCAATATGGGGTTTCTTCGACTTCT TCTTTGGGAAAATCTCCAGAGAAATCGACAATCTTGATGTGATCATCAGAAGTATACAGATAAAGCTTGCTGTTCTTATAAGCCAGATCCCAAAAA CCCAAATTTGCGCCTCTTGTATACCAATTCAACCAACAATCATCTCCTTTCTTGTGTGTGAACAACTCAAGCTTATTGAAAATTCAAGCAACGACAT AATCTCCTGTTCTTTCGTTTTATCCACAGAACAGCTGCTCTTTCGCTTCCGAAGATATCTTTCGACACACGTTCCTTATGATAAGACTCAAGGAAAT AAGCTCACTCATCACCACCGCTTCGTTTGAATCTCACTTGTCCACCGCGTATTGACGTCTCCATTGATGGAAGATTGATTCTTTCGCCCGGTAAA ATGTTAAAGATAGAAAAGTCGAGATAAGAATCTACCATCAATAGACAGTTTCCGAAGCTAGCCATTATATAATAAGTCTTGTCAGAGATTCCAAA GAGTTTCGCGTTGTAAATCTTGTCTTCTCCGGGATCGAACAACATCGAAGATTCATTATGGTAATATGAATTGCCATGGATACAGAGGCCGCTTC ACGCATGTTTTCCAAACGGAGAACCACTCTAAGCAAACGTTTTTTATACGGTGAAAATCTATAGTGCTCAAGCTTTCGATTATCGGACGTAAAAGA TTGGACCAATCTTGATTTTCAGACATTTTATTCACTCATCAAATGTCTGCGCTTACATCACACTAATGGGAGCTATAGGGTTAGGGTTTTAGGGTT TACAACCAACCCTGTACATTTCCCTTTTTGATGTTCTGCTTTTCACTTTGATAATTAACTGTATTTTTATATTTTACTTTCTTCCTAAATACAATAAATT TCTTAACACATTCAAAATCCCATGAATTAGAAAATTGGTTTTCACATGTTTCAGATTGATGAATGAAAATTCTCATACACCTGATTACGTAGTTTCTT ATGAGATACATTTATAATCATATTTATATAAGATTTAGGCAAAGGTTCTAATCTTGATTTTTATGAAGTAGTTGTGGAATATAGAAGATAAGAAGAGA CAGTTATGTAACAAAAATTAAAATCCCAGAAATTCCTCTCAACCTTTGTCTTAAATCGCCTAAATGTTGTTTTTACTTTAAATACCAAAATTACAGACA CGTTTTGGTCCTAAACTTTCTCCCTATCAACAAAACCCTATAATCCTTTGAACCGAATACCAACGAAGAAAAAAAAACATTCTATTCATAGCCATTCA CAACCAAAACACCTAACGTTTTCTTGCTAAAGAAGAAAACAAAAACCAGAGGAAGAA | 1894 bp | Chr5 3935130 to 3937024 |
| AtCPk34 | CATATTTATTCAGTCTTCTTTGAGTGGTTAACTTGTAAATTTCTCTCACACTTGGACTCGTTACCCAAAAGAAACCCATAATGTTTAGAGTGACCGGA ACTCTCTCGGCGGCATCATCTCCAGCGGTGGCGGCGGCGTCTTTCTCCGCCGCTCTAAGACTTTCTACAAACCACTACAAACCACTGCTGCCTCT CCTCCCCCATCTCCGTTGGTTCTCGAAGTTTTCTCGTCAATTTCTCGGCGGACGCATCTCTTCACTCCGACCACGCATTCCTTCTCCGTGCCCGA TTCGACTCTCTGGTTTCCCGGCTCTCAAAATGAGAGGTACAACAAAACATGAACTGTATCAAAGACTTGACTGCGAGAGCATCTTCTCAGATTTC TAGTAGGAACGAATTTACCTTTACTCTGATCTGATTTTTCTTCAAAATTCATGTGCGATTACCGTTGCTCTGTGAACTGTGTATATCTGAATTATTG ATTATTTAGTTTCTGATTGGGTTGATTTCTGAATTATCTTTTGTTTTTTCAGCTTCGTTTAGCAGTGGGAGTAGTGGAAGTAGTGCAAGTAGAGAGA TTCTGGTGCAGCATTTGCTTGTTAAGAATAATGATGTTGAGCTCTTTGCCGAACTCCAAAAGAAATTCTTGGATGGTCTGATTCTTGAATCCTTTT TTACTCGTTTGATATTTCAACATACATAGTAGCTAGGCATGTAAGCTTTTATAGTTGTTTCTGCCATGAGCTTTGTCGCTGTGGAGCTGAACTTAT ATTATGTGTCGCAATGTCTATAGTCAATGGCGTCTCGCTTACTCGATTTTAATTGGTTTTGGGGTTTGGTTAGCACATATCACTGACACATAT ATGAATTTTTCATTACTCGACTGTCACTTTGTTTTGTTTGTATTAGAAGATTAAGGCTGTCTACTGTTTTTTCAAGTAACTGTTTCTTGTGTAGGAG AGGAGATGAGTGATCTTGCGGCTGAGTATTCGATTTGTCCTTCCAAAAAGGACGGTGGTATACTTGGATGGGTAAAACTGGGTCAAATGGTATG CTCTCTCTTTTTATAATATTCTCTATTTATTTTGGTTATATTGTTGCATAGACACCTTCAGATCTATAGGACCCTTGCTGTCTTAGCCTTGATATTTCA CAAACAATTGGGTGATGTAACAGTCCCTGGTGAATAAAGTTTCAGTATTGACATCCATGTGCAAATGGAAGTAGTAGGATGTAACCAATGATAAA GTAATTCCAAAATGTTTTGGACTTTGGCAAGTTCCTGCTGCGATATGAACATGGCAGACCAATTAAAATTGTTTTCAGGTACCAGAGTTTGAGGA AGCTGCGTTTAAAGCAGAGCTGAATCAGGTGGTGAGGTGTAGAACCCAATTTGGCTTGCACTTATTGCAAGTTCTATCTGAGAGGTAATTCTTC TGAAAACTATCCTCCGCGTTAGCGACATTTACTATCTGACTGTTGTAAATTTCTTGACGAAAAGTTTTATATTAACAGAGAACCGGTGAAAGATA TCCAGGTAGAGGAGCTGCATTCCAAAATGCAAGATCCAGTTTTCATGGATGAGGCTCAGTTAATTGATGTCAGAGAACCCAATGAGATGTAAGT TCCTTTCATTTCTCATCTTGACATACTTTTCCTTCTTCAATGAAGCCCAATGGATTAGCTTCATACTTTCTCAACATTATGATTCATGTGTATCTTG CTGAGTTTGATTGCTCTGCCATTTAGATTCATGTGTACCTTGCACATTTTTATCTTAGAAGGAGGACAAAGTATCCATAGTCAACTCACATTTGCT TTCTTCTATTCTGTTTCACTACTTGTTGTACCAGAGAGATAGCATCCTTGCCAGGATTCAAAGTGTTCCCACTCCGCCAATTTGGAACATGGGCA CCAGACATCACTTCAAAGCTGAACCCTGAGAAGGATACATTTGTCTTGGTATGTATATCAGCCTCTTCTCAACTGTTAGATATAGGTATACTGTAT CTAGCCTAGCACAATTTAGAGGTGTTGTAACCAGAGAGATAGCATCCTTGCCAGGATTCAAAGTGTTCCCACTCCGCCAATTTGGAACATGGGCA TATTTATCTGGTTAGGAGTTGAGAGCCTCTCATAAGAAATCTAAGATCATCTGAAACAAAGGAACCTCTGATGTTGAAATGACTTTTTAAAGTTCT TCTCTACTAAGTCGCGGCCTTTGTTACAGTGCAAGGTTGGTGGCAGGTCAATGCAAGGTTGCCAACTGGTTACAGTCTCAGGTAAAAAGCTTTCC TGCTCCTTTATAAGTATATATATATATATATATATATATATACATATGCAATTATGATATCTTCATGGCTTGTTCTCTTGTTTCCTCCAGGGCTTCAA GAGTGTGTACAACATTACTGGTGGGATACAAGCTTATTCTCTCAAGGTTGACCCATCAATTCCTACTTACTGAAATTGAGCATCTTAGTCTCTTTT TCTTGTATCCTCCTAATTCCAAATTAGATTTGCAAATTCGTATTCAATAAACTATATAATATAAGGTTCATTTTTGAGAATATAGATAACAAATTTAG AATATT | 2587 bp | Chr5 6526707 to 6524120 |

237

**Figure 5.20 Upstream flanking regions and neighbouring genes of (a) AtCPK3 and (b) AtCPK17 and 34.** Figures generated and modified from NCBI Map viewer http://www.ncbi.nlm.nih.gov/projects/mapview/ and NCBI Gene http://www.ncbi.nlm.nih.gov/gene/ on April 5, 2013. Regions were coloured as follows: purple box, upstream neighbouring gene; red box, upstream region taken for promoter analysis; light green box with black arrowheads, 5' or 3' UTR; green box with white arrowheads, exons; green line with grey arrowheads, introns.

Transcription factor binding site analysis of AtCPK 3, 17 and 34 using Mat Inspector as described in section 5.2.3 resulted in an extensive list of binding sites and their sequence similarity (Appendix 32). TF binding sites that are shared by all group IIb AtCPKs (Group IIb.1 and IIb.2), those that are unique to AtCPK3 (Group IIb.1) and uniquely shared by AtCPK17 and 34 (Group IIb.2) are summarised in Table 5.2. Some TF binding sites common to all group IIb CPKs are involved in housekeeping, metabolic, cellular division processes in plants such as the CCAAT-box (Laloum et al. 2013), M phase-specific activators (Ito et al. 2001), nodulins (Denancé et al. 2014), and GATA factors (Reyes et al. 2004), as well as in development such as the homeodomain protein WUSCHEL (Leibfried et al. 2005), plant specific floral meristem identity LEAFY (Engelhorn et al. 2014), SBF-1 (Lawton et al. 1991), and squamosa promoter binding protein (Preston and Hileman 2013). Some are involved in stress response such as ER stress response element I (ERSE) (Liu et al. 2007), and jasmonate- and elicitor-responsive elements (Menke et al. 1999) . CCAAT-boxes were also found to form a complex with CONSTANS domains which promote flowering in Arabidopsis (Wenkel et al. 2006). Nodulins and nodulin-like proteins have also been found to function in plant development and pathogen response (Denancé et al. 2014).

TF binding sites that are unique to AtCPK3, which belong to the stress-responsive CPK group IIb.1, include ABA inducible transcriptional activator, NAC domain containing protein 92 (ATNAC2/6) and bZIP transcription factor which are all involved in abiotic and biotic stress responses (Table 5.2). Similarity to a developmental TF binding site was also detected, which is the rice transcription activator-1 (RITA) that is highly expressed during seed development in rice (Izawa et al. 1994). There were only four TF binding sites that are uniquely shared by AtCPK17 and 34, which belong to the developmental CPK group IIb.2. These include sites involved in development such as the anther-specific myb gene, protodermal factor 2, in storage such as GCN4 and STK, in flavonoid and carotenoid synthesis such as RAP2. Interestingly, many TF binding sites involved in myb transcription factor genes were also detected, which controls various processes such as abiotic and biotic stress, development, metabolism and defence.

It can be noted that in the stress-responsive Group IIb.1 CPKs most of the TF binding sites identified were involved in stress, but there were also some TF binding sites involved in development and metabolism. Likewise, in the developmental Group IIb.2 CPKs, the TF-binding sites detected include those involved in development, but also those involved in stress.

**Table 5.2 Shared and unique transcription factor binding sites detected in AtCPK3, 17 and 34 upstream flanking regions determined using Mat Inspector Analysis (Genomatix Software Inc. http://www.genomatix.de/)**

| Shared by AtCPK3, 17 and 34 (IIb.1 and IIb.2) |
|---|
| CA-rich element |
| CCAAT-box in plant promoters |
| Class I GATA factors |
| DNA-binding protein of sweet potato that binds to the SP8a (ACTGTGTA) and SP8b (TACTATT) sequences of sporamin and beta-amylase genes |
| Dof3 - single zinc finger transcription factor |
| ERSE I (ER stress-response element I)-like motif |
| Ethylene-responsive elements (ERE) and jasmonate- and elicitor-responsive elements (JERE) |
| Homeodomain protein WUSCHEL |
| M-phase-specific activators (NtmybA1, NtmybA2, NtmybB) |
| Nodulin consensus sequence 1 |
| Plant specific floral meristem identity gene LEAFY (LFY) |
| SBF-1 |
| Squamosa promoter binding protein-like 14 |
| **Unique to AtCPK3 (IIb.1 only)** |
| ABA (abscisic acid) inducible transcriptional activator |
| Arabidopsis NAC domain containing protein 92 (ATNAC2/ATNAC6) |
| Rice transcription activator-1 (RITA), basic leucin zipper protein, highly expressed during seed development |
| Tobacco bZip transcription activator (TAF-1) |
| **Shared by AtCPK 17 and 34 only (IIb.2 only)** |
| ABA insensitive protein 4 (ABI4) |
| Anther-specific myb gene from tobacco |
| Arabidopsis 6b-interacting protein 1-like 1 |
| Arabidopsis leucine zipper protein TGA1 |
| Arabidopsis thaliana class A heat shock factor 1a |
| CAACTC regulatory elements, GA-inducible |
| GAAA motif involved in pollen specific transcriptional activation |
| GA-regulated myb gene from barley |
| GCN4, conserved in cereal seed storage protein gene promoters, similar to yeast GCN4 and vertebrate AP-1 |
| GT1-Box binding factors with a trihelix DNA-binding domain |
| ICE (inducer of CBF expression 1), AtMYC2 (rd22BP1) |
| Myb domain protein 96 (MYBCOV1) |
| Myb-domain transcription factor werewolf |
| Myb-like protein of Petunia hybrida |
| P1BS, PHR1 binding sequences |
| Protodermal factor 2 |
| Putative cis-acting element in various PAL and 4CL gene promoters |
| RAP2.2, involved in carotenoid and tocopherol biosynthesis and in the expression of photosynthesis-related genes |
| Recognition site for BZIP transcription factors that belong to the group of Opaque-2 like proteins |
| Ribosomal protein box, appears unique to plant RP genes and genes associated with gene expression |
| Sequence motif from the promoters of different sugar-responsive genes |
| Storekeeper (STK), plant specific DNA binding protein important for tuber-specific and sucrose-inducible gene expression |
| Sunflower homeodomain leucine-zipper protein Hahb-4 |
| TEIL (tobacco EIN3-like) |
| Trihelix DNA-binding factor GT-3a |
| Zea mays MYB-related protein 1 (transfer cell specific) |

The TF binding site analysis was also performed in rice CPKs OsCPK1, 15, 2, 14, 25 and 26 (Table 5.3). Transcription factor binding site analysis of these genes also resulted in an extensive list of binding sites and their sequence similarity (Appendix 32). Sites that are shared by all Group IIb OsCPKs, those that are uniquely shared by the OsCPK1 and 15 and uniquely shared by OsCPK2, 14, 25 and 26 are noted in Table 5.3. Some TF binding sites that match those that were detected in AtCPKs, while some are unique to OsCPKs, such as TF binding sites on OsCPKs that match those detected in AtCPKs were shaded as follows: green, if shared with all group IIb AtCPKs; pink, if shared with all group IIb.1 AtCPKs (stress-responsive); and blue, if shared with all group IIb.2 (developmental).

Similar to the AtCPK analysis, TF binding sites involved in developmental and stress-response were detected in OsCPKs in all cases; whether shared by all group IIb OsCPKs, uniquely shared by Group IIb.1 OsCPKs, or by Group IIb.2 OsCPKs. It can be noted that many of those identified in the developmental Group IIb.1 AtCPKs (Table 5.2) were identified as shared by all Group IIb OsCPKs (Table 5.3, shaded in blue). Likewise, number of TF binding sites shared by all Group IIb AtCPKs were identified as uniquely shared by Group IIb.1 OsCPKs or Group IIb.2 CPKs. There are also TF binding sites identified in Group IIb.1 AtCPKs that were identified in Group IIb.2 OsCPKs and vice versa.

In combining the two TF binding analyses, there are sites that were common to all Group IIb AtCPKs and OsCPKs. These are Class I GATA factors, sweet potato DNA-binding proteins to sporamin and beta-amylase genes (SPF1), and SBF-1. GATA factors are known to be broadly distributed in eukaryotes and are involved in light-dependent and nitrate-dependent control of transcription (Reyes et al. 2004). SPF1 was reported to bind to 5' upstream sequences of various genes in tubers that encode for sporamin and beta-amylase (Ishiguro and Nakamura 1994). Sporamin has multiple biological functions related to stress in sweet potato (Senthilkumar and Yeh 2012). Beta-amylase, an enzyme that breaks down starch and other polysaccharides in to maltose, was reported be involved in abiotic stress responses such as temperature stress (Kaplan and Guy 2004). SBF-1 was reported as a factor that represses the transcription of a bean defense gene, chalcone synthase 15 (Lawton et al. 1991).

On the other hand, the TF binding sites that were common to all Group IIb.1 AtCPKs and OsCPKs include ABA inducible transcriptional activator, RITA, and ATNAC proteins while those that were common to all Group IIb.2 AtCPKs and OsCPKs include the GT1-Box binding factors and Trihelix DNA-binding factor GT-3a. ABA inducible transcription factors function in response to ABA and related responses, mainly in late and adaptive responsive during stress in plants (Nakashima and Yamaguchi-Shinozaki 2010). RITA-1 is a basic leucine zipper (bZIP) transcriptional activator which is highly expressed in aleurone and endosperm cells of developing rice seeds (Izawa et al. 1994). AtNAC2 was reported to be involved in ethylene and auxin signalling pathways during salt stress response and lateral root development in Arabidopsis(He et al. 2005). In rice, a member of the NAC family of transcription factors OsNAP was reported to confer abiotic stress response through ABA signalling (Chen et al. 2014).

All of the three transcription factors described above appear to function mainly in stress responses and to some extent in development. Conversely, the transcription factors that were uniquely common to all Group IIb.2 AtCPKs and OsCPKs mostly function in pollen and floral tissue, and to some extent in stress. GT-1 was reported to interact with a pollen-specific promoter in tobacco (Hochstenbach et al. 1996) while GT-3a was reported to be a distinct group of GT-factors that are predominantly expressed in floral buds and roots (Ayadi et al. 2004). GT factors were first known to be upregulated in light response, but are also known to be involved in other functions such as pollen development and salt stress (Kaplan-Levy et al. 2013).

In summary, there are a number of TF binding sites that show similarity between Group IIb.1 and IIb.2 CPKs, which may be involved in the overlapping and redundant functions between CPKs. There were a few TF binding sites identified that were unique to Group IIb.1, which may be important promoter regions that contribute to CPK functional specificity to stress. Likewise, there were also a few TF binding sites that were unique to Group IIb.2, which may be important promoter regions that contribute to CPK functional specificity to development.

**Table 5.3 Shared and unique transcription factor binding sites detected in OsCPK1, 15, 2, 14, 25 and 26 upstream flanking regions determined using Mat Inspector Analysis (Genomatix Software Inc. http://www.genomatix.de/).** Rows that match Group IIb AtCPK transcription factor binding sites have been highlighted: pink, if common with Group IIb.1 AtCPKs; blue, if common with Group IIb.2 AtCPKs, and green, if common with all Group IIb AtCPKs

| Shared by all Group IIb OsCPKs (IIb.1 and IIb.2) |
|---|
| Anther-specific myb gene from tobacco |
| Arabidopsis NAC domain containing protein 19 |
| Arabidopsis thaliana signal-responsive gene1, Ca2+/ calmodulin binding protein homolog to NtER1 (tobacco early ethylene-responsive gene) |
| AS1/AS2 repressor complex binding motif II |
| Brassinazole-resistant 1 |
| Circadian clock associated 1 |
| Cis-element in the GAPDH promoters conferring light inducibility |
| Class I GATA factors |
| DNA-binding protein of sweet potato that binds to the SP8a (ACTGTGTA) and SP8b (TACTATT) sequences of sporamin and beta-amylase genes (SPF-1) |
| Drosophila initiator motifs |
| Evening element |
| Flowering locus C |
| GAAA motif involved in pollen specific transcriptional activation |
| GA-regulated myb gene from barley |
| HDZip class I protein ATHB5 |
| Heat shock element |
| Homeodomain protein WUSCHEL |
| ICE (inducer of CBF expression 1), AtMYC2 (rd22BP1) |
| L1-specific homeodomain protein ATML1 (A. thaliana meristem layer 1) |
| Myb domain protein 96 (MYBCOV1) |
| Myb-like protein of Petunia hybrida |
| Nodulin consensus sequence 1 |
| Nodulin consensus sequence 2 |
| RAP2.2, involved in carotenoid and tocopherol biosynthesis and in the expression of photosynthesis-related genes |
| S1F, site 1 binding factor of spinach rps1 promoter |
| SBF-1 |
| Sequence motif from the promoters of different sugar-responsive genes |
| Storekeeper (STK), plant specific DNA binding protein important for tuber-specific and sucrose-inducible gene expression |
| Sunflower homeodomain leucine-zipper protein Hahb-4 |
| TEIL (tobacco EIN3-like) |
| Transcriptional repressor BELLRINGER |
| Wheat bZIP transcription factor HBP1B (histone gene binding protein 1b) |
| **Shared by OsCPK1 and 15 only (IIb.1 only)** |
| ABA (abscisic acid) inducible transcriptional activator |
| AC-type motifs, MYB46/MYB83-responsive elements |
| AGL15, Arabidopsis MADS-domain protein AGAMOUS-like 15 |
| Arabidopsis leucine zipper protein TGA1 |
| Avian C-type LTR TATA box |
| Cellular and viral TATA box elements |
| Cis-element involved in SA (salicylic acid) induction of secretion-related genes via NPR1 |
| Floral homeotic protein APETALA1 |
| HBP-1a, suggested to be involved in the cell cycle-dependent expression |
| High mobility group I/Y-like protein isolated from pea |
| Homeodomain glabrous 9 |
| LIM domain protein binding to a PAL-box like sequence |
| MADS-box protein SQUAMOSA |
| Maize C1 myb-domain protein |
| Mammalian C-type LTR TATA box |
| Oryza sativa bZIP protein 8 |
| PBF (MPBF) |
| Phosphate starvation response 1 |
| Phytochrome interacting factor3-like 5 |
| Plant TATA box |
| Prolamin box, conserved in cereal seed storage protein gene promoters |
| Promoter elements involved in MgProto (Mg-protoporphyrin IX) and light-mediated induction |
| Protodermal factor 2 |
| Recognition site for BZIP transcription factors that belong to the group of Opaque-2 like proteins |
| Rice iron-related transcription factor 2 |
| Rice transcription activator-1 (RITA), basic leucin zipper protein, highly expressed during seed development |
| RY and Sph motifs conserved in seed-specific promoters |

| |
|---|
| TATA-binding protein, general transcription factor that interacts with other factors to form the preinitiation complex at promoters |
| TEF cis acting elements in both RNA polymerase II-dependent promoters and rDNA spacer sequences |
| Tobacco bHLH transcription factor MYC2 |
| Transcription factor of rice and barley binding to the iron deficiency-responsive cis-acting element 2 (IDE2) |
| ABA response elements |
| Arabidopsis NAC domain containing protein 92 (ATNAC2/ATNAC6) |
| Auxin Response Element |
| Coupling element 3 (CE3), non-ACGT ABRE |
| C-repeat/dehydration response element |
| Dof1 / MNB1a - single zinc finger transcription factor |
| Dof2 - single zinc finger transcription factor |
| Dof3 - single zinc finger transcription factor |
| ERSE I (ER stress-response element I)-like motif |
| High mobility group I/Y-like proteins |
| Maize activator P of flavonoid biosynthetic genes |
| MYB protein from wheat |
| Myb-domain transcription factor werewolf |
| Nodulin consensus sequence 3 |
| **shared by OsCPK14, 2, 25 and 26 (IIb.2 only)** |
| Agamous, required for normal flower development, similarity to SRF (human) and MCM (yeast) proteins |
| AS1/AS2 repressor complex binding motif I |
| bZIP transcription factor from Antirrhinum majus |
| CCAAT-box in plant promoters |
| E2F class I sites |
| GT1-Box binding factors with a trihelix DNA-binding domain |
| M-phase-specific activators (NtmybA1, NtmybA2, NtmybB) |
| R2R3-type myb-like transcription factor (I-type binding site) |
| Rice MYB proteins with single DNA binding domains, binding to the amylase element (TATCCA) |
| Secondary wall NAC binding elements |
| Squamosa promoter binding protein-like 14 |
| Transcription factor NAC2 |
| Trihelix DNA-binding factor GT-3a |
| WRINKLED 1 |
| ABORTED MICROSPORES |

### 5.3.3 Does the tissue localisation of Group IIb CPKs contribute to functional specificity?

*5.3.3.1 Development of AtCPK34 overexpressor lines*

The full AtCPK34 gene was amplified using RT-PCR from an Arabidopsis flower RNA using AtCPK34 gene specific primers that were flanked with Gateway attB sequence and *Pfx* polymerase, as described in section 5.2.4. Unlike AtCPK3 (Chapter 4), the initial attempt to amplify AtCPK34 was successful. A PCR product of the expected size (1.6 Kb) was achieved (Figure 5. 21a). After purification using Diffinity RapidTip (Sigma-Aldrich, St. Louis, MO, USA), some of the lower molecular weight products and smearing was reduced, but not all the primer-dimers were removed.
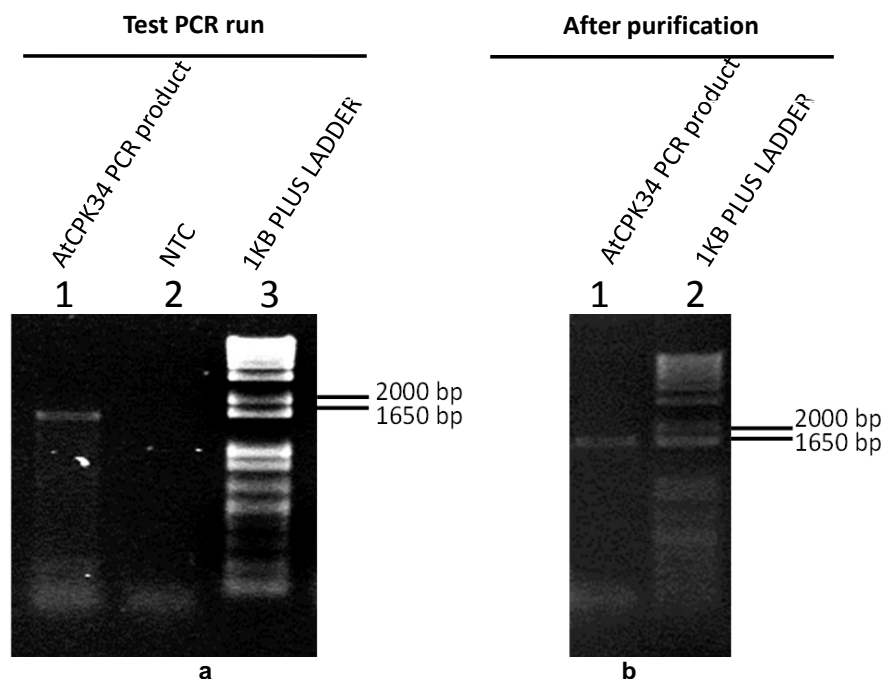
**Figure 5.21 AtCPK34 RT-PCR products before purification and after purification.** **(a) Before amplification.** Lane 1: AtCPK34 RT-PCR product amplified from Arabidopsis flower RNA; Lane 2: NTC; Lane 3: 1 Kb Plus DNA ladder (Invitrogen); **(b) After purification.** Lane 1: AtCPK34 PCR product after purification using Diffinity RapidTip (Sigma-Aldrich, St. Louis, MO, USA); Lane 2: 1 Kb Plus DNA ladder (Invitrogen).

A BP reaction to clone AtCPK34 PCR products into the entry clone pDONR resulted in about 3.88 x $10^5$ CFU/µg insert DNA. The insert: vector molar ratio used was 1:1 as suggested in the manufacturer's protocol. A total of 16 colonies were screened by colony PCR using AtCPK34-specific primers. Seven colonies appeared to contain an insert approximately the size of AtCPK34 (1569 bp) (Figure 5.22). Three colonies were randomly selected for plasmid isolation and sequencing, which were all verified to have the correct AtCPK34 sequence.
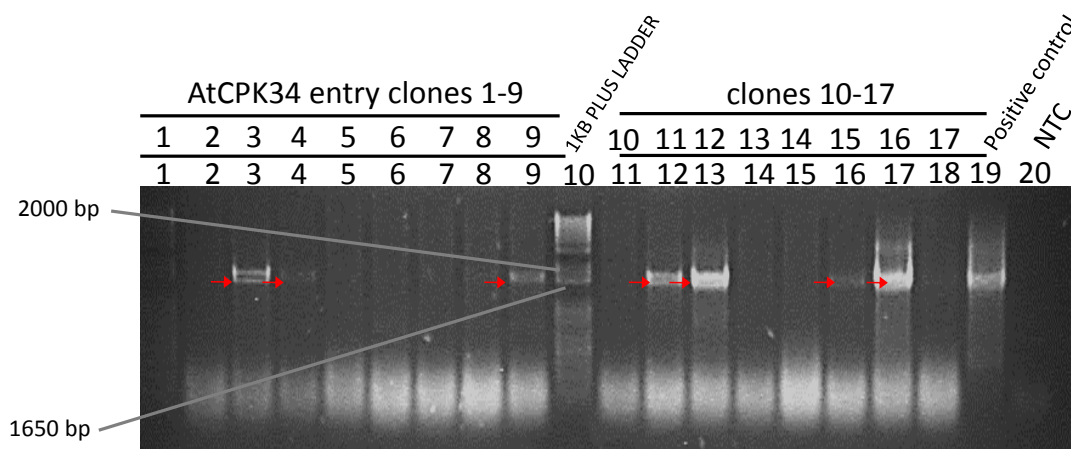


**Figure 5.22 Colony PCR of AtCPK34 entry clones.** Lanes 1 to 9: AtCPK34 entry clones 1-9; Lane 10: 1 Kb DNA Ladder (Invitrogen); Lanes 11 to 18: AtCPK34 entry clones 10 – 17; Lane 19: AtCPK34 positive control flower cDNA; Lane 20 NTC. Positive colonies are on lanes 3, 4, 9, 12, 13, 16 and 17 as indicated by the red arrows.

The LR recombination reaction successfully transferred the AtCPK34 coding sequence from the entry clone (pDONR-AtCPK34) to the destination vector (pHEX2). This cloning procedure resulted in 4.1 x 10$^6$ CFU/μg donor DNA. Fifteen colonies screened by colony PCR all appeared to contain an insert of the correct size (Figure 5.23). These constructs were named pHEX2-AtCPK34.1 to .15. All three pHEX2-AtCPK34 constructs that were selected for plasmid isolation and sequencing contained correct AtCPK34 sequence in the right orientation.
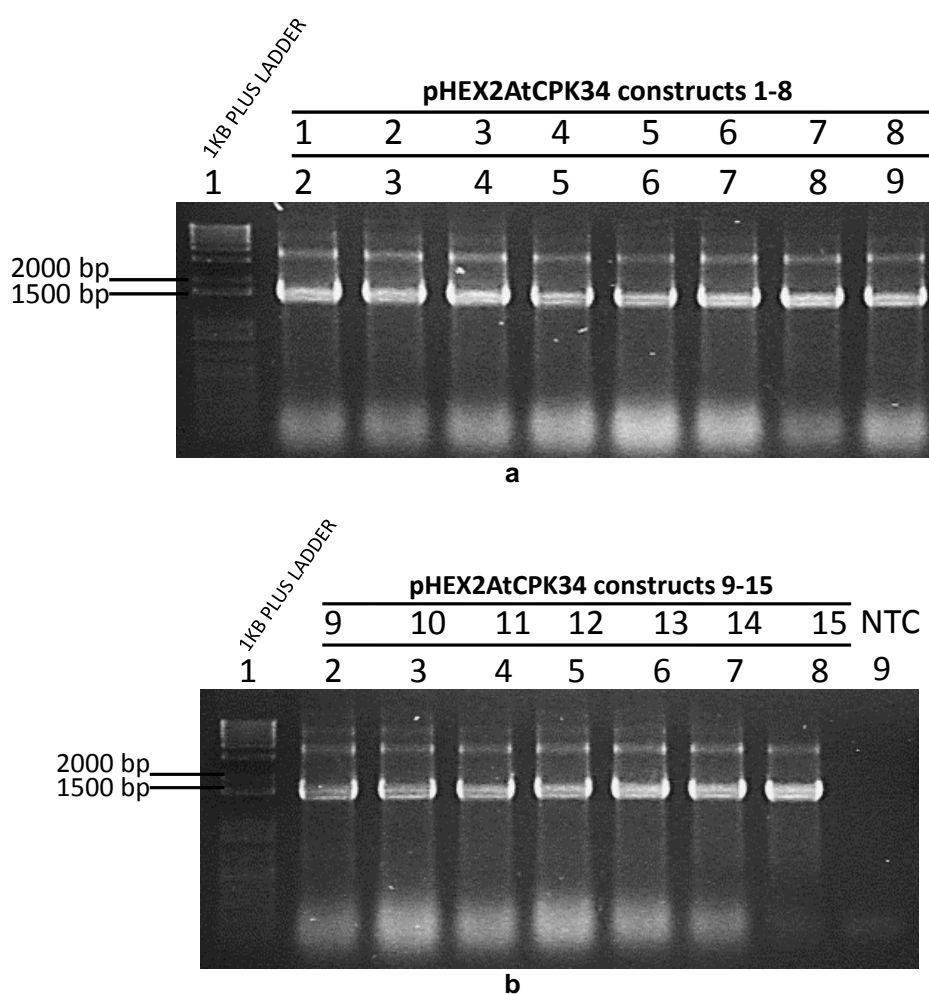


**Figure 5.23 Colony PCR of AtCPK34 destination clones.** All colonies tested were positive as shown on lanes 2 to 9 and 12 to 18. Lane 19 shows the negative control

One construct was randomly selected (pHEX2-AtCPK34.1 construct) and transformed into Agrobacterium GV3101. This was done by electroporation and resulted in 5.66 x 10$^6$ CFU/μg plasmid construct. Fifteen colonies were screened by colony PCR, and seven colonies showed to contain the construct (Figure 5.24). Three positive Agrobacterium colonies were selected for plasmid extraction and re-cloning into *E. coli* for sequencing. All constructs had the correct AtCPK34 sequence.
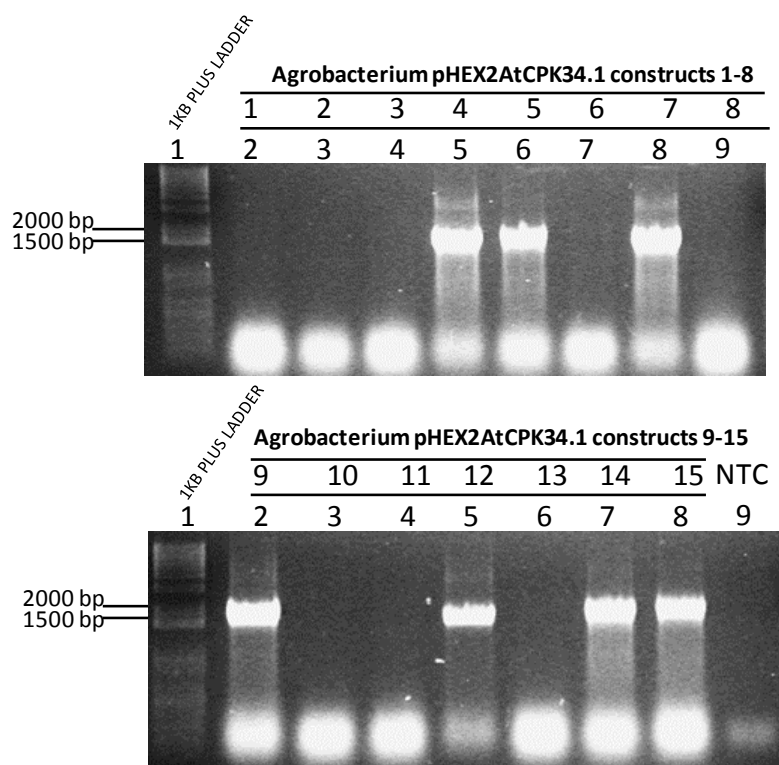
**Figure 5.24. Colony PCR of Agrobacterium GV3101 transformed with pHEX2-AtCPK34.1 construct**. Positive Agrobacterium clones were 5, 6, 8, 9, 12, 14 and 15.

One of the clones was randomly selected and transformed into three pots of *A. thaliana* with about 20-30 individual plants by floral dipping. The cycle of transformation of *A. thaliana* and collection of first (T1) generation seeds was done twice as the first transformation attempt did not produce any successful T1 transformants. The second transformation resulted in 12 transformants out of approximately 300 seeds sown. These lines were propagated under kanamycin selection until the third (T3) generation. Three lines (pHEX2AtCPK34.1, .2 and .3) were further propagated under kanamycin selection until the fourth (T4) generation to ensure homozygosity.

### 5.3.3.2 Verification of AtCPK34 overexpressor and knockout lines

The AtCPK34 overexpressor lines that were developed and the T-DNA knockout lines obtained from NASC were tested using AtCPK34 qPCR primers (section 4.2.2.4a, Table 4.14) by end-point PCR and qRT-PCR to check for AtCPK34 expression in the plants. PCR products of about 120 bp were observed for the three overexpressor pHEX2AtCPK34 leaf samples, one pHEX2AtCPK34.1 flower sample and the WT flower sample. All the leaf and flower samples from the pHEX2AtCPK34 lines appeared to be brighter than the WT flower sample, whereas the flower sample from T-DNA knockout lines showed a very weak PCR product, assuming equal amounts of template used.
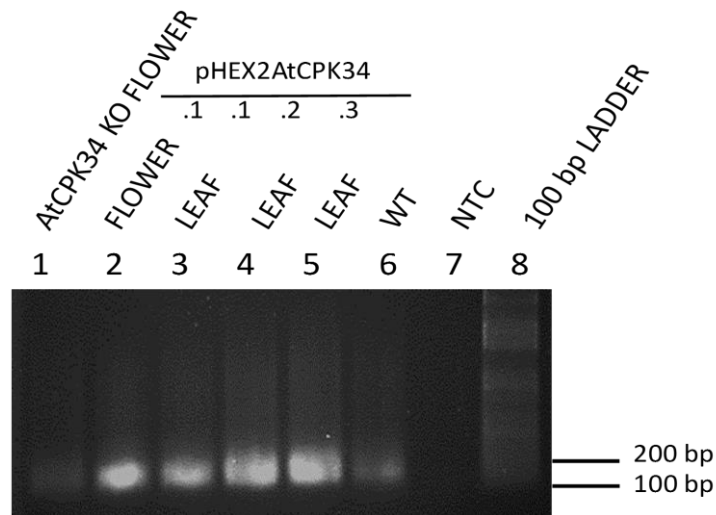
**Figure 5.25 PCR results comparing AtCPK34 wildtype, overexpressors and knockouts.** Lane 1 AtCPK34 knockout from NASC, flower. Lane 2 pHEX2AtCPK34.1, flower. Lane 3 pHEX2AtCPK34.1 leaf, Lane 4 pHEX2AtCPK34.2 leaf. Lane 5 pHEX2AtCPK34.3 leaf. Lane 6 Wild type. Lane 7 negative control. Lane 8 1-Kb ladder.

To further verify the difference in expression, qRT-PCR was performed to determine relative expression of AtCPK34 (Figure 5.26). The flower tissue from AtCPK34 T-DNA knockouts showed a three-fold decrease in transcript accumulation compared to the wild type flower, while the floral tissue from pHEX2AtCPK34.1 showed an eight-fold increase and the leaf tissue showed a 19-fold increase. The two other leaf tissue from AtCPK34 overexpression lines showed 11-fold and 31-fold increases in transcript accumulation compared to the wild type, respectively. Leaf samples were not tested for AtCPK34 from wild type and knockouts as it is not expressed in the leaves based on microarray, qRT-PCR data and previous literature, as described in sections 3.3.2.2 and 4.3.2.1.
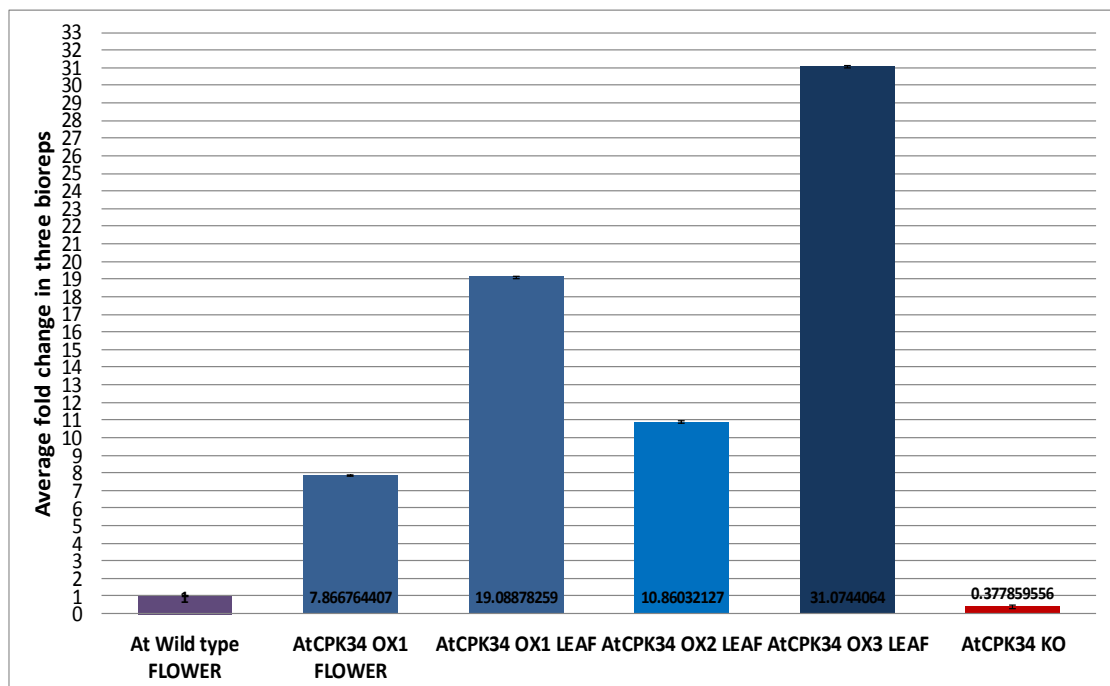


**Figure 5.26 AtCPK34 transcript accumulation in AtCPK34 WT, OX and KO plants.**

### 5.3.3.3 Seed germination assay

To determine similarity in function between AtCPK3 and 34 in relation to salt stress response, a seed germination assay with different concentrations of salt (0, 75, 150 and 300 mM NaCl) was performed with Arabidopsis WT, OX and KO lines of both genes, in three experimental replicates (Figure 5.27). Germinating and non-germinating seeds are shown in Figure 5.28.

In all the lines tested, the seed germination rate generally decreased as the salt concentration increased. None of the lines showed seed germination in 300 mM salt, while very little difference was seen among the lines at 75 mM. Marked differences in response to 150 mM salt were observed for the AtCPK3 overexpressing lines compared to WT plants.

At 150 mM salt, AtCPK3 overexpressors (SAIL and pHEX2 lines) showed three times greater seed germination rates than WT (60% vs 20% germination rate). Both SAIL and pHEX2 lines showed strong statistical difference compared to WT (both Tukey's P= 0.000, Fisher's LSD P= 0.000). AtCPK34 overexpressors had marginally greater seed germination rates (40% and 25%) than WT. The statistical support was strong for pHEX2AtCPK34 OX2 (Tukey's P= 0.000, Fisher's LSD P= 0.000) but weak for the pHEX2AtCPK34 OX3 line.

AtCPK3 knockouts had marginally greater seed germination rates (25% and 33%) than WT at 150 mM salt. The *atcpk3-2* KO lines showed strong statistical evidence (Tukey's P=0.005, Fisher's LSD P=0.000) while *atcpk3-1* KO lines showed weak statistical evidence (Fisher's LSD P= 0.095). On the other hand, AtCPK34 knockouts did not show significant difference from WT.
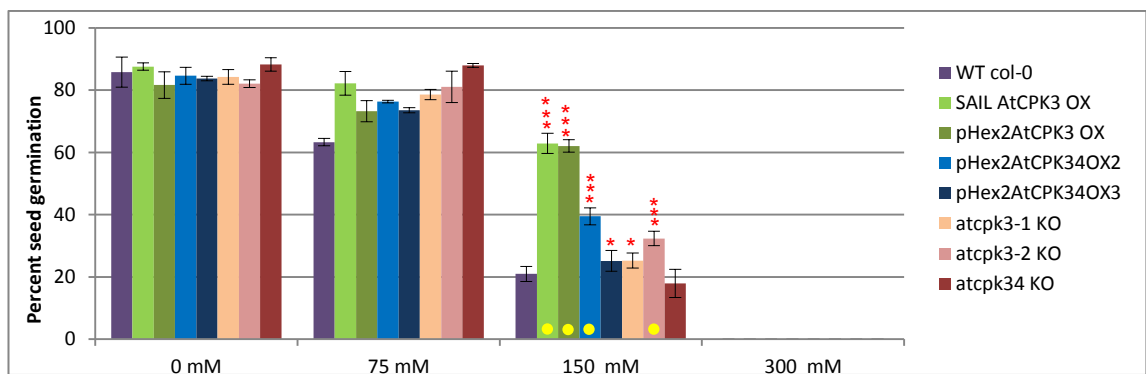


**Figure 5.27 Seed germination rates for AtCPK3 WT, OX, KO's, and AtCPK34 WT, OX, and KO plants.** Colour of bars match the plant lines: purple, WT Arabidopsis; light green, SAIL AtCPK3 OX; moss green, pHEx2AtCPK3 OX; blue, pHEx2AtCPK3 OX1; dark blue, pHEx2AtCPK3 OX2; peach, atcpk3-1 KO; pink, atcpk3-2 KO; and red, atcpk34. Line bars indicate SE of the mean. Yellow dots indicate a marked difference between the transgenic line and WT. Statistical support is indicated as: strong (***, P≤ 0.01), good (**, 0.01<P≤ 0.05) or weak (*, 0.05<P<~0.10).
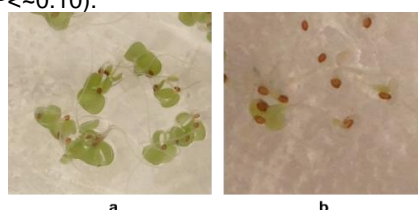


**Figure 5.28. Arabidopsis seeds (a) germinating and (b) non-germinating.**

### 5.3.3.4 Pollen germination assay

To determine similarity in function between AtCPK3 and 34 in relation to pollen development, a pollen germination assay was performed among wild type, knockout and overexpression lines of both genes, in three experimental replicates (Figure 5.29). Pollen germination was observed in all samples treated with the pollen germination medium. WT pollen showed 65% germination in the pollen germination medium and no germination in water (negative control) as expected.

All the knockout and overexpression lines tested showed a lower pollen germination rate compared to the WT in pollen germination medium. AtCPK34 overexpression lines showed lower pollen germination rates (45 to 50%) than WT with good statistical support (Fisher's LSD P= 0.003 and 0.035). Similarly, the AtCPK3 overexpression lines showed lower pollen germination rates than WT; the SAIL lines with 22% germination had strong statistical support (Tukey's and Fisher's LSD P= 0.000) while the pHEX2AtCPK3 lines with 52% germination had good statistical support (Fisher's LSD P= 0.037). AtCPK34 knockouts also showed significantly lower pollen germination rate with very strong statistical support (32% germination, Tukey's P=0.003, Fisher's LSD P=0.000) while AtCPK3 knockouts showed marginally lower pollen germination rates (41-50% germination) with good statistical support (Tukey's P= 0.167, 0.021, 0.035; Fisher's LSD P= 0.013, 0.001, and 0.002).
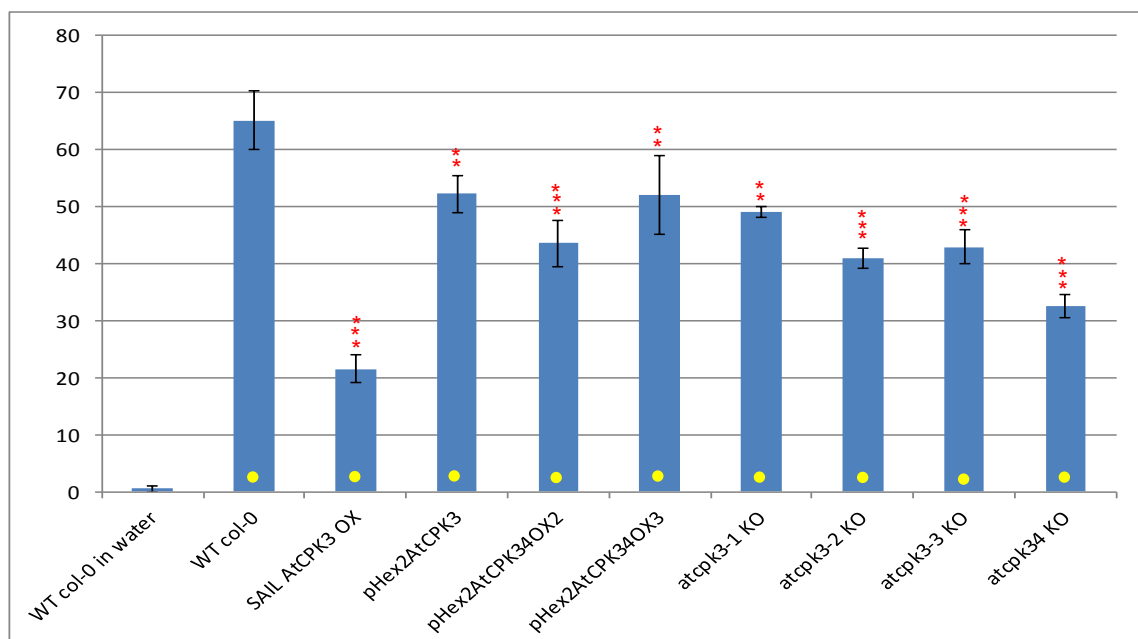


**Figure 5.29 Pollen germination rate for AtCPK3 WT, OX, KO, AtCPK34 WT, OX, and KO plants.**
Yellow dots indicate a marked difference between the transgenic line and WT. Statistical support is indicated as: strong (***, P≤ 0.01), good (**, 0.01<P≤ 0.05) or weak (*, 0.05<P<~0.10).

Most of the germinated pollen showed similar morphology, except for the AtCPK3 SAIL overexpressing lines and AtCPK34 knock-out lines (Table 5.54), which showed tetrad formations, characteristic of delayed pollen development. pHEX2AtCPK3 overexpression lines also showed tetrad formation, but only in some of the pollen. Very long pollen tubes were observed in some pollen from the AtCPK34 overexpression lines and AtCPK3 knockout lines.

**Table 5.4 Pollen morphology among WT, OX and KO lines of AtCPK3 and AtCPK34**

| Transgenic lines | Distinct morphology observed in some pollen |
|---|---|
| WT col-0 in water | normal |
| WT col-0 | normal |
| Atcpk3-1 KO | very long pollen tubes present in about 5 out of 218 pollen, 400-600uM |
| Atcpk3-2 KO | very long pollen tubes present in about 5 out of 202 pollen, 400-600uM |
| Atcpk3-3 KO | very long pollen tubes present in about 5 out of 196 pollen, 400-600uM |
| SAIL AtCPK3 OX | mostly tetrads |
| pHEX2AtCPK3 OX | some tetrads |
| At cpk34 KO | mostly tetrads |
| pHEX2AtCPK34OX2 | very long pollen tubes present in about 5 out of 233 pollen, 400-600 uM |
| pHEX2AtCPK34OX3 | very long pollen tubes present in about 5 out of 237 pollen, 400-600 uM |
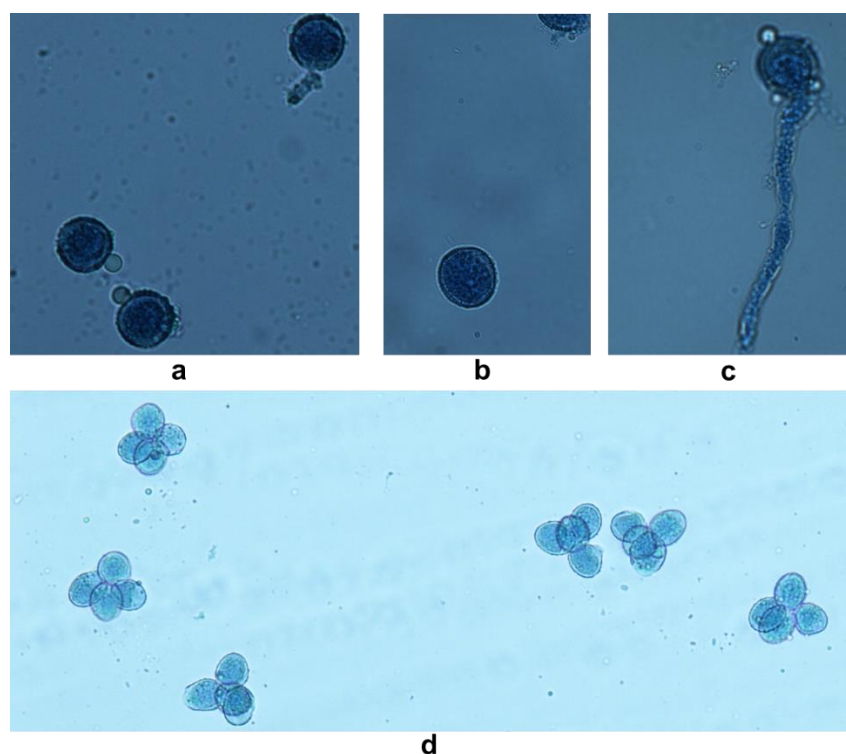


**Figure 5.30 Arabidopsis pollen morphologies observed.** (a) Germinating, (b) Non-germinating, (c) Long pollen tubes, and (d) Immature tetrads.

**5.4 Discussion**

Four factors that may influence functional specificity among CPKs were examined in this chapter. These factors were primary and secondary protein structure, tertiary protein structure, gene regulatory regions, and tissue localisation. Analyses of the four factors were done among Group IIb CPKs, which were noted as the most conserved group based on the phylogenetic analysis of protein sequences performed in Chapter 3. Remarkably, despite having the least difference from their common ancestor compared with other CPK groups, Group IIb showed distinct differences in reported functions between its two subgroups, IIb.1 (AtCPK3 and its orthologues) and IIb.2 (AtCPK17 and 34 and their orthologues). As emphasised in the previous chapter, Group IIb.1 CPKs respond to biotic and abiotic stresses, whereas Group IIb.2 respond to stimuli involved in pollen and floral development. In the analyses performed, factors that potentially influence CPKs to differentiate to stress response functions and floral development were identified (summarised in Table 5.5). Several hypotheses were developed from the analyses performed, which require further investigation.

Firstly, there are protein motifs identified in the analyses that may be important to differentiate CPKs into either stress-response or floral development. In separate primary and predicted secondary structure analyses in Arabidopsis and rice, sixteen motifs (motif At 1 to 16) were identified among AtCPKs, while fourteen motifs (motif Os 1 to 14) were identified among OsCPKs. These motifs were found throughout all the CPK domains: At motifs 1 and 2 and Os motifs 1 to 3 in the N-VD; At motifs 3 to 11 and Os motifs 4 to 9 within the PK; Os motifs 10 and 11 within the AJ region; At motifs 12-15 and Os motifs 12-13 within the CAD domain; and At motif 16 and Os motif 14 within the CT region. With the combined primary and secondary structure analysis of CPKs in these two representative species, fourteen group-specific motifs (AtOs motif 1 to 14) were found to be common between Arabidopsis and rice described in section 5.3.1.1.

Secondly, the predicted tertiary structure of the identified motifs provided stronger evidence for potential correlation with functional specificity, but only in a number of motifs. Subgroup-specificity in the predicted protein tertiary structure was observed in only eight of the

**Table 5.5 Factors that may influence CPK function specificity**

| Group IIb CPKs | Shared by Group IIb.1 CPKs | Shared by Group IIb.2 CPKs |
|---|---|---|
| **Function represented** | Stress response functions | Developmental functions |
| **Primary/secondary structure** | Group IIb.1 AtOs motif 1 to 14 | Group IIb.2 AtOs motif 1 to 14 |
| **Tertiary structure (domains)** | N-VD does not cover the entire left hemisphere at the angle shown in Figure 5.4 | N-VD spans the middle part of the molecule as an S-shaped structure, from the right hemisphere across to the left hemisphere at the angle shown in Figure 5.4 |
| | Opening towards the active sites in PK are pointed or v-shaped | Opening towards the active sites in PK are more curved or irregularly-shaped |
| | CT domain appears as a pointed protrusion at the base of the protein | CT domain appears globular or irregularly shaped |
| **Tertiary structure (motif-specific)** | AtOs motif 2 (NVD)-no prominent shape<br>AtOs motif 3 (PK) -projections at the tip of motif<br>AtOs motif 4 (PK) - located towards the left side in the angle shown<br>AtOs motif 9 (PK)- general size at CT<br><br>AtOs motif 11 (CAD) - projections less prominent<br><br>AtOs motif 12 (CAD) - more internal location<br>AtOs motif 13 (CAD)- more narrow<br>AtOs motif 14 (CT)- protruded | AtOs motif 2 (NVD)- u-shaped structure<br>AtOs motif 3 (PK)- no evident projections<br>AtOs motif 4 (PK)- located towards the middle in the angle shown<br>AtOs motif 9 (PK) - sharper ends<br>AtOs motif 11 (CAD) - short finger-like projections prominent<br>AtOs motif 12 (CAD)- more exposed at the protein surface<br>AtOs motif 13 (CAD) - wider<br>AtOs motif 14 (CT)- more attached to the whole protein |
| **Promoter regions** | ABA inducible transcriptional activator<br>RITA<br>ATNAC proteins | GT1-Box binding factors<br>Trihelix DNA-binding factor GT-3a |
| **Localisation** | gene expression throughout the plant anatomy | gene expression in floral tissue/ pollen |

motifs analysed: AtOs motif 2 in the NV-D, AtOs motif 3, 4, and 9 in the PK, AtOs motif 11, 12 and 13 in the CAD, and AtOs 14 in the CT. These motifs may prove important in substrate specificity and calcium sensitivity and thus are good targets for future experiments to investigate CPK functional specificity. AtOs motif 3, 4, and 9 in the PK can be analysed in relation to the target substrates of CPKs in each subgroup as these influence the shape of the region within and around the active sites. Motifs that are within the CAD or adjacent to EF hands can be investigated and compared in terms of the calcium binding activity of CPKs in each subgroup. Since the N-VD has been reported to be responsible for the membrane association of CPKs (Cheng et al. 2002; Dammann et al. 2003; Lu and Hrabak 2002; Martin and Busconi 2001), AtOs motif 2 must be important in understanding and/or predicting subcellular localisation of CPKs, particularly where there are myristoylation and palmitoylation sites. Moreover, this region may also contribute to substrate recognition (Asai et al. 2013). Group IIb.1 CPK motifs may also be analysed to differentiate between stress-specific responses among CPKs; however, at present the number of CPKs in Arabidopsis, rice and other plant species with reported roles in virus, bacteria, drought and salt stress do not seem to be enough to be able to draw strong inferences.

Thirdly, there were elements in the promoter region of Group IIb CPKs in Arabidopsis and rice, which specifically correlate with their reported function, particularly between Group IIb.1 and IIb.2, which have stress response and developmental functions, respectively. There were examples where candidate transcription factor binding sites were similar between all members of Group IIb, or almost all of the members, while some were uniquely shared by the CPKs belonging to either Group IIb.1 or IIb.2. The *in silico* analysis performed is valuable in opening further research prospects and experiments looking at promoter regions of CPKs their functional specificity. The promoter regions identified in this chapter (section 5.3.2) that are shared by Group IIb.1 and IIb.2, namely Class I GATA factors, SPF1 and SBF-1 are good initial targets in understanding why CPK genes can overlap in function. On the other hand, those that are uniquely shared by Group IIb.1, such as ABA inducible transcriptional activator, RITA and ATNAC proteins, and those that are uniquely shared by Group IIb.2, such as GT-1 and GT-3a factors are elements that can be studied in order to elucidate CPK functional specificity to stress response or to development, respectively.

Lastly, Group IIb.1 and Group IIb.2 potentially show similarity in function when overexpressed in a similar tissue location, but with reduced effects as compared to the other gene. Under salt stress, in comparison to WT Arabidopsis, seed germination rates were significantly higher in AtCPK3 OX lines and marginally higher in AtCPK34 lines; whereas seed germination rates were marginally higher in AtCPK3 KO lines and did not significantly differ in AtCPK34 KO lines. This indicates that tolerance to salt stress can be conferred by the overexpression of AtCPK3 and to a lower extent by the overexpression of AtCPK34, and thus some similarity in their functions when expressed throughout the plant's anatomy. The pollen germination rates were decreased, either marginally or significantly, among AtCPK3 and AtCPK34 OX and KO lines, compared to WT. This indicates that both genes play important roles in pollen development, and that the changes in the expression of these genes, whether upregulation or downregulation, may affect pollen germination. Interestingly, the abnormalities observed in developing pollen were similar between AtCPK34 OX and AtCPK3 KO lines, having some extremely long pollen tubes, and between AtCPK34 KO and AtCPK3 OX lines, having delayed pollen development The findings in section 5.3.3 provide preliminary evidence that AtCPK3 may function similarly to AtCPK34 and vice versa, and that the function of each protein is influenced by the location where the gene is expressed, but with less efficiency compared to the gene that naturally has the function. This hypothesis may be tested further by using double mutants, such as a line that has AtCPK3 knocked out and AtCPK34 overexpressed, and a line that has AtCPK34 and 17 knocked out and AtCPK3 overexpressed. These lines must be subjected to various stress treatments and pollen development assays, and the number of plant samples and biological replicates must be increased for stronger statistical support.

In summary, CPK functional specificity to stress-response or pollen development are potentially influenced by the protein motifs, structure, promoter binding sites and tissue localisation listed in Table 5.5. The findings in this chapter are only limited to the predicted secondary and tertiary structures, predicted transcription binding sites, number of plant lines available for experimental analysis. These findings open future research investigating the structure of Group IIb plant CPKs determined using X-ray crystallography and understanding functional specificity using biological experiments such as target substrate identification, protein-protein interaction studies, promoter exchange, and site directed mutagenesis in the identified motifs and/or promoter regions, among a wider range of plant species and in bigger sample size.

# Chapter Six

# General Discussion and Future Directions

This thesis project made an original contribution to science by providing novel, ecologically valuable and agriculturally useful insights to three main questions about the functional diversification and specificity of CPKs.

Firstly, how did CPKs diversify and what is the most conserved CPK group in plants? The phylogenetic analysis of CPKs from lower to higher plants showed that CPKs diversified in parallel with the transition of plants into terrestrial life and that the most conserved members of this gene family in plants are those that belong to evolutionary Group IIb. The findings described and discussed in Chapter 3 led to a hypothesis that CPK evolution was one of the significant processes that allowed plants to colonise terrestrial environments, as plants evolved.

Secondly, what is the role of the most conserved CPKs in plant stress and pathogen responses? CPKs from subgroup IIb.1 (AtCPK3 and its orthologues) change in transcript accumulation in response to most abiotic stresses and pathogens, as inferred from meta-analysis of publicly available transcript data and as validated from biological experiments in Arabidopsis, rice and kiwifruit that were described in Chapter 4. Knocking out or overexpressing these genes appeared to change the way plants respond to stress and pathogens. On the other hand, CPKs from subgroup IIb.2 (AtCPK17 and 34 and orthologues) function mainly in pollen development. From these findings, it can be hypothesised that the most conserved CPKs in subgroup IIb.1 are useful target genes for the indirect detection of plant pathogens and for the improvement of plant disease resistance or stress tolerance; while CPKs in subgroup IIb.2 are important gene targets

for understanding or improving pollen development. However, further investigation is needed to provide stronger scientific support, as discussed in section 4.4.

Thirdly, what influences CPK functional specificity? The functional specificity of the most conserved CPKs is determined by a combination of several factors such as gene structure, protein structure, and tissue localisation. The similarities in promoter regions, the high level of protein sequence conservation among CPKs, the presence of few protein motifs correlated to function, and the variations in tissue localisation and expression levels explain why most CPK functions are usually redundant and overlapping, and why they are useful as plant signalling hubs. On the other hand, subgroup specificity was also determined in all of the four factors for Group IIb.1 and IIb.2. Factors that may influence CPKs into having stress response functions or developmental functions are determined and described in Chapter 5 of this thesis.

In light of the questions answered, this chapter expounds on the new insights, new hypotheses and new questions posed with regards the evolution of CPKs as well as the use of group IIb.1 CPKs for novel molecular and diagnostic approaches in managing plant biotic and abiotic stress across a broad range of plant species. The relevant results from chapter three to five of this thesis will be written as publications (Appendix 33).

**CPK evolution and functional diversification**

The phylogenetic analysis presented in Chapter 3 provided insights into how the CPKs identified in representative species from lower to higher plants were evolutionarily related, and into how CPK diversification correlated with the timing of plant terrestrial transition and adaptation. As mentioned in Chapters 1 and 2, it was hypothesised by earlier authors that CPK genes have a common ancestor, which was believed to be a fusion between a CaM and CaMK-like protein (Zhang and Choi 2001). CPK basic architecture has been conserved among all organisms that possess this gene family: protists, oomycetes, and all Viridiplantae including green algae and land plants (Hamel et al. 2014). However, as presented by Bilker, Lourido and Sibley (2009), protist CPKs have evolved into several groups, independent of plant CPKs. In this thesis, green algae it was determined that CPKs diversified into several groups through evolutionary events that were also independent of land plant CPK evolution. This suggests that the current land plant CPKs have evolved into their evolutionary groupings concomitantly with the terrestrial transition of plants. This was supported by the molecular clock analysis presented in this thesis that showed

that the diversification into four groups occurred after the land colonisation of plants. The findings of the phylogenetic analysis in this thesis was published as an original paper in Plant Physiology in 2014 (online publication December 2013) and has been a useful reference by several succeeding papers that studied the evolution and function of CPKs, as well as other genes involved in calcium signalling. Recent papers that have cited the said publication are listed in Appendix 34.

Notably, the insights presented in this thesis agree with the findings of a concurrent independent study on the comparative genomics of green plant CPKs reported by a research group in the USA (Hamel et al. 2014). Their study included 132 land plant and 48 putative green algal CPK sequences from two eudicots (*A. thaliana* and *P. trichocarpa*), one monocot (*O. sativa*), one pteridophyte (*S. moellendorfii*), one bryophyte (*P. patens*) and eight green algae (C. *reinhardtii, V. carteri, Coccomyxa subellipsoidea, Chlorella variabilis, Ostreococcus lucimarinus, Ostreococcus tauri Micromonas pusilla* (RCC299) and *Micromonas pusilla* (CCMP1545)). Similar to what was presented in this thesis, their paper suggested that the current architecture of the CPK gene family in plants was formed during the colonisation of land by plants, and that the ancestral green algae CPK evolved independently from those of land plants. This publication also used CPK protein sequences and ClustalW alignment, and reported the presence of atypical CPKs with fewer than four EF hands or extra long NV-D, which were also mentioned in Chapter 3. However, their study gave more emphasis to green algae CPKs and only included a few representative species for land plants. On the other hand, in this thesis project, emphasis was given to land plants as there were five eudicots, four monocots, one gymnosperm, one pteridophyte, one bryophyte, and two green algae species included in the analysis. Moreover, Hamel et al. (2014) used unrooted NJ trees and drew conclusions based on the clustering of sequences in the phylogenetic tree, whereas this thesis project used five protist sequences as the outgroup for a rooted tree, and included a molecular clock analysis to support our inferences. Both approaches were scientifically valid and demonstrated comprehensive analyses; however, rooted trees provide stronger representation of the evolutionary relationship between sequences as these reflect basal lineages. Their publication also gave emphasis to the domain organisation among CPK sequences to correlate CPK architecture with land colonisation by plants, while this thesis project gave emphasis to phylogenetic analysis and functional information about CPKs. Their publication defined four major clades of green alga CPKs, with a distinct lineage from the

land plant groups, whereas this thesis project identified only two. With the data presented by two independent groups, there is therefore strong scientific evidence to affirm that land plant CPKs evolved concurrently with the terrestrial transition of plants. It was also suggested in both studies that the expansion and diversification of the CPK family is mainly due to whole-genome duplications or polyploidisation in plants and recent gene duplication events. Single and or multiple gene duplications as well as gene losses were also observed in some of the species as presented in both studies.

Further genome-wide identification and phylogenetic analysis of CPKs in other plant species was done by other research groups in the subsequent years. Mohanta et al (2015) performed a vast genome-wide identification and phylogenetic analysis of CPKs, which included a total of 950 sequences from five green algae species, one bryophyte moss, one pteridophyte fern, one gymnosperm, six monocots and 26 dicots.This publication used an unrooted NJ tree, similar to Hamel et al (2014). Their phylogenetic analysis also described four major evolutionary groups of CPKs. However, it was not mentioned whether the green algae CPKs were included within the four groups, or whether they form a separate lineage from land plants. This paper also determined a difference between the EF domain structures between higher plants and lower plants. Higher plants have four D-X-D and two D-E-L motifs in the EF hand domains, while lower plants have two D-X-D and one D-X-E motifs in the EF hand domains. It was also determined using Tajima's neutrality test that the CPK gene family is under balancing selection, which means that multiple alleles are being maintained as it provides an advantage. This may help explain the functional redundancy among multiple CPK genes within a species. This publication also suggested a new numbering system and nomenclature for CPKs, which is based on the numbering of Arabidopsis CPKs for dicots and rice CPKs for monocots. This may be good approach particularly in predicting function based on orthologous sequences. However, it may pose some difficulty in cases when there are several orthologues for one Arabidopsis or rice CPK gene, or if there are more than one Arabidopsis or rice CPK that are potentially orthologous to the gene being named.

As many plant genomes have been recently completed, many other genome-wide or transcriptome-based identification of CPKs in different plant species have been reported since the publication of the results of this thesis in Chapter 3. This includes CPKs from *Vitis spp, Rafflesia cantleyi, Glycine max, Hordeum vulgare, Cajanus cajan, Brassica rapa, Hevea*

*brasiliensis,* and *Cucumis melo* (Amini et al. 2016; Fedorowicz-Strońska et al. 2017; Hettenhausen et al. 2016; Wankhede et al. 2017; Wu et al. 2017; Xiao et al. 2017; Zhang et al. 2017; Zhang et al. 2015). All of these publications grouped CPKs into four evolutionary groups and found out that CPKs are highly conserved, with redundant and complementary functions. Some have also included CRKs (Zhang et al. 2017) and other members of the CDPK-SnRK superfamily (Xiao et al. 2017). Most of the recent papers also performed intron-exon analyses and expression profiling. The study on genome-wide identification of grape CDPKs (Zhang et al. 2015) also reported dicot and monocot CPKs forming distinct groups within subgroups, and mentioned that many duplicated grape CPKs arose before the divergence of grapevine and Arabidopsis. The study on *Rafflesia* was based on RNA-seq data instead of the whole genome and have studied expression profiles of 14 unique transcripts in different bud stages, but did not present a phylogenetic analysis. For *G. max*, in this thesis, 45 CPK genes were detected, but in the recent study (Liu et al. 2016), only 39 were confirmed as CPK genes; which were also classified into four evolutionary groups. For *H. vulgare*, 29 CPK genes were identified from its genome data and classified into six evolutionary groups--- with groups V and VI clustering with *C. reinhardtii* and *P. patens.* However, the authors mentioned low confidence regarding these new groupings and used Human calmodulin-dependent kinase 1 as an outgroup, which may not be a good outgroup for this analysis. For *C. cajan*, 23 CPK genes were identified and grouped into four evolutionary groups, however the full text could not be retrieved due to limitations in access. For *H. brasiliensis,* 49 BrCPKs were identified, together with other members of the CDPK-SnRK gene family. In this study, CPKs were also grouped into four evolutionary groupings, and it was suggested that the expansion of the CDPK-SnRK gene family started from angiosperms, with segmental duplication being the main driver for the gene family expansion. In the study with *C. melo*, 18 CPK genes were identified and also clustered into four group, and it was reported that CPKs and CRKs have a common ancestor.

From these recent studies, CPK genes that belong to Group IIb.1 and potentially orthologous to AtCPK3 include VvCPK1, GmCPK2, 12, 17 & 23, BrCPK21, 28, 29, & 30 and CmCPK11. On the otherhand, CPK genes that belong to Group IIb.2 and potentially orthologous to AtCPK17 & 34 include VvCPK6, GmCPK5, 14, 29 & 37, BrCPK4, 10, 14, 31 & 36 and CmCPK9. GmCPK2 and 23 showed high expression in response to wound and herbivore attack, while GmCPK5 and 29 showed high expression in dry seed (Liu et al. 2016). BrCPK 4, 10, 21, 28 and

36 showed upregulation in response to salt while BrCPK30 showed downregulation in response to salt. BrCPK10, 29 and 36 showed downregulation to PEG (osmotic stress). BrCPK4 showed specific expression to flower tissue, but showed upregulation in response to PEG. CmCPK11 showed some degree of downregulation in response to salt. CmCPK9 is highly expressed in flowers and also showed downregulation in response to the powdery mildew *P. xanthii.* These findings provide further support to the importance of Group IIb.1 CPKs in stress and the importance of Group IIb.2 CPKs in development, but also potential functional overlaps between these genes.

More research work have also reported the importance of certain CPKs in abiotic stresses such as drought, salinity, cold and heat stress. To mention some of these reports are: *Zea mays* CPKs ZMCK3 which endows tolerance to heat and drought stress, and ZmCPK1, a negative regulator in cold stress; rice OsCPK9 which regulates drought stress tolerance; grape VaCPK20 which mediates cold and drought stress, VaCPK29 which is important in heat and osmotic stress and VaCPK21 which is involved in salt stress response; and barley CPK2a which is important in drought response (Ciésla et al. 2016; Dubrovina et al. 2015; 2016a; 2016b; Wang and Song 2014; Weckwerth et al. 2014; Wei et al. 2014). Likewise, more research regarding the importance of CPKs in biotic stresses have been reported since then, such as potato StCDPK7 which is induced upon Phytophthora infection, Arabidopsis CPK1 which is important in plant defense response, and wheat TaCDPK7 which regulates resistance to sharp eyespot disease (Fantino et al. 2017; Nie et al. 2015; Wei et al. 2016).

Several questions with regards CPK evolution remain to be answered. Firstly, why are there so many CPK gene family members? After the whole genome duplication events in land plants, why did gene loss not occur on a large scale among CPKs (except for some species such as *V. vinifera*) and why have so many CPK genes been maintained? Secondly, while the redundancy and overlaps in CPK functions were explained in this thesis by high sequence conservation and multiple factors defining function, it is still unclear why functional redundancy and overlaps occurred during evolution. Why did land plant evolution favour the occurrence of CPKs with redundant and overlapping functions? Why do the evolutionary groupings (Group I-IV) not correlate with group-specific functions? Lastly, why are there primarily two divergent functions among land plant CPKs; that is pollen development and stress response? Was the divergence a

result of neofunctionalisation or subfunctionalisation of the ancestral gene of all land plant CPKs? Or were these the results of independent evolutionary events?

With regards these questions, it can be hypothesised that most extant plants have so many CPK gene family members because this characteristic provided a selective advantage to plants as they evolved towards thriving in environments with less water supply. Since the transition from aquatic to terrestrial environments may pose a drastic change in physiological requirements, ancestral plants would have needed to develop new biological processes for a terrestrial niche, but at the same time maintain the fundamental processes that they used to have in an aquatic niche. In this regard, there may have been selective pressure for gene duplications followed by subfunctionalisations and neofunctionalisations among genes to occur. Moreover, having genes with redundant and overlapping functions could potentially have provided security for a specific function to be maintained in spite of the pressures that an abiotic stress or pathogen infection pose to a particular gene. Similar to CPKs, a study on the evolution of CaM and CML reported that both evolved with striking diversity during plant terrestrial colonisation and that the CML gene family expanded as a result of selective pressures in adapting to a land environment (Zhu et al. 2015). The results of their study, the report of Hamel et al. (2014) and the findings of this thesis support the importance of calcium signalling as a significant cellular process that evolved during the terrestrial transition of plants. This hypothesis is also supported by a comprehensive review by Plattner and Verkhratsky (2015) on the ancient roots of the calcium signalling evolutionary tree.

As mentioned in Chapter 3, the expansion of CPK gene family occurred as a result of whole genome duplication (WGD) events and many cases of single gene duplication events. WGD, or polyploidisation is considered as an important evolutionary force among all organisms, especially plants (Jiao et al. 2011b). A large number of plant genes appear to have arisen as a result of polyploidisation (Jiao et al. 2011b). As mentioned in section 3.4.2 examples of other gene families that have shown expansion among higher plants include CaM (Zhu et al. 2015), floral MADS-box (Nam et al. 2003) and prolamin genes (Xu and Messing 2008). Moreover, WGDs have been correlated with the diversification of regulatory genes important to the development of seeds and flowers (Jiao et al. 2011b), which are important adaptive processes for flourishing in terrestrial environments. From this information, it can also be hypothesised that gene family evolution in

plants is commonly a result of polyploidisation and adaptation to terrestrial life, and one good example of this is the evolution of the CPK gene family.

Similar to other calcium signalling proteins, CPKs function both in plant seed/floral development and in plant biotic and abiotic stress response. The diversification into these two different functions did not correlate with CPK evolutionary groupings, nor did specific types of stress response. The branching between developmental CPKs and stress-responsive CPKs was only evident among Group IIb CPKs wherein the two subgroups IIb.1 and IIb.2 showed distinct functions. Moreover, the lowest level of plants with reports about CPK function in development were bryophytes (Nishiyama et al. 1999). No reports about CPK developmental function among green algae has been reported yet. However, because research on CPKs among green algae is limited, this inference could not be considered final. Assuming that CPKs do not function in development among green algae, two hypotheses can be drawn. Firstly, it is possible that the land plant ancestral CPK gene originally arose having both of the functions, and have subsequently diversified into the ancestral genes of the four groups still maintaining both functions, which then independently diversified and subfunctionalised into either stress or developmentally responsive. Secondly, it is also possible that the ancestral gene originally functioned in response to stress, diversified into the four evolutionary groups, which then have all neofunctionalised for seed and pollen development in correlation with WGD and further adaptation to terrestrial environments. These hypotheses may be tested by determining the possible ancestral CPK sequence among land plants and analysing its functions using both bioinformatic and experimental approaches.

It can also be hypothesised that when there are multiple CPKs working within the same pathway (e.g. bacterial response), with redundant function (CPKs can promiscuously phosphorylate proteins), the specific target may be less important for each CPK than the likelihood of a CPK finding any of its possible targets within the cell. It then becomes more important for the cell to maintain the CPK amounts and variations at high level to increase the likelihood of an interaction with a target when required. Since there are different CPKs that can interact with the target substrates, the effect of a different stimulus to a CPK can be compensated by other CPKs that are present. This leads to further questions such as: Is having multiple genes with overlapping/redundant function more efficient than having one gene producing sufficient protein to equal what the multiple genes might produce? What amount of CPK proteins are needed by

different kinds of cells to ensure they are likely to be activated and interact with targets when required? How much CPK protein is present in cells at any given time?

In the case of Group IIb CPKs presented in this thesis, Subgroup IIb.1 (AtCPK3 and orthologues) and IIb.2 (AtCPK17 and 34 and orthologues) do not show overlapping functions in a normal plant. However, as demonstrated in sections 5.3.3.3 and 5.3.3.4, single gene mutant plants appeared to show similarity in the responses to salt stress and pollen germination, but to a lesser extent. It is therefore possible that CPKs from group IIb.1 can operate within pathways that CPKs from group IIb.2 operate in, and provide redundant function, but with reduced efficiency. The CPK tertiary structures were highly conserved, but motifs that differentiate group IIb.1 from IIb.2 were also identified. Therefore, some molecular functions and target substrates must be shared or conserved between the two subgroups. Likewise, some molecular functions and target substrates must be different between the two and specific to either stress response or pollen development. It can be hypothesised that while ACPK17 and 34 have become pollen specific, they may still have substrates in common with AtCPK3, albeit in a limited cell type. As there are two genes in Arabidopsis and four genes in rice performing the same developmental function, are these genes going to undergo subfunctionalisation or neofunctionalisation? Or are they currently undergoing these processes? Moreover, there were two orthologues of AtCPK3 in rice, OsCPK1 and 15, which match in transcript accumulation in some stresses but show opposite response or unique response to certain stresses. Has subgroup IIb.1 in monocots undergone subfunctionalisation or neofunctionalisation? These questions may be verified in future studies including substrate identification, gene promoter exchange and protein interaction studies as mentioned in section 5.4.

The single gene knockout experiments provided some insights to the overlapping functions but did not clearly answer the question with regards CPK functional specificity. By knocking out or lowering the expression of a single CPK, any function to which it relates to would have been reduced. Further clarification may be gathered from performing experiments with double mutants. For example, plants with double knockouts of AtCPK17 and 34, together with overexpression of AtCPK3 in pollen may demonstrate that AtCPK3 can also function in pollen development, if normal pollen development occurs in the combination mutants. On the other hand, knockouts of AtCPK3 with either AtCPK17 or 34 overexpressed may demonstrate that AtCPK17 or 34 may function similarly to AtCPK3 in response to different stresses. Double knockouts of

AtCPK17 and 34 transformed with mutant forms of either CPK17 or 34 could be developed to identify the regions necessary for development. Likewise, knockout mutants of AtCPK3 transformed with mutant forms of AtCPK3 could be developed to identify the regions necessary for stress response. It could be hypothesised that these regions would be the motifs identified in section 5.3.1 as being specific to Group IIb.1 and Group IIb.2. These should be the first regions to be analysed for these type of studies. Likewise, mutations in the promoter regions and promoter exchange experiments between Group IIb.1 and Group IIb.2 may also elucidate specific elements that control CPK functions.

**AtCPK3 and its orthologues as target genes for molecular diagnostics of plant disease and managing plant stress tolerance/disease resistance**

Focusing on the role of Group IIb.1 in plant stress responses and pathogen infections provided information about the range of stimuli to which the most conserved CPKs respond to and how, in terms of transcript accumulation. Additionally, insights to how these genes could be used as targets for diagnosing or managing plant disease were gathered. In Chapter 4 the role of Group IIb.1 CPKs AtCPK3, OsCPK1, OsCPK15 and AcCPK16 in response to biotic and abiotic stress was explored. It was demonstrated that in leaves, the mRNA accumulation of Group IIb.1 CPKs significantly increases in response to virus, while it decreases in response to osmotic stress, fungal and bacterial infections. As this was demonstrated in all Group IIb.1 CPKs tested except for OsCPK1 (although OsCPK15 follows the pattern), it can be hypothesised that Group IIb.1 CPKs may be good gene targets for designing molecular diagnostic tools for plant pathogens, particularly viruses. An elevation of mRNA accumulation is suggestive of infection particularly by viruses and may be used as molecular markers to diagnose infection instead of directly detecting the virus. Direct detection of virus requires previous knowledge of a viral sequence which is not available for a number of pathogenic viruses.

In light of this, a collaborative study is being carried out with another researcher in New Zealand which proposes a molecular diagnostic tool to detect plants that have virus infections (Lilly 2014; Valmonte et al. 2015). Virus infections in plants pose a major threat to agricultural, environmental and economic security globally. However, the available detection tools are limited, particularly with regards plant virus detection. Current assays are limited to symptom observations, which may be inaccurate; and molecular tests to detect virus particles or nucleic acids, which require some prior knowledge about the virus. NGS, where no *a priori* knowledge is

required, is currently too expensive for routine diagnostics. Rather than detecting specific viruses, this study proposes a potential tool to detect virus infection in plants by looking at host cellular responses to viruses. The study explored molecular indicators of virus infection which included measurement of the amount of low molecular weight RNA and transcript accumulation of AtCPK3 and AtSGS3 (Suppressor of Gene Silencing 3). AtSGS3 has been determined to decrease about two-fold in response to virus, with no significant change in other stimuli except for drought, to which it drastically decreases (far more than two-fold) (Lilly 2014). The ratio between AtCPK3 and AtSGS3 transcript accumulation in combination with the amount of low molecular weight RNA may provide a virus indication value, which can be used as a measure of the presence or absence of virus infection. These molecular indicators were identified in Arabidopsis and will be validated in other plant species such as rice and kiwifruit. Further experimentation using different kinds of viruses, including latent infections, needs to be carried out.

The overexpressors of AtCPK3 and AcPK16 appeared to have some degree of tolerance or resistance to virus and fungal infections as these plants showed lower symptom severity compared to wild types and knock outs. However, as mentioned earlier, this needs further investigation as the number of samples available for analysis and that were manageable at the time of experiment was not enough to make definite conclusions. Samples need to be increased to at least 30 per treatment and to be performed in at least three trials, to improve statistical support. However, the findings of this thesis in this regard are still very valuable in opening research into AtCPK3 and its orthologues, with a hypothesis that overexpressing these genes may confer tolerance or resistance to plant diseases or environmental stress; or, alternatively, plants that are naturally able to produce more of these proteins may have higher tolerance and resistance to disease or stress.

**AtCPK3 and extremophiles**

In light of their role in osmotic stress and tolerance to drought and high salinity, it is important to investigate CPK genes and their functions, as well as other genes involved in calcium signalling, among extremophilic species particularly among plants and/or green algae that thrive in salty environments, high temperatures, deserts, and polar regions. Identification of stress responsive genes in halophytes, thermophytes and xerophytes, and exploring their cellular signalling networks, particularly involving calcium may provide valuable insights in understanding how plants adapt to drastic changes in the environment, how they have evolved for terrestrial life,

and how they could evolve in response to climate change that is currently occurring. No studies have yet explored CPKs among these plants and green algae species, nor their calcium signalling pathways. Related studies are limited to comparisons of the transcript profiles between Arabidopsis and its halophilic relative, *Thelungiella halophile* analyses of ROS homeostasis and ion transport in halophytes, and characterisation of a CPK from a halotolerant green alga, *Dunaliella tertiolecta* (Gong et al. 2005; Taji et al. 2004; Yuasa and Muto 1992). Comparison of the transcriptome using microarray data identified genes and pathways that were shared and divergent among the two species, in relation to salt stress response (Gong et al. 2005). Shared genes included those involved in ABA responses, growth regulation, calcium signalling genes and other genes known to be involved in abiotic stress. Certain stress-induced genes were reported to be highly expressed in *T. halophile* even in the absence of stress, such as Fe-SOD, P5CS, PDF1.2, AtNCED, P-protein, β-glucosidase, and SOS1 (Taji et al. 2004). However, these studies focused only on upregulated genes and genes that have intensity differences between the two species, and did not explore genes that were downregulated. Salt-tolerant plant species were reported to not allow excessive ROS production and have $H_2O_2$ signatures in addition to calcium signatures (Bose, Rodrigo-Moreno, and Shabala, 2013), and possess specific anatomical, morphological, and cellular mechanisms that are highly orchestrated towards regulation of ion transport and sequestration (Shabala and Mackay, 2011). Therefore, the identification of CPKs in extremophilic plants and algae and analysis of their sequence, expression and function must be done in future research.

**Future Directions**

Further investigations can provide support for the insights suggested from each chapter or test the hypotheses that were made from the findings of this thesis. Firstly, the hypothesis that CPK evolution was one of the significant processes that allowed plants to colonise terrestrial environments as plants evolved can be further supported in future research by including CPKs from additional plant genomes and from the other genome-wide analyses of CPKs that were reported recently. The phylogenetic study in Chapter 3 was limited to the genomes available at the time when the analysis was done in 2012. As mentioned earlier, new genomes have been completed and new genome-wide analyses of CPKs have been published since then. Including all of these CPK sequences can provide a more robust analysis of CPK evolution.

Further analysis looking at evolutionary rates and determining natural selection processes may also help explain how CPKs were evolutionarily maintained and how they have diversified with redundant and overlapping functions functions. In the $K_a/K_s$ analysis performed in this thesis, it was detected that CPKs (at least from At and Os) are under purifying selection, while in the Tajima's test in three sets of three random sequences performed by Mohanta et al (2016) it was found out that CPKs are under balancing selection. It will be interesting to determine whether these two processes were undergone by CPKs at particular time points or only by a certain group of CPKs, and whether there are other selective processes that play in the evolution of this gene family.

Furthermore, studies that specifically aim to identify all CPK gene family members in different plant species that do not have a completed genome yet may also be useful. This is particularly important with the lower plant species which can be representative of major plant taxonomic groups but do not have current genome sequencing projects. In this regard, a collaborative study with another researcher in New Zealand (Arthur et al. 2012) was carried out within the duration of this thesis project to determine sequence motifs that are characteristic of each of the CPK evolutionary groups, and to design degenerate primers that can pick up CPK sequences in a wide range of plant species. Testing of the usefulness of these primers is ongoing.

Secondly, the insights regarding the importance of Group IIb CPKs in abiotic and biotic stress response as well as in development as reported in Chapters 4 and can be supported by analysing the function of the other group IIb CPKs identified in this thesis project and in the other CPK genome-wide identification recently published. Also, while the results of this thesis project showed trends of marked differences in the transcript accumulation of AtCPK3 and its orthologues in rice and kiwifruit in response to abiotic and biotic, the statistical support is not strong in many of the experiments. This is because the sample sizes were small. Larger sample sizes would have given stronger statistical support; however, due to limitations in time, resources and manageability this was not possible. Nevertheles, the findings of this thesis provide an indication that these genes show differential expression in response to various stresses, follow-up experiments with bigger sample sizes and focusing on specific stresses can be made.

Similarly, there is a need to provide stronger statistical support for the hypothesised stress tolerance among overexpression lines and knockout lines. Increase in sample size and focusing

on a specific stress response is also necessary. As it was also determined that Group IIb.1 CPKs (i.e. AtCPK3) and Group IIb.2 CPKs (i.e. AtCPK17 and 34) analysing double or multiple mutants for these orthologous genes may also provide more insight regarding CPK function in response to stress and development. As developing mutant plants from floral dip methods and tissue culture take a very long time, virus-induced gene silencing (VIGS) can provide a faster and simple way in determining phenotypic responses, particularly in biotic stress treatments.

Lastly, the functional specificity of CPKs can be further analysed looking at double mutants, identifying specific substrates using yeast two-hybrid approaches, bimolecular complementation assays, laser dissection followed by RNA seq, and promoter exchange experiments. Studying double mutants and promoter exchange experiments for each of the orthologues in Group IIb.1 and IIb.2 can verify whether functional specificity is influenced by localisation. Identifying substrates for CPKs in each group using molecular approaches such as yeast two hybrid or bimolecular complementation assays can provide insights regarding similarities and differences in subcellular targets and cellular function of these CPKs. Laser dissection followed by RNA seq can help determine whether tissue and/or cell type specificity account for the functional overlaps observed.

**Summary**

In summary, this thesis suggests that CPKs diversified in parallel with the terrestrial transition of plants and that the most conserved members of this gene family are useful targets for molecular diagnosis of plant disease and for molecular approaches in managing plant biotic and abiotic stress. This thesis also suggests that the functional specificity of CPKs is determined by a combination of its protein structure, gene regulatory regions and tissue localisation, which explains their functional redundancy and overlaps. This thesis also calls for further research on the identification and characterisation of CPKs in other plant species important to agriculture and ecology, which include both crop plants and extremophilic plants and green algae.

# References

Abbasi F, Onodera H, Toki S, Tanaka H, Komatsu S. 2004. OsCDPK13, a calcium-dependent protein kinase gene from rice, is induced by cold and gibberellin in rice leaf sheath. Plant Molecular Biology. 55(4):541-552. doi:10.1007/s11103-004-1178-y.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. Journal of Molecular Biology. 215(3):403-410. doi:10.1006/jmbi.1990.9999.

Amini S, Goh HH, Wan KL. 2016. Identification of calcium-dependent protein kinase (CDPK): A multi-functional gene family in *Rafflesia cantleyi*. Paper presented at: AIP Conference Proceedings doi:101063/14966713

Anil VS, Harmon AC, Rao KS. 2003. Temporal association of Ca(2+)-dependent protein kinase with oil bodies during seed development in *Santalum album* L.: its biochemical characterization and significance. Plant Cell Physiol. 44(4):367-376.

Arimura G-i, Maffei ME. 2010. Calcium and secondary CPK signaling in plants in response to herbivore attack. Biochemical and Biophysical Research Communications. 400(4):455-460. doi:10.1016/j.bbrc.2010.08.134.

Arimura G-i, Sawasaki T. 2010. Arabidopsis CPK3 plays extensive roles in various biological and environmental responses. Plant signaling & behavior. 5(10):1263-1265.

Arthur K, Valmonte G, Higgins C, MacDiarmid RM. 2012. A classification tool for calcium dependent protein kinases (CPKs) based on motif analysis. XV International Congress on Molecular Plant- Microbe Interactions. Kyoto, Japan.

Asai S, Ichikawa T, Nomura H, Kobayashi M, Kamiyoshihara Y, Mori H, Kadota Y, Zipfel C, Jones JDG, Yoshioka H. 2013. The Variable Domain of a Plant Calcium-dependent Protein Kinase (CDPK) Confers Subcellular Localization and Substrate Recognition for NADPH Oxidase. Journal of Biological Chemistry. 288(20):14332-14340. doi:10.1074/jbc.M112.448910.

Asano T, Hakata M, Nakamura H, Aoki N, Komatsu S, Ichikawa H, Hirochika H, Ohsugi R. 2011. Functional characterisation of OsCPK21, a calcium-dependent protein kinase that confers salt tolerance in rice. Plant Molecular Biology. 75(1):179-191. doi:10.1007/s11103-010-9717-1.

Asano T, Hayashi N, Kobayashi M, Aoki N, Miyao A, Mitsuhara I, Ichikawa H, Komatsu S, Hirochika H, Kikuchi S et al. 2012. A rice calcium-dependent protein kinase OsCPK12 oppositely modulates salt-stress tolerance and blast disease resistance. Plant Journal. 69(1):26-36. doi:10.1111/j.1365-313X.2011.04766.x.

Asano T, Tanaka N, Yang G, Hayashi N, 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 and Cell Physiology. 46(2):356-366. doi:10.1093/pcp/pci035.

Asano T, Wakayama M, Aoki N, Komatsu S, Ichikawa H, Hirochika H, Ohsugi R. 2010. Overexpression of a calcium-dependent protein kinase gene enhances growth of rice under low-nitrogen conditions. Plant Biotechnology. 27(4):369-373.

Ashraf M, Akram NA, Arteca RN, Foolad MR. 2010. The physiological, biochemical and molecular roles of brassinosteroids and salicylic acid in plant processes and salt tolerance. Critical Reviews in Plant Sciences. 29(3):162-190. doi:10.1080/07352689.2010.483580.

Ayadi M, Delaporte V, Li Y-F, Zhou D-X. 2004. Analysis of GT-3a identifies a distinct subgroup of trihelix DNA-binding transcription factors in Arabidopsis. FEBS Letters. 562(1-3):147-154. doi:10.1016/S0014-5793(04)00222-4.

Babu M, Griffiths J, Huang T-S, Wang A. 2008. Altered gene expression changes in Arabidopsis leaf tissues and protoplasts in response to Plum pox virus infection. BMC Genomics. 9(1):325.

Baillie BK, Belda-Baillie CA, Maruyama T. 2000. A calcium-dependent protein kinase functions in wound healing in *Ventricaria ventricosa* (chlorophyta). Journal of Phycology. 36(6):1145-1152. doi:10.1046/j.1529-8817.2000.00050.x.

Baldauf SL. 2003. Phylogeny for the faint of heart: a tutorial. Trends in Genetics. 19(6):345-351. doi:10.1016/s0168-9525(03)00112-4.

Banks JA, Nishiyama T, Hasebe M, Bowman JL, Gribskov M, DePamphilis C, Albert VA, Aono N, Aoyama T, Ambrose BA et al. 2011. The selaginella genome identifies genetic changes associated with the evolution of vascular plants. Science. 332(6032):960-963.

Barrett T, Edgar R. 2006. Mining microarray data at NCBI's Gene Expression Omnibus (GEO). Methods in molecular biology (Clifton, NJ). 338:175-190.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2010. GenBank. Nucleic Acids Research. 38(suppl 1):D46-D51. doi:10.1093/nar/gkp1024.

Berman HM, Henrick K, Nakamura H. 2003. Announcing the worldwide protein data bank. Nature Structural Biology. 10(12):980.

Biddington NL. 1986. The effects of mechanically-induced stress in plants — a review. Plant Growth Regulation. 4(2):103-123. doi:10.1007/BF00025193.

Billker O, Lourido S, Sibley LD. 2009. Calcium-dependent signaling and kinases in apicomplexan parasites. Cell Host and Microbe. 5(6):612-622. doi:10.1016/j.chom.2009.05.017.

Bininda-Edmonds O. 2009. An introduction to phylogenetic analysis. Modern computational science 10 Lecture notes from the International Summer School Oldenburg.227-243. http://www.molekularesystematik.uni-oldenburg.de/download/Publications/phyloMCS.pdf.

Birkett MA, Campbell CAM, Chamberlain K, Guerrieri E, Hick AJ, Martin JL, Matthes M, Napier JA, Pettersson J, Pickett JA et al. 2000. New roles for cis-jasmone as an insect semiochemical and in plant defense. Proceedings of the National Academy of Sciences of the United States of America. 97(16):9329-9334. doi:10.1073/pnas.160241697.

Blouin A, Pearson MN, Chavan RR, Woo ENY, Lebas BSM, Veerakone S, C. R, Biccheri R, MacDiarmid RM. 2013. Viruses of kiwifruit (Actinidia species) Journal of plant pathology. (doi: 10.4454/JPP.V95I2.013).

Boorstein WR, Ziegelhoffer T, Craig EA. 1994. Molecular evolution of the Hsp70 multigene family. Journal of Molecular Evolution. 38(1):1-17.

Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T. 2009. Protein structure homology modeling using SWISS-MODEL workspace. Nature Protocols. 4(1):1-13. doi:10.1038/nprot.2008.197.

Botella JR, Arteca JM, Somodevilla M, Arteca RN. 1996. Calcium-dependent protein kinase gene expression in response to physical and chemical stimuli in mungbean (Vigna radiata). Plant Mol Biol. 30(6):1129-1137.

Boudsocq M, Droillard MJ, Regad L, Laurière C. 2012. Characterization of Arabidopsis calcium-dependent protein kinases: activated or not by calcium? Biochemical Journal. 447(2):291-299. doi:10.1042/bj20112072.

Boudsocq M, Sheen J. 2013. CDPKs in immune and stress signaling. Trends Plant Sci. 18(1):30-40. doi:10.1016/j.tplants.2012.08.008.

Boudsocq M, Willmann M, McCormack M, Lee H, Shan L, He P, Bush J, Cheng S-H, Sheen J. 2010. Differential immune signalling via calcium sensor protein kinases. Nature Letters. 464:418-423. doi:10.1038/nature08794.

Bowers JE, Chapman BA, Rong J, Paterson AH. 2003. Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. Nature. 422(6930):433-438.

Brooker RJ, Widmaier EP, Graham LE, Stiling PD. 2007. Biology. United States: McGraw-Hill Higher Education.

Buell CR, Joardar V, Lindeberg M, Selengut J, Paulsen IT, Gwinn ML, Dodson RJ, Deboy RT, Durkin AS, Kolonay JF et al. 2003. The complete genome sequence of the Arabidopsis and tomato pathogen Pseudomonas syringae pv. tomato DC3000. Proceedings of the National Academy of Sciences. 100(18):10181-10186. doi:10.1073/pnas.1731982100.

Bulley SM, Rassam M, Hoser D, Otto W, Schuenemann N, Wright M, MacRae E, Gleave A, Laing W. 2009. Gene expression studies in kiwifruit and gene over-expression in Arabidopsis indicates that GDP-L-galactose guanyltransferase is a major control point of vitamin C biosynthesis. Journal of Experimental Botany. 60(3):765-778. doi:10.1093/jxb/ern327.

Campbell N, Reece J. 2008. Biology. San Francisco: Pearson Benjamin Cummings.

Campos-Soriano L, Gomez-Ariza J, Bonfante P, San Segundo B. 2011. A rice calcium-dependent protein kinase is expressed in cortical root cells during the presymbiotic phase of the arbuscular mycorrhizal symbiosis. BMC Plant Biology. 11:90. doi:10.1186/1471-2229-11-90.

Cândido ES, Carmo LSTd, Campos PF, Santana B, Ávila ACd, Quirino BF. 2006. Evaluation of *Arabidopsis thaliana* response to infection by *Tomato spotted wilt virus* and *Groundnut ringspot virus*. Fitopatologia Brasileira. 31:101-101.

Cao J, Shi F. 2012. Evolution of the RALF Gene Family in Plants: Gene Duplication and Selection Patterns. Evolutionary Bioinformatics. 8:271-292. doi:10.4137/ebo.s9652.

Cecchini E, Al-Kaff NS, Bannister A, Giannakou ME, McCallum DG, Maule AJ, Milner JJ, Covey SN. 1998. Pathogenic interactions between variants of *Cauliflower mosaic virus* and *Arabidopsis thaliana*. Journal of Experimental Botany. 49(321):731-737. doi:10.1093/jxb/49.321.731.

Chandran V, Stollar EJ, Lindorff-Larsen K, Harper JF, Chazin WJ, Dobson CM, Luisi BF, Christodoulou J. 2006. Structure of the regulatory apparatus of a calcium-dependent protein kinase (CDPK): a novel mode of calmodulin-target recognition. Journal of Molecular Biology. 357(2):400-410. doi:10.1016/j.jmb.2005.11.093.

Chang W, Fu G, Chen X, Zhu J, Zhang Z. 2011. Biochemical characterization of a calcium-sensitive protein kinase LeCPK2 from tomato. Indian Journal of Biochemistry and Biophysics. 48(3):148-153.

Chang WJ, Su HS, Li WJ, Zhang ZL. 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(11):2427-2431.

Chaves MM, Maroco JP, Pereira JS. 2003. Understanding plant responses to drought - from genes to the whole plant. Functional Plant Biology. 30(3):239-264. doi:10.1071/fp02076.

Chen X, Wang Y, Lv B, Li J, Luo L, Lu S, Zhang X, Ma H, Ming F. 2014. The NAC family transcription factor OsNAP confers abiotic stress response through the ABA pathway. Plant and Cell Physiology. 55(3):604-619. doi:10.1093/pcp/pct204.

Cheng S-H, Willmann M, Chen H-C, Sheen J. 2002. Calcium signaling through protein kinases: The Arabidopsis calcium-dependent protein kinase gene family. Plant Physiology. 129:469-485. doi:10.1104/pp.005645.

Chico JM, Raices M, Tellez-Inon MT, Ulloa RM. 2002. A calcium-dependent protein kinase is systemically induced upon wounding in tomato plants. Plant Physiology. 128(1):256-270.

Chisholm ST, Coaker G, Day B, Staskawicz BJ. 2006. Host-microbe interactions: Shaping the evolution of the plant immune response. Cell. 124(4):803-814. doi:10.1016/j.cell.2006.02.008.

Choi HI, Park HJ, Park JH, Kim S, Im MY, Seo HH, Kim YW, Hwang I, Kim SY. 2005. Arabidopsis calcium-dependent protein kinase AtCPK32 interacts with ABF4, a transcriptional regulator of abscisic acid-responsive gene expression, and modulates its activity. Plant Physiology. 139(4):1750-1761. doi:10.1104/pp.105.069757.

Christodoulou J, Malmendal A, Harper JF, Chazin WJ. 2004. Evidence for differing roles for each lobe of the calmodulin-like domain in a calcium-dependent protein kinase. Journal of Biological Chemistry. 279(28):29092-29100. doi:10.1074/jbc.M401297200.

Chung E, Park JM, Oh SK, Joung YH, Lee S, Choi D. 2004. Molecular and biochemical characterization of the Capsicum annuum calcium-dependent protein kinase 3 (CaCDPK3) gene induced by abiotic and biotic stresses. Planta. 220(2):286-295. doi:10.1007/s00425-004-1372-9.

Ciésla A, Mituła F, Misztal L, Fedorowicz-Strónska O, Janicka S, Tajdel-Zielińska M, Marczak M, Janicki M, Ludwików A, Sadowski J. 2016. A role for barley calcium-dependent protein kinase CPK2a in the response to drought. Frontiers in Plant Science. 7(Oct). doi:10.3389/fpls.2016.01550.

Coca M, San Segundo B. 2010. AtCPK1 calcium-dependent protein kinase mediates pathogen resistance in Arabidopsis. Plant Journal. 63(3):526-540. doi:10.1111/j.1365-313X.2010.04255.x.

Cohn J, Sessa G, Martin GB. 2001. Innate immunity in plants. Current Opinion in Immunology. 13(1):55-62.

Conte MG, Gaillard S, Lanau N, Rouard M, Périn C. 2008. GreenPhylDB: A database for plant comparative genomics. Nucleic Acids Research. 36(SUPPL. 1):D991-D998. doi:10.1093/nar/gkm934.

Contreras-Porcia L, Dennett G, Gonzalez A, Vergara E, Medina C, Correa JA, Moenne A. 2011. Identification of Copper-Induced Genes in the Marine Alga Ulva compressa (Chlorophyta). Marine Biotechnology. 13(3):544-556. doi:10.1007/s10126-010-9325-8.

Contreras-Porcia L, Dennett G, González A, Vergara E, Medina C, Correa JA, Moenne A. 2010. Identification of Copper-Induced Genes in the Marine Alga Ulva compressa (Chlorophyta). Marine Biotechnology.1-13.

Cousson A. 2011. Arabidopsis Ca(2+)-dependent protein kinase CPK3 mediates relationship of putative inositol triphosphate receptor with slow-type anion channel. Biologia Plantarum. 55(3):507-521. doi:10.1007/s10535-011-0117-4.

Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible WR. 2005. Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. Plant Physiology. 139(1):5-17. doi:10.1104/pp.105.063743.

Dammann C, Ichida A, Hong B, Romanowsky S, Hrabak E, Harmon A, Pickard B, Harper J. 2003. Subcellular targeting of nine calcium-dependent protein kinase isoforms from Arabidopsis. Plant Physiology. 132(4):1840-1848. doi:10.1104/pp.103.020008.

Dar WD, Laxmipathi Gowda CL. 2013. Declining agricultural productivity and global food security. Journal of Crop Improvement. 27(2):242-254.

Dardick CD, Golem S, Culver JN. 2000. Susceptibility and symptom development in *Arabidopsis thaliana* to *Tobacco mosaic virus* is influenced by virus cell-to-cell movement. Molecular Plant-Microbe Interactions. 13(10):1139-1144. doi:10.1094/MPMI.2000.13.10.1139.

Das R, Pandey GK. 2010. Expressional analysis and role of calcium regulated kinases in abiotic stress signaling. Current Genomics. 11(1):2-13.

Daugaard M, Rohde M, Jaattela M. 2007. The heat shock protein 70 family: Highly homologous proteins with overlapping and distinct functions. Febs Letters. 581(19):3702-3710. doi:10.1016/j.febslet.2007.05.039.

Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, Thon M, Kulkarni R, Xu J-R, Pan H et al. 2005. The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature. 434(7036):980-986. doi:http://www.nature.com/nature/journal/v434/n7036/suppinfo/nature03449_S1.html.

Deeken R, Ache P, Kajahn I, Klinkenberg J, Bringmann G, Hedrich R. 2008. Identification of Arabidopsis thaliana phloem RNAs provides a search criterion for phloem-based transcripts hidden in complex datasets of microarray experiments. The Plant Journal. 55(5):746-759. doi:10.1111/j.1365-313X.2008.03555.x.

DeFalco TA, Bender KW, Snedden WA. 2010. Breaking the code: Ca2+ sensors in plant signalling. Biochemical Journal. 425(1):27-40.

Denancé N, Szurek B, Noël LD. 2014. Emerging functions of nodulin-like proteins in non-nodulating plant species. Plant and Cell Physiology. 55(3):469-474. doi:10.1093/pcp/pct198.

Des Marais DL, Juenger TE. 2010. Pleiotropy, plasticity, and the evolution of plant abiotic stress tolerance. Annals of the New York Academy of Sciences. 1206(1):56-79. doi:10.1111/j.1749-6632.2010.05703.x.

Deveaux Y, Toffano-Nioche C, Claisse G, Thareau V, Morin H, Laufs P, Moreau H, Kreis M, Lecharny A. 2008. Genes of the most conserved WOX clade in plants affect root and flower development in Arabidopsis. BMC Evolutionary Biology. 8(1). doi:10.1186/1471-2148-8-291.

Di C, Xu W, Su Z, Yuan JS. 2010. Comparative genome analysis of PHB gene family reveals deep evolutionary origins and diverse gene function. BMC Bioinformatics. 11. doi:10.1186/1471-2105-11-s6-s22.

Dodd AN, Kudla J, Sanders D. 2010. The Language of Calcium Signaling. In: Merchant S, Briggs WR, Ort D, editors. Annual Review of Plant Biology, Vol 61. Palo Alto: Annual Reviews. p. 593-620.

Dreher TW. 2004. *Turnip yellow mosaic virus:* transfer RNA mimicry, chloroplasts and a C-rich genome. Molecular Plant Pathology. 5(5):367-375. doi:10.1111/j.1364-3703.2004.00236.x.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evolutionary Biology. 7(1). doi:10.1186/1471-2148-7-214.

Du H, Wang Y-B, Xie Y, Liang Z, Jiang S-J, Zhang S-S, Huang Y-B, Tang Y-X. 2013. Genome-Wide Identification and Evolutionary and Expression Analyses of MYB-Related Genes in Land Plants. DNA Research. 20(5):437-448. doi:10.1093/dnares/dst021.

Dubrovina AS, Kiselev KV, Khristenko VS, Aleynova OA. 2015. VaCPK20, a calcium-dependent protein kinase gene of wild grapevine *Vitis amurensis* Rupr., mediates cold and drought stress tolerance. Journal of Plant Physiology. 185:1-12. doi:10.1016/j.jplph.2015.05.020.

Dubrovina AS, Kiselev KV, Khristenko VS, Aleynova OA. 2016a. The calcium-dependent protein kinase gene VaCPK29 is involved in grapevine responses to heat and osmotic stresses. Plant Growth Regulation.1-11. doi:10.1007/s10725-016-0240-5.

Dubrovina AS, Kiselev KV, Khristenko VS, Aleynova OA. 2016b. VaCPK21, a calcium-dependent protein kinase gene of wild grapevine *Vitis amurensis* Rupr., is involved in grape response to salt stress. Plant Cell, Tissue and Organ Culture. 124(1):137-150. doi:10.1007/s11240-015-0882-4.

Durrant WE, Dong X. 2004. Systemic acquired resistance. Annual Review of Phytopathology. 42(1):185-209. doi:doi:10.1146/annurev.phyto.42.040803.140421.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research. 32(5):1792-1797. doi:10.1093/nar/gkh340.

Eklöf JM, Brumer H. 2010. The XTH Gene Family: An Update on Enzyme Structure, Function, and Phylogeny in Xyloglucan Remodeling. Plant Physiology. 153(2):456-466. doi:10.1104/pp.110.156844.

Engelhorn J, Moreau F, Fletcher JC, Carles CC. 2014. ULTRAPETALA1 and LEAFY pathways function independently in specifying identity and determinacy at the Arabidopsis floral meristem. Annals of Botany. doi:10.1093/aob/mcu185.

Engstrom EM. 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 signaling & behavior. 6(6):850-854.

Erb M, Glauser G. 2010. Family Business: Multiple Members of Major Phytohormone Classes Orchestrate Plant Stress Responses. Chemistry-a European Journal. 16(34):10280-10289. doi:10.1002/chem.201001219.

Estruch JJ, Kadwell S, Merlin E, Crossland L. 1994. Cloning and characterization of a maize pollen-specific calcium-dependent calmodulin-independent protein kinase. Proceedings of the National Academy of Sciences of the United States of America. 91(19):8837-8841.

Fantino E, Segretin ME, Santin F, Mirkin FG, Ulloa RM. 2017. Analysis of the potato calcium-dependent protein kinase family and characterization of StCDPK7, a member induced upon infection with *Phytophthora infestans*. Plant Cell Reports.1-21. doi:10.1007/s00299-017-2144-x.

Fedorowicz-Strońska O, Koczyk G, Kaczmarek M, Krajewski P, Sadowski J. 2017. Genome-wide identification, characterisation and expression profiles of calcium-dependent protein kinase genes in barley (*Hordeum vulgare* L.). Journal of Applied Genetics. 58(1):11-22. doi:10.1007/s13353-016-0357-2.

Finer J, Taniya D. 2008. Transgenic plant production. In: Stewart C, editor. Plant biotechnology and genetics: Principles, techniques and applications. Hoboken, New Jersey: John Wiley & Sons, Inc. p. 245-274.

Franz S, Ehlert B, Liese A, Kurth J, Cazalé AC, Romeis T. 2011. Calcium-dependent protein kinase CPK21 functions in abiotic stress response in Arabidopsis thaliana. Molecular Plant. 4(1):83-96. doi:10.1093/mp/ssq064.

Frattini M, Morello L, Breviario D. 1999. Rice calcium-dependent protein kinase isoforms OsCDPK2 and OsCDPK11 show different responses to light and different expression patterns during seed development. Plant Molecular Biology. 41(6):753-764. doi:10.1023/a:1006316422400.

Fujita M, Fujita Y, Noutoshi Y, Takahashi F, Narusaka Y, Yamaguchi-Shinozaki K, Shinozaki K. 2006. Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. Current Opinion in Plant Biology. 9(4):436-442. doi:10.1016/j.pbi.2006.05.014.

Fukuzawa H, Kubo T, Yamano T. 2008. Genome of a green alga, Chlamydomonas reinhardtii, lights up key functions of plant and animal cells. Tanpakushitsu kakusan koso Protein, nucleic acid, enzyme. 53(9):1133-1143.

Galindo FG, Sjoholm I, Rasmusson AG, Widell S, Kaack K. 2007. Plant stress physiology: Opportunities and challenges for the food industry. Critical Reviews in Food Science and Nutrition. 47(8):749-763. doi:10.1080/10408390601062211.

Gargantini PR, Giammaria V, Grandellis C, Feingold SE, Maldonado S, Ulloa RM. 2009. Genomic and functional characterization of StCDPK1. Plant Molecular Biology. 70(1-2):153-172. doi:10.1007/s11103-009-9462-5.

Gechev TS, Hille J. 2012. Molecular basis of plant stress. Cellular and Molecular Life Sciences. 69(19):3161-3163. doi:10.1007/s00018-012-1086-2.

Gensel PG. 2008. The Earliest Land Plants. Annual Review of Ecology Evolution and Systematics. 39:459-477. doi:10.1146/annurev.ecolsys.39.110707.173526.

Giammaria V, Grandellis C, Bachmann S, Gargantini PR, Feingold SE, Bryan G, Ulloa RM. 2011. StCDPK2 expression and activity reveal a highly responsive potato calcium-dependent protein kinase involved in light signalling. Planta. 233(3):593-609. doi:10.1007/s00425-010-1319-2.

Gill N, Findley S, Walling JG, Hans C, Ma J, Doyle J, Stacey G, Jackson SA. 2009. Molecular and chromosomal evidence for allopolyploidy in soybean. Plant Physiology. 151(3):1167-1174.

Gogarten JP. 1994. Which is the most conserved group of proteins? Homology-orthology, paralogy, xenology, and the fusion of independent lineages. Journal of Molecular Evolution. 39(5):541-543.

Gong Q, Li P, Ma S, Indu Rupassara S, Bohnert HJ. 2005. Salinity stress adaptation competence in the extremophile *Thellungiella halophila* in comparison with its relative *Arabidopsis thaliana*. The Plant Journal. 44(t5):826-839. doi:10.1111/j.1365-313X.2005.02587.x.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N et al. 2012. Phytozome: a comparative platform for green plant genomics. Nucleic Acids Research. 40(D1):D1178-D1186. doi:10.1093/nar/gkr944.

Govrin EM, Levine A. 2002. Infection of Arabidopsis with a necrotrophic pathogen, Botrytis cinerea, elicits various defense responses but does not induce systemic acquired

resistance (SAR). Plant Molecular Biology. 48(3):267-276. doi:10.1023/a:1013323222095.

Grabarek Z. 2006. Structural Basis for Diversity of the EF-hand Calcium-binding Proteins. Journal of Molecular Biology. 359(3):509-525. doi:10.1016/j.jmb.2006.03.066.

Graves PR, Haystead TAJ. 2002. Molecular biologist's guide to proteomics. Microbiology and Molecular Biology Reviews. 66(1):39-63. doi:10.1128/mmbr.66.1.39-63.2002.

Gray J. 1985. The Microfossil Record of Early Land Plants - Advances in Understanding of Early Terrestrialization, 1970-1984. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences. 309(1138):167-&. doi:10.1098/rstb.1985.0077.

Gu C, Chen S, Liu Z, Shan H, Luo H, Guan Z, Chen F. 2011. Reference gene selection for quantitative real-time PCR in chrysanthemum subjected to biotic and abiotic stress. Molecular Biotechnology. 49(2):192-197.

Guex N, Peitsch MC, Schwede T. 2009. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. Electrophoresis. 30:S162-S173. doi:10.1002/elps.200900140.

Gullino ML, Gilardi G, Sanna M, Garibaldi A. 2009. Epidemiology of *Pseudomonas syringae* pv. *syringae* on tomato. Phytoparasitica. 37(5):461-466. doi:10.1007/s12600-009-0055-2.

Gust AA, Brunner F, Nürnberger T. 2010. Biotechnological concepts for improving plant innate immunity. Current Opinion in Biotechnology. 21(2):204-210.

Halim VA, Vess A, Scheel D, Rosahl S. 2006. The role of salicylic acid and jasmonic acid in pathogen defence. Plant Biology. 8(3):307-313. doi:10.1055/s-2006-924025.

Hamel L-P, Sheen J, Seguin A. 2014. Ancient signals: comparative genomics of green plant CDPKs. Trends in Plant Science. 19(2):79-89. doi:10.1016/j.tplants.2013.10.009.

Hanks SK, Hunter T. 1995. Protein Kinases .6. The Eukaryotic Protein-Kinase Superfamily - Kinase (Catalytic) Domain-Structure and Classification. Faseb Journal. 9(8):576-596.

Harmon AC, Gribskov M, Gubrium E, Harper JF. 2001. The CDPK superfamily of protein kinases. New Phytologist. 151(1):175-183. doi:10.1046/j.1469-8137.2001.00171.x.

Harmon AC, Gribskov M, Harper JF. 2000. CDPKs - A kinase for every Ca2+ signal? Trends in Plant Science. 5(4):154-159. doi:10.1016/s1360-1385(00)01577-6.

Harper JR, Breton G, Harmon A. 2004. Decoding calcium signals through plant protein kinases. p. 263-288.

Hasanuzzaman M, Nahar K, Fujita M. 2013. Extreme temperature responses, oxidative stress and antioxidant defense in plants. In: Vahdati K, Leslie C, editors. Abiotic Stress - Plant Responses and Applications in Agriculture. InTech.

Hashimoto K, Kudla J. 2011. Calcium decoding mechanisms in plants. Biochimie. 93:2054-2059. doi:10.1016/j.biochi.2011.05.019.

He X-J, Mu R-L, Cao W-H, Zhang Z-G, Zhang J-S, Chen S-Y. 2005. AtNAC2, a transcription factor downstream of ethylene and auxin signaling pathways, is involved in salt stress response and lateral root development. The Plant Journal. 44(6):903-916. doi:10.1111/j.1365-313X.2005.02575.x.

Hedges SB. 2002. The origin and evolution of model organisms. Nature Reviews Genetics. 3(11):838-849. doi:10.1038/nrg929.

Hedges SB, Blair JE, Venturi ML, Shoe JL. 2004. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. BMC Evolutionary Biology. 4.

Heidarvand L, Maali Amiri R. 2010. What happens in plant molecular responses to cold stress? Acta Physiologiae Plantarum. 32(3):419-431. doi:10.1007/s11738-009-0451-8.

Hettenhausen C, Sun G, He Y, Zhuang H, Sun T, Qi J, Wu J. 2016. Genome-wide identification of calcium-dependent protein kinases in soybean and analyses of their transcriptional responses to insect herbivory and drought stress. Scientific Reports. 6. doi:10.1038/srep18973.

Hochstenbach R, de Groot P, Jacobs J, Schrauwen J, Wullems G. 1996. The promoter of a gene that is expressed only in pollen interacts with ubiquitous transcription factors. Sexual Plant Reproduction. 9(4):197-202. doi:10.1007/BF02173098.

Hrabak E, Chan C, Gribskov M, Harper J, Choi J, Halford N, Kudla J, Luan S, Nimmo H, Sussman M et al. 2003. The Arabidopsis CDPK-SnRK superfamily of protein kinases. Plant Physiology. 132(2):666-680. doi:10.1104/pp.102.011999.

Hruz T, Laule O, Szabo G, Wessendorp F, Bleuler S, Oertle L, Widmayer P, Gruissem W, Zimmermann P. 2008. Genevestigator V3: a reference expression database for the meta-analysis of transcriptomes. Advances in Bioinformatics.420747.

Hu JS, Ferreira S, Xu MQ, Lu M, Iha M, Pflum E, Wang M. 1994. Transmission, movement and inactivation of *Cymbidium mosaic* and *Odontoglossum ringspot viruses*. Plant Disease. 78(6):633-636. doi:10.1094/PD-78-0633.

Hubbard KE, Siegel RS, Valerio G, Brandt B, Schroeder JI. 2012. Abscisic acid and CO2 signalling via calcium sensitivity priming in guard cells, new CDPK mutant phenotypes and a method for improved resolution of stomatal stimulus–response analyses. Annals of Botany. 109(1):5-17. doi:10.1093/aob/mcr252.

Hughes AL. 1994. The evolution of functionally novel proteins after gene duplication. Proceedings of the Royal Society B:Biological Sciences. 256(1346):119-124. doi:10.1098/rspb.1994.0058.

Immanuel TM, Greenwood DR, MacDiarmid RM. 2012. A critical review of translation initiation factor eIF2 alpha kinases in plants - regulating protein synthesis during stress. Functional Plant Biology. 39(9):717-735. doi:10.1071/fp12116.

Iriti M, Faoro F. 2007. Review of innate and specific immunity in plants and animals. Mycopathologia. 164(2):57-64.

Ishida S, Yuasa T, Nakata M, Takahashi Y. 2008. A tobacco calcium-dependent protein kinase, CDPK1, regulates the transcription factor REPRESSION OF SHOOT GROWTH in response to gibberellins. Plant Cell. 20(12):3273-3288. doi:10.1105/tpc.107.057489.

Ishiguro S, Nakamura K. 1994. Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5' upstream regions of genes coding for sporamin and beta-amylase from sweet potato. Molecular & general genetics : MGG. 244(6):563-571.

Ito M, Araki S, Matsunaga S, Itoh T, Nishihama R, Machida Y, Doonan JH, Watanabe A. 2001. G2/M-phase–specific transcription during the plant cell cycle is mediated by c-Myb–like transcription factors. The Plant Cell. 13(8):1891-1905. doi:10.1105/tpc.010102.

Ivashuta S, Liu J, Lohar DP, Haridas S, Bucciarelli B, VandenBosch KA, Vance CP, Harrison MJ, Gantt JS. 2005. RNA interference identifies a calcium-dependent protein kinase involved in Medicago truncatula root development. Plant Cell. 17(11):2911-2921. doi:10.1105/tpc.105.035394.

Izawa T, Foster R, Nakajima M, Shimamoto K, Chua NH. 1994. The rice bZIP transcriptional activator RITA-1 is highly expressed during seed development. The Plant Cell. 6(9):1277-1287. doi:10.1105/tpc.6.9.1277.

Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C et al. 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature. 449(7161):463-U465. doi:10.1038/nature06148.

Jain M, Pathak BP, Harmon AC, Tillman BL, Gallo M. 2011. Calcium dependent protein kinase (CDPK) expression during fruit development in cultivated peanut (Arachis hypogaea) under Ca2+-sufficient and -deficient growth regimens. Journal of Plant Physiology. 168(18):2272-2277. doi:10.1016/j.jplph.2011.07.005.

Jaworski K, Pawełek A, Szmidt-Jaworska A, Kopcewicz J. 2010. Expression of calcium-dependent protein kinase gene (PnCDPK1) is affected by various light conditions in Pharbitis nil seedlings. Journal of Plant Growth Regulation. 29(3):316-327.

Jia Y, Valent B, Lee FN. 2003. Determination of Host Responses to Magnaporthe grisea on Detached Rice Leaves Using a Spot Inoculation Method. Plant Disease. 87(2):129-133. doi:10.1094/PDIS.2003.87.2.129.

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS et al. 2011a. Ancestral polyploidy in seed plants and angiosperms. Nature. 473(7345):97-U113. doi:10.1038/nature09916.

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS et al. 2011b. Ancestral polyploidy in seed plants and angiosperms. Nature. 473(7345):97-100.

Jones JDG, Dangl JL. 2006. The plant immune system. Nature. 444(7117):323-329.

Kanchiswamy CN, Takahashi H, Quadro S, Maffei ME, Bossi S, Bertea C, Zebelo SA, Muroi A, Ishihama N, Yoshioka H et al. 2010. Regulation of Arabidopsis defense responses against Spodoptera littoralis by CPK-mediated calcium signaling. BMC Plant Biology. 10.

Kaplan-Levy RN, Brewer PB, Quon T, Smyth DR. 2013. The trihelix family of transcription factors: light, stress and development. Trends in Plant Science. 17(3):163-171. doi:10.1016/j.tplants.2011.12.002.

Kaplan F, Guy CL. 2004. β-amylase induction and the protective role of maltose during temperature shock. Plant Physiology. 135(3):1674-1684. doi:10.1104/pp.104.040808.

Katagiri F, Thilmony R, He SY. 2002. The *Arabidopsis thaliana-Pseudomonas syringae* interaction. The Arabidopsis Book / American Society of Plant Biologists. 1:e0039. doi:10.1199/tab.0039.

Katagiri F, Tsuda K. 2010. Understanding the Plant Immune System. Molecular Plant-Microbe Interactions. 23(12):1531-1536. doi:10.1094/mpmi-04-10-0099.

Kaur N, Gupta AK. 2005. Signal transduction pathways under abiotic stresses in plants. Current Science. 88(11):1771-1780.

Kawasaki H, Nakayama S, Kretsinger RH. 1998. Classification and evolution of EF-hand proteins. BioMetals. 11(4):277-295. doi:10.1023/a:1009282307967.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28(12):1647-1649.

Kenrick P, Crane PR. 1997. The origin and early evolution of plants on land. Nature. 389(6646):33-39. doi:10.1038/37918.

Kiefer F, Arnold K, Kuenzli M, Bordoli L, Schwede T. 2009. The SWISS-MODEL Repository and associated resources. Nucleic Acids Research. 37:D387-D392. doi:10.1093/nar/gkn750.

Kilian J, Peschke F, Berendzen KW, Harter K, Wanke D. 2012. Prerequisites, performance and profits of transcriptional profiling the abiotic stress response. Biochimica et biophysica acta. 1819(2):166-175.

Kilian J, Whitehead D, Horak J, Wanke D, Weinl S, Batistic O, D'Angelo C, Bornberg-Bauer E, Kudla J, Harter K. 2007. The AtGenExpress global stress expression data set: protocols, evaluation and model data analysis of UV-B light, drought and cold stress responses. The Plant Journal. 50(2):347-363. doi:10.1111/j.1365-313X.2007.03052.x.

Kim B, Masuta C, Matsuura H, Takahashi H, Inukai T. 2008. Veinal necrosis induced by *Turnip mosaic virus* infection in Arabidopsis is a form of defense response accompanying HR-like cell death. Molecular Plant-Microbe Interactions. 21(2):260-268. doi:10.1094/MPMI-21-2-0260.

Kiselev KV, Turlenko AV, Zhuravlev YN. 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, Tissue and Organ Culture. 103(2):197-204.

Klimecka M, Muszyńska G. 2007. Structure and functions of plant calcium-dependent protein kinases. Acta Biochimica Polonica. 54(2):219-233.

Kobayashi M, Ohura I, Kawakita K, Yokota N, Fujiwara M, Shimamoto K, Doke N, Yoshioka H. 2007. Calcium-dependent protein kinases regulate the production of reactive oxygen species by potato NADPH oxidase. Plant Cell. 19(3):1065-1080. doi:10.1105/tpc.106.048884.

Komatsu S, Yang G, Khan M, Onodera H, Toki S, Yamaguchi M. 2007. Over-expression of calcium-dependent protein kinase 13 and calreticulin interacting protein 1 confers cold tolerance on rice plants. Molecular Genetics and Genomics. 277(6):713-723. doi:10.1007/s00438-007-0220-6.

Konrad A, Teufel AI, Grahnen JA, Liberles DA. 2011. Toward a General Model for the Evolutionary Dynamics of Gene Duplicates. Genome Biology and Evolution. 3:1197-1209. doi:10.1093/gbe/evr093.

Kotak S, Larkindale J, Lee U, von Koskull-Döring P, Vierling E, Scharf K-D. 2007. Complexity of the heat stress response in plants. Current Opinion in Plant Biology. 10(3):310-316. doi:http://dx.doi.org/10.1016/j.pbi.2007.04.011.

Kugelstadt D, Winter D, Pluckhahn K, Lehmann WD, Kappes B. 2007. Raf kinase inhibitor protein affects activity of Plasmodium falciparum calcium-dependent protein kinase 1. Molecular and Biochemical Parasitology. 151(1):111-117. doi:10.1016/j.molbiopara.2006.10.012.

Kung S-d, Wu R. 1993. Transgenic plants. San Diego: Academic Press.

Lachaud C, Prigent E, Thuleau P, Grat S, Da Silva D, Briere C, Mazars C, Cotelle V. 2013. 14-3-3-Regulated Ca2+-dependent protein kinase CPK3 is required for sphingolipid-induced cell death in Arabidopsis. Cell Death and Differentiation. 20(2):209-217. doi:10.1038/cdd.2012.114.

Laloum T, De Mita S, Gamas P, Baudin M, Niebel A. 2013. CCAAT-box binding transcription factors in plants: Y so many? Trends in Plant Science. 18(3):157-166. doi:10.1016/j.tplants.2012.07.004.

Lanfear R, Welch JJ, Bromham L. 2010. Watching the clock: Studying variation in rates of molecular evolution between species. Trends in Ecology & Evolution. 25(9):495-503. doi:http://dx.doi.org/10.1016/j.tree.2010.06.007.

Lapierre H, Signoret P. 2004. Virus and virus diseases of Poaceae. American Phytopathological Society Press.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R et al. 2007. Clustal W and Clustal X version 2.0. Bioinformatics. 23(21):2947-2948.

Latz A, Mehlmer N, Zapf S, Mueller TD, Wurzinger B, Pfister B, Csaszar E, Hedrich R, Teige M, Becker D. 2013. Salt Stress Triggers Phosphorylation of the Arabidopsis Vacuolar K

Channel TPK1 by Calcium-Dependent Protein Kinases (CDPKs). Molecular Plant. 6(4):1274-1289. doi:10.1093/mp/sss158.

Laude AJ, Simpson AWM. 2009. Compartmentalized signalling: Ca2+ compartments, microdomains and the many facets of Ca2+ signalling. Febs Journal. 276(7):1800-1816. doi:10.1111/j.1742-4658.2009.06927.x.

Lawton MA, Dean SM, Dron M, Kooter JM, Kragh KM, Harrison MJ, Yu L, Tanguay L, Dixon RA, Lamb CJ. 1991. Silencer region of a chalcone synthase promoter contains multiple binding sites for a factor, SBF-1, closely related to GT-1. Plant Molecular Biology. 16(2):235-249. doi:10.1007/BF00020555.

Lee SC, Luan S. 2012. ABA signal transduction at the crossroad of biotic and abiotic stress responses. Plant, Cell and Environment. 35(1):53-60. doi:10.1111/j.1365-3040.2011.02426.x.

Leibfried A, To JPC, Busch W, Stehling S, Kehle A, Demar M, Kieber JJ, Lohmann JU. 2005. WUSCHEL controls meristem function by direct regulation of cytokinin-inducible response regulators. Nature. 438(7071):1172-1175. doi:http://www.nature.com/nature/journal/v438/n7071/suppinfo/nature04270_S1.html.

Li A, Wang X, Leseberg CH, Jia J, Mao L. 2008a. Biotic and abiotic stress responses through calcium-dependent protein kinase (CDPK) signaling in wheat (Triticum aestivum L.). Plant Signaling and Behavior. 3(9):654-656.

Li AL, Zhu YF, Tan XM, Wang X, Wei B, Guo HZ, Zhang ZL, Chen XB, Zhao GY, Kong XY et al. 2008b. Evolutionary and functional study of the CDPK gene family in wheat (Triticum aestivum L.). Plant Molecular Biology. 66(4):429-443.

Li H, Cai J, Liu F, Jiang D, Dai T, Cao W. 2012. Generation and scavenging of reactive oxygen species in wheat flag leaves under combined shading and waterlogging stress. Functional Plant Biology. 39(1):71-81. doi:10.1071/fp11165.

Li J, Liang D, Li M, Ma F. 2013. Light and abiotic stresses regulate the expression of GDP-L-galactose phosphorylase and levels of ascorbic acid in two kiwifruit genotypes via light-responsive and stress-inducible cis-elements in their promoters. Planta. 238(3):535-547. doi:10.1007/s00425-013-1915-z.

Li M, Ma F, Liang D, Li J, Wang Y. 2010. Ascorbate Biosynthesis during Early Fruit Development Is the Main Reason for Its Accumulation in Kiwi. PLoS ONE. 5(12):e14281. doi:10.1371/journal.pone.0014281.

Lilly S. 2014. An investigation of molecular indicators of plant virus infection. [Auckland, New Zealand]: University of Auckland.

Lilly ST, Drummond RSM, Pearson MN, MacDiarmid RM. 2011. Identification and Validation of Reference Genes for Normalization of Transcripts from Virus-Infected Arabidopsis thaliana. Molecular Plant-Microbe Interactions. 24(3):294-304. doi:10.1094/mpmi-10-10-0236.

Lim DC, Cooke BM, Doerig C, Saeij JPJ. 2012. Toxoplasma and Plasmodium protein kinases: Roles in invasion and host cell remodelling. International Journal for Parasitology. 42(1):21-32.

Liu G, Chen J, Wang X. 2006. VfCPK1, a gene encoding calcium-dependent protein kinase from Vicia faba, is induced by drought and abscisic acid. Plant Cell Environ. 29(11):2091-2099. doi:10.1111/j.1365-3040.2006.01582.x.

Liu J-X, Srivastava R, Che P, Howell SH. 2007. Salt stress responses in Arabidopsis utilize a signal transduction pathway related to endoplasmic reticulum stress signaling. The Plant Journal. 51(5):897-909. doi:10.1111/j.1365-313X.2007.03195.x.

Lovdal T, Lillo C. 2009. Reference gene selection for quantitative real-time PCR normalization in tomato subjected to nitrogen, cold, and light stress. Analytical Biochemistry. 387(2):238-242. doi:10.1016/j.ab.2009.01.024.

Lu B, Ding R, Zhang L, Yu X, Huang B, Chen W. 2006. Molecular cloning and characterization of a novel calcium-dependent protein kinase gene IiCPK2 Responsive to polyploidy from tetraploid Isatis indigotica. J Biochem Mol Biol. 39(5):607-617.

Lu SX, Hrabak E. 2002. An Arabidopsis calcium-dependent protein kinase is associated with the endoplasmic reticulum. Plant Physiol. 128(3):1008-1021. doi:10.1104/pp.010770.

Ma SY, Wu WH. 2007. AtCPK23 functions in Arabidopsis responses to drought and salt stresses. Plant Molecular Biology. 65(4):511-518. doi:10.1007/s11103-007-9187-2.

Magallon S, Castillo A. 2009. Angiosperm Diversification through Time. American Journal of Botany. 96(1):349-365. doi:10.3732/ajb.0800060.

Maksimov IV. 2009. Abscisic acid in the plants-pathogen interaction. Russian Journal of Plant Physiology. 56(6):742-752. doi:10.1134/s102144370906003x.

Maksup S, Supaibulwatana K, Selvaraj G. 2013. High-quality reference genes for quantifying the transcriptional responses of Oryza sativa L. (ssp indica and japonica) to abiotic stress

conditions. Chinese Science Bulletin. 58(16):1919-1930. doi:10.1007/s11434-013-5726-1.

Mantri NL, Ford R, Coram TE, Pang ECK. 2010. Evidence of unique and shared responses to major biotic and abiotic stresses in chickpea. Environmental and Experimental Botany. 69(3):286-292. doi:10.1016/j.envexpbot.2010.05.003.

Martin ML, Busconi L. 2001. A rice membrane-bound calcium-dependent protein kinase is activated in response to low temperature. Plant Physiol. 125(3):1442-1449.

Martinière A, Zancarini A, Drucker M. 2009. Aphid transmission of *Cauliflower mosaic virus*. Plant Signaling & Behavior. 4(6):548-550. doi:10.4161/psb.4.6.8712.

Mascia T, Santovito E, Gallitelli D, Cillo F. 2010. Evaluation of reference genes for quantitative reverse-transcription polymerase chain reaction normalization in infected tomato plants. Molecular Plant Pathology. 11(6):805-816. doi:10.1111/j.1364-3703.2010.00646.x.

McCurdy DW, Harmon AC. 1992. Calcium-dependent protein kinase in the green alga Chara. Planta. 188(1):54-61.

McLaughlin SB, Wimmer R. 1999. Tansley Review No. 104 - Calcium physiology and terrestrial ecosystem processes. New Phytologist. 142(3):373-417. doi:10.1046/j.1469-8137.1999.00420.x.

Mehlmer N, Wurzinger B, Stael S, Hofmann-Rodrigues D, Csaszar E, Pfister B, Bayer R, Teige M. 2010. The Ca2+-dependent protein kinase CPK3 is required for MAPK-independent salt-stress acclimation in Arabidopsis. Plant Journal. 63(3):484-498. doi:10.1111/j.1365-313X.2010.04257.x.

Melcher U. 1989. Symptoms of *Cauliflower mosaic virus* infection in *Arabidopsis thaliana* and turnip. Botanical Gazette. 150(2):139-147.

Menke FL, Champion A, Kijne JW, Memelink J. 1999. A novel jasmonate- and elicitor-responsive element in the periwinkle secondary metabolite biosynthetic gene Str interacts with a jasmonate- and elicitor-inducible AP2-domain transcription factor, ORCA2. The EMBO Journal. 18(16):4455-4463. doi:10.1093/emboj/18.16.4455.

Migocka M, Papierniak A. 2011. Identification of suitable reference genes for studying gene expression in cucumber plants subjected to abiotic stress and growth regulators. Molecular Breeding. 28(3):343-357. doi:10.1007/s11032-010-9487-0.

Ming R, Yu Q, Moore PH, Paull RE, Chen NJ, Wang ML, Zhu YJ, Schuler MA, Jiang J, Paterson AH. 2012. Genome of papaya, a fast growing tropical fruit tree. Tree Genetics and Genomes. 8(3):445-462.

Minitab 17 Statistical Software. 2010. State College: PA: Minitab, Inc. (www.minitab.com).

Mitra D, Johri MM. 2000. Enhanced expression of a calcium-dependent protein kinase from the moss Funaria hygrometrica under nutritional starvation. Journal of Biosciences. 25(4):331-338.

Miura K, Furumoto T. 2013. Cold Signaling and Cold Response in Plants. International Journal of Molecular Sciences. 14(3):5312.

Mohanta TK, Mohanta N, Mohanta YK, Bae H. 2015. Genome-wide identification of calcium dependent protein kinase gene family in plant lineage shows presence of novel D-x-D and D-E-L motifs in EF-hand domain. Frontiers in Plant Science. 6(DEC). doi:10.3389/fpls.2015.01146.

Morello L, Frattini M, Giani S, Christou P, Breviario D. 2000. Overexpression of the calcium-dependent protein kinase OsCDPK2 in transgenic rice is repressed by light in leaves and disrupts seed development. Transgenic Research. 9(6):453-462.

Mori IC, Murata Y, Yang Y, Munemasa S, Wang YF, Andreoli S, Tiriac H, Alonso JM, Harper JF, Ecker JR et al. 2006. CDPKs CPK6 and CPK3 function in ABA regulation of guard cell S-type anion- and Ca(2+)-permeable channels and stomatal closure. PLOS Biology. 4(10). doi:10.1371/journal.pbio.0040327.

Munemasa S, Hossain MA, Nakamura Y, Mori IC, Murata Y. 2011. The Arabidopsis Calcium-Dependent Protein Kinase, CPK6, Functions as a Positive Regulator of Methyl Jasmonate Signaling in Guard Cells. Plant Physiology. 155(1):553-561. doi:10.1104/pp.110.162750.

Murillo I, Jaeck E, Cordero MJ, San Segundo B. 2001. Transcriptional activation of a maize calcium-dependent protein kinase gene in response to fungal elicitors and infection. Plant Molecular Biology. 45(2):145-158.

Murphy ME. 2013. The HSP70 family and cancer. Carcinogenesis. 34(6):1181-1188. doi:10.1093/carcin/bgt111.

Myers C, Romanowsky SM, Barron YD, Garg S, Azuse CL, Curran A, Davis RM, Hatton J, Harmon AC, Harper JF. 2009. Calcium-dependent protein kinases regulate polarized tip growth in pollen tubes. Plant Journal. 59(4):528-539. doi:10.1111/j.1365-313X.2009.03894.x.

Nagamune K, Sibley LD. 2006. Comparative genomic and phylogenetic analyses of calcium ATPases and calcium-regulated proteins in the Apicomplexa. Molecular Biology and Evolution. 23(8):1613-1627.

Nakashima K, Yamaguchi-Shinozaki K. 2010. Promoters and transcription factors. In: Pareek A, Sopory S, Bohnert H, Govindjee, editors. Abiotic stress adaptation in plants. Dordrecht, The Netherlands: Springer.

Nam J, dePamphilis CW, Ma H, Nei M. 2003. Antiquity and Evolution of the MADS-Box Gene Family Controlling Flower Development in Plants. Molecular Biology and Evolution. 20(9):1435-1447. doi:10.1093/molbev/msg152.

Narsai R, Ivanova A, Ng S, Whelan J. 2010. Defining reference genes in Oryza sativa using organ, development, biotic and abiotic transcriptome datasets. BMC Plant Biology. 10. doi:10.1186/1471-2229-10-56.

Nie L, Wang R, Xia Y, Li G. 2015. CDPK1, an *Arabidopsis thaliana* calcium-dependent protein kinase, is involved in plant defense response. Russian Journal of Plant Physiology. 62(6):866-874. doi:10.1134/S1021443715070018.

Nishiyama R, Mizuno H, Okada S, Yamaguchi T, Takenaka M, Fukuzawa H, Ohyama K. 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(2):205-212.

Ohshima K, Yamaguchi Y, Hirota R, Hamamoto T, Tomimura K, Tan Z, Sano T, Azuhata F, Walsh JA, Fletcher J et al. 2002. Molecular evolution of *Turnip mosaic virus*: evidence of host adaptation, genetic recombination and geographical spread. Journal of General Virology. 83(6):1511-1521. doi:doi:10.1099/0022-1317-83-6-1511.

Ozfidan C, Turkan I, Sekmen AH, Seckin B. 2012. Abscisic acid-regulated responses of aba2-1 under osmotic stress: the abscisic acid-inducible antioxidant defence system and reactive oxygen species production. Plant Biology. 14(2):337-346.

Paolacci AR, Tanzarella OA, Porceddu E, Ciaffi M. 2009. Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. Bmc Molecular Biology. 10. doi:10.1186/1471-2199-10-11.

Parfrey LW, Grant J, Tekle YI, Lasek-Nesselquist E, Morrison HG, Sogin ML, Patterson DJ, Katz LA. 2010. Broadly sampled multigene analyses yield a well-resolved eukaryotic tree of life. Systematic Biology. 59(5):518-533. doi:10.1093/sysbio/syq037.

Parfrey LW, Lahr DJG, Knoll AH, Katz LA. 2011. Estimating the timing of early eukaryotic diversification with multigene molecular clocks. Proceedings of the National Academy of Sciences of the United States of America. 108(33):13624-13629. doi:10.1073/pnas.1110633108.

Parkinson H, Kapushesky M, Kolesnikov N, Rustici G, Shojatalab M, Abeygunawardena N, Berube H, Dylag M, Emam I, Farne A et al. 2009. ArrayExpress update - From an archive of functional genomics experiments to the atlas of gene expression. Nucleic Acids Research. 37(SUPPL. 1):D868-D872.

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A et al. 2009. The Sorghum bicolor genome and the diversification of grasses. Nature. 457(7229):551-556.

Paterson AH, Freeling M, Tang H, Wang X. 2010. Insights from the comparison of plant genome sequences. Annual Review of Plant Biology. 61:349-372.

Patharkar OR, Cushman JC. 2000. A stress-induced calcium-dependent protein kinase from Mesembryanthemum crystallinum phosphorylates a two-component pseudo-response regulator. Plant J. 24(5):679-691.

Pattengale N, Alipour M, Bininda-Edmonds O, Moret B, Stamatakis A. 2010. How many bootstrap replicates are necessary? Journal of Computational Biology. 17(3):337-354. doi:10.1089/cmb.2009.0179.

Pearson MN, Clover GRG, Guy PL, Fletcher JD, Beever RE. 2006. A review of the plant virus, viroid and mollicute records for New Zealand. Australasian Plant Pathology. 35(2):217-252. doi:10.1071/AP06016.

Pevsner J. 2009. Bioinformatics and functional genomics. New Jersey: John Wiley & Sons, Inc.

Plattner H, Verkhratsky A. 2015. The ancient roots of calcium signalling evolutionary tree. Cell Calcium. 57(3):123-132. doi:http://dx.doi.org/10.1016/j.ceca.2014.12.004.

Preston JC, Hileman LC. 2013. Functional Evolution in the Plant SQUAMOSA-PROMOTER BINDING PROTEIN-LIKE (SPL) Gene Family. Frontiers in Plant Science. 4:80-80. doi:10.3389/fpls.2013.00080.

Qin F, Shinozaki K, Yamaguchi-Shinozaki K. 2011. Achievements and Challenges in Understanding Plant Abiotic Stress Responses and Tolerance. Plant and Cell Physiology. 52(9):1569-1582. doi:10.1093/pcp/pcr106.

Quandt K, Frech K, Karas H, Wingender E, Werner T. 1995. MatInd and MatInspector: New fast and versatile tools for detection of consensus matches in nucleotide sequence data. Nucleic Acids Research. 23(23):4878-4884. doi:10.1093/nar/23.23.4878.

Radhamony RN, Prasad AM, Srinivasan R. 2005. T-DNA insertional mutagenesis in Arabidopsis: A tool for functional genomics. Electronic Journal of Biotechnology. 8(1):82-106.

Raices M, Chico JM, Tellez-Inon MT, Ulloa RM. 2001. Molecular characterization of StCDPK1, a calcium-dependent protein kinase from Solanum tuberosum that is induced at the onset of tuber development. Plant Molecular Biology. 46(5):591-601.

Ralph SG, Chun HJE, Kolosova N, Cooper D, Oddy C, Ritland CE, Kirkpatrick R, Moore R, Barber S, Holt RA 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.

Raven JA, Edwards D. 2001. Roots: evolutionary origins and biogeochemical significance. Journal of Experimental Botany. 52:381-401.

Ray S, Agarwal P, Arora R, Kapoor S, Tyagi AK. 2007. Expression analysis of calcium-dependent protein kinase gene family during reproductive development and abiotic stress conditions in rice (Oryza sativa L. ssp. indica). Molecular Genetics and Genomics. 278(5):493-505. doi:10.1007/s00438-007-0267-4.

Reddy A, Ali G, Celesnik H, Day I. 2011. Coping with Stresses: Roles of Calcium- and Calcium/Calmodulin-Regulated Gene Expression. Plant Cell. 23(6):2010-2032. doi:10.1105/tpc.111.084988.

Reddy V, Reddy A. 2004. Proteomics of calcium-signaling components in plants. Phytochemistry. 65(12):1745-1776. doi:10.1016/j.phytochem.2004.04.033.

Rensing SA, Lang D, Zimmer AD, Terry A, Salamov A, Shapiro H, Nishiyama T, Perroud P-F, Lindquist EA, Kamisugi Y et al. 2008. The Physcomitrella genome reveals evolutionary insights into the conquest of land by plants. Science. 319(5859):64-69. doi:10.1126/science.1150646.

Reyes JC, Muro-Pastor MI, Florencio FJ. 2004. The GATA family of transcription factors in Arabidopsis and rice. Plant Physiology. 134(4):1718-1732. doi:10.1104/pp.103.037788.

Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G, Montoya M et al. 2003. The Arabidopsis Information Resource (TAIR): A model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. Nucleic Acids Research. 31(1):224-228.

Roberti M, Bruni F, Loguercio Polosa P, Manzari C, Gadaleta MN, Cantatore P. 2006. MTERF3, the most conserved member of the mTERF-family, is a modular factor involved in mitochondrial protein synthesis. Biochimica et Biophysica Acta - Bioenergetics. 1757(9-10):1199-1206. doi:10.1016/j.bbabio.2006.04.026.

Rodriguez M, Canales E, Borras-Hidalgo O. 2005. Molecular aspects of abiotic stress in plants. Biotecnologia Applicada. 22(1):1-10.

Rouard M, Guignon V, Aluome C, Laporte MA, Droc G, Walde C, Zmasek CM, Périn C, Conte MG. 2011. GreenPhylDB v2.0: Comparative and functional genomics in plants. Nucleic Acids Research. 39(SUPPL. 1):D1095-D1102. doi:10.1093/nar/gkq811.

Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protocols. 5(4):725-738.

Rutschmann F, Stalder U, Piotrowski M, Oecking C, Schaller A. 2002. LeCPK1, a calcium-dependent protein kinase from tomato. Plasma membrane targeting and biochemical characterization. Plant Physiology. 129(1):156-168. doi:10.1104/pp.000869.

Sadowski MI, Jones DT. 2009. The sequence-structure relationship and protein function prediction. Current Opinion in Structural Biology. 19(3):357-362.

Saijo Y, Hata S, Kyozuka J, Shimamoto K, Izui K. 2000. Over-expression of a single Ca2+-dependent protein kinase confers both cold and salt/drought tolerance on rice plants. Plant Journal. 23(3):319-327. doi:10.1046/j.1365-313X.2000.00787.x.

Saijo Y, Kinoshita N, Ishiyama K, Hata S, Kyozuka J, Hayakawa T, Nakamura T, Shimamoto K, Yamaya T, Izui K. 2001. A Ca2+-dependent protein kinase that endows rice plants with cold- and salt-stress tolerance functions in vascular bundles. Plant and Cell Physiology. 42(11):1228-1233.

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J et al. 2010. Genome sequence of the palaeopolyploid soybean. Nature. 463(7278):178-183.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA et al. 2009. The B73 maize genome: Complexity, diversity, and dynamics. Science. 326(5956):1112-1115.

Schoelz JE, Shepherd RJ. 1988. Host range control of *Cauliflower mosaic virus*. Virology. 162(1):30-37. doi:http://dx.doi.org/10.1016/0042-6822(88)90391-1.

Scholthof K-BG. 2004. *Tobacco mosaic virus*: a model system for plant biology. Annual Review of Phytopathology. 42(1):13-34. doi:10.1146/annurev.phyto.42.040803.140322.

Senthilkumar R, Yeh K-W. 2012. Multiple biological functions of sporamin related to stress tolerance in sweet potato (*Ipomoea batatas* Lam). Biotechnology Advances. 30(6):1309-1317. doi:http://dx.doi.org/10.1016/j.biotechadv.2012.01.022.

Shao H-B, Jiang S-Y, Li F-M, Chu L-Y, Zhao C-X, Shao M-A, Zhao X-N, Li F. 2007. Some advances in plant stress physiology and their implications in the systems biology era. Colloids and Surfaces B-Biointerfaces. 54(1):33-36. doi:10.1016/j.colsurfb.2006.05.011.

Sheen J. 1996. Ca2+-dependent protein kinases and stress signal transduction in plants. Science. 274(5294):1900-1902. doi:10.1126/science.274.5294.1900.

Sherwood J, German T, Moyer J, Ullman D. 2003. *Tomato spotted wilt virus*. The Plant Health Instructor. http://www.apsnet.org/edcenter/intropp/lessons/viruses/Pages/TomatoSpottedWilt.aspx. doi:10.1094/PHI-I-2003-0613-02.

Shiboleth YM, Haronsky E, Leibman D, Arazi T, Wassenegger M, Whitham SA, Gaba V, Gal-On A. 2007. The conserved FRNK box in HC-Pro, a plant viral suppressor of gene silencing, is required for small RNA binding and mediates symptom development. Journal of Virology. 81(23):13135-13148. doi:10.1128/JVI.01031-07.

Shin DH, Hou J, Chandonia JM, Das D, Choi IG, Kim R, Kim SH. 2007. Structure-based inference of molecular functions of proteins of unknown function from Berkeley Structural Genomics Center. Journal of Structural and Functional Genomics. 8(2-3):99-105.

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, dePamphilis CW, Wall PK, Soltis PS. 2009. Polyploidy and Angiosperm Diversification. American Journal of Botany. 96(1):336-348. doi:10.3732/ajb.0800079.

Sousa AdS, Rego EJL, Santos FdARd. 2013. Viability and Action of CPL Lectin on &lt;i&gt;in Vitro&lt;/i&gt; Germinability of Pollen Grains of &lt;i&gt;Malpighia emarginata&lt;/i&gt; DC.¡ª(Malpighiaceae). American Journal of Plant Sciences. Vol.04No.07:6. doi:10.4236/ajps.2013.47A1007.

Staats M, van Kan JAL. 2012. Genome update of *Botrytis cinerea* strains B05.10 and T4. Eukaryotic Cell. 11(11):1413-1414. doi:10.1128/ec.00164-12.

Stratmann JW. 2003. Long distance run in the wound response - jasmonic acid is pulling ahead. Trends in Plant Science. 8(6):247-250. doi:10.1016/s1360-1385(03)00106-7.

Sugiyama K-i, Mori IC, Takahashi K, Muto S, Shihira-Ishikawa I. 2000. A calcium-dependent protein kinase functions in wound healing in *Ventricaria ventricosa* (Chlorophyta). Journal of Phycology. 36(6):1145-1152. doi:10.1046/j.1529-8817.2000.00050.x.

Suntio T, MÄKinen K. 2012. Abiotic stress responses promote Potato virus A infection in Nicotiana benthamiana. Molecular Plant Pathology.no-no. doi:10.1111/j.1364-3703.2012.00786.x.

Swarbreck D, Wilks C, Lamesch P, Berardini TZ, Garcia-Hernandez M, Foerster H, Li D, Meyer T, Muller R, Ploetz L et al. 2008. The Arabidopsis Information Resource (TAIR): Gene structure and function annotation. Nucleic Acids Research. 36(SUPPL. 1):D1009-D1014. doi:10.1093/nar/gkm965.

Swatek KN, Wilson RS, Ahsan N, Tritz RL, Thelen JJ. 2014. Multisite phosphorylation of 14-3-3 proteins by calcium-dependent protein kinases. Biochemical Journal. 459:15-25. doi:10.1042/bj20130035.

Syam Prakash SR, Jayabaskaran C. 2006. Expression and localization of calcium-dependent protein kinase isoforms in chickpea. Journal of Plant Physiology. 163(11):1135-1149. doi:10.1016/j.jplph.2006.04.002.

Szczegielniak J, Klimecka M, Liwosz A, Ciesielski A, Kaczanowski S, Dobrowolska G, Harmon AC, Muszyńska G. 2005. A wound-responsive and phospholipid-regulated maize calcium-dependent protein kinase. Plant Physiology. 139(4):1970-1983. doi:10.1104/pp.105.066472.

Tai Ss, Liu Gs, Sun Yh, Chen J. 2009. Cloning and Expression of Calcium-Dependent Protein Kinase (CDPK) Gene Family in Common Tobacco (Nicotiana tabacum). Agricultural Sciences in China. 8(12):1448-1457. doi:10.1016/s1671-2927(08)60358-2.

Taji T, Seki M, Satou M, Sakurai T, Kobayashi M, Ishiyama K, Narusaka Y, Narusaka M, Zhu J-K, Shinozaki K. 2004. Comparative genomics in salt tolerance between Arabidopsis and Arabidopsis-related halophyte salt cress using Arabidopsis microarray. Plant Physiology. 135(3):1697-1709. doi:10.1104/pp.104.039909.

Takezawa D, Patil S, Bhatia A, Poovaiah BW. 1996. Calcium-dependent protein kinase genes in corn roots. Journal of Plant Physiology. 149:329-335.

Tang H, Wang X, Bowers JE, Ming R, Alam M, Paterson AH. 2008. Unraveling ancient hexaploidy through multiply-aligned angiosperm gene maps. Genome Research. 18(12):1944-1954.

Taylor TN, Kerp H, Hass H. 2005. Life history biology of early land plants: Deciphering the gametophyte phase. Proceedings of the National Academy of Sciences of the United States of America. 102(16):5892-5897.

TeBeest D, Guerber C, Ditmore M. 2007. Rice blast. The Plant Health Instructor. http://www.apsnet.org/edcenter/intropp/lessons/fungi/ascomycetes/Pages/RiceBlast.aspx. doi:10.1094/PHI-I-2007-0313-07.

Tsai TM, Chen YR, Kao TW, Tsay WS, Wu CP, Huang DD, Chen WH, Chang CC, Huang HJ. 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(10):1899-1908. doi:10.1007/s00299-007-0389-5.

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A et al. 2006. The Genome of Black Cottonwood, Populus trichocarpa (Torr. & Gray). Science. 313(5793):1596-1604. doi:10.1126/science.1128691.

Tzfira T, Li J, Lacroix Bt, Citovsky V. 2004. Agrobacterium T-DNA integration: molecules and models. Trends in Genetics. 20(8):375-383. doi:http://dx.doi.org/10.1016/j.tig.2004.06.004.

Ulloa RM, Raices M, MacIntosh GC, Maldonado S, Tellez-Inon MT. 2002a. Jasmonic acid affects plant morphology and calcium-dependent protein kinase expression and activity in Solanum tuberosum. Physiol Plant. 115(3):417-427.

Ulloa RM, Raíces M, MacIntosh GC, Maldonado S, Téllez-Iñón MT. 2002b. Jasmonic acid affects plant morphology and calcium-dependent protein kinase expression and activity in solanum tuberosum. Physiologia Plantarum. 115(3):417-427.

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM. 2007. Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Research. 35:W71-W74. doi:10.1093/nar/gkm306.

Valmonte GR, Arthur K, Higgins CM, MacDiarmid RM. 2014. Calcium-Dependent Protein Kinases in Plants: Evolution, Expression and Function. Plant and Cell Physiology. 55(3):551-569. doi:10.1093/pcp/pct200.

Valmonte GR, Lilly S, Higgins CM, MacDiarmid RM. 2015. Can plants tell you if they are infected with viruses? Paper presented at: Australasian Plant Pathology Society Conference 2015. Fremantle, Western Australia.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biology. 3(7):research0034.0031 - research0034.0011.

Walley JW, Coughlan S, Hudson ME, Covington MF, Kaspi R, Banu G, Harmer SL, Dehesh K. 2007. Mechanical stress induces biotic and abiotic stress responses via a novel *cis*-element. PLoS Genetics. 3(10):e172. doi:10.1371/journal.pgen.0030172.

Walton EF, Wu R-M, Richardson AC, Davy M, Hellens RP, Thodey K, Janssen BJ, Gleave AP, Rae GM, Wood M et al. 2009. A rapid transcriptional activation is induced by the dormancy-breaking chemical hydrogen cyanamide in kiwifruit (Actinidia deliciosa) buds. Journal of Experimental Botany. 60(13):3835-3848. doi:10.1093/jxb/erp231.

Wang CT, Song W. 2014. ZmCK3, a maize calcium-dependent protein kinase gene, endows tolerance to drought and heat stresses in transgenic Arabidopsis. Journal of Plant Biochemistry and Biotechnology. 23(3):249-256. doi:10.1007/s13562-013-0208-8.

Wang H, Mei W, Qin Y, Zhu Y. 2011a. 1-Aminocyclopropane-1-carboxylic acid synthase 2 is phosphorylated by calcium-dependent protein kinase 1 during cotton fiber elongation. Acta Biochimica Et Biophysica Sinica. 43(8):654-661. doi:10.1093/abbs/gmr056.

Wang H, Mei W, Qin Y, Zhu Y. 2011b. 1-Aminocyclopropane-1-carboxylic acid synthase 2 is phosphorylated by calcium-dependent protein kinase 1 during cotton fiber elongation. Acta Biochim Biophys Sin (Shanghai). 43(8):654-661. doi:10.1093/abbs/gmr056.

Wankhede DP, Kumari M, Richa T, Aravind J, Rajkumar S. 2017. Genome wide identification and characterization of calcium dependent protein kinase gene family in *Cajanus cajan*. Journal of Environmental Biology. 38(1):167-177.

Warren AS, Anandakrishnan R, Zhang L. 2010. Functional bias in molecular evolution rate of Arabidopsis thaliana. BMC Evolutionary Biology. 10. doi:10.1186/1471-2148-10-125.

Wasilewska A, Vlad F, Sirichandra C, Redko Y, Jammes F, Valon C, Frey NFd, Leung J. 2008. An update on abscisic acid signaling in plants and more. Molecular Plant. 1(2):198-217. doi:10.1093/mp/ssm022.

Weckwerth P, Ehlert B, Romeis T. 2014. ZmCPK1, a calcium-independent kinase member of the *Zea mays* CDPK gene family, functions as a negative regulator in cold stress signalling. Plant, Cell and Environment. doi:10.1111/pce.12414.

Wei S, Hu W, Deng X, Zhang Y, Liu X, Zhao X, Luo Q, Jin Z, Li Y, Zhou S et al. 2014. A rice calcium-dependent protein kinase OsCPK9 positively regulates drought stress tolerance and spikelet fertility. BMC Plant Biology. 14(1). doi:10.1186/1471-2229-14-133.

Wei X, Shen F, Hong Y, Rong W, Du L, Liu X, Xu H, Ma L, Zhang Z. 2016. The wheat calcium-dependent protein kinase TaCPK7-D positively regulates host resistance to sharp eyespot disease. Molecular Plant Pathology. 17(8):1252-1264. doi:10.1111/mpp.12360.

Weljie AM, Clarke TE, Juffer AH, Harmon AC, Vogel HJ. 2000. Comparative modeling studies of the calmodulin-like domain of calcium-dependent protein kinase from soybean. Proteins. 39(4):343-357.

Weljie AM, Gagne SM, Vogel HJ. 2004. Solution structure and backbone dynamics of the N-terminal region of the calcium regulatory domain from soybean calcium-dependent protein kinase alpha. Biochemistry. 43(48):15131-15140. doi:10.1021/bi048751r.

Weljie AM, Robertson KM, Vogel HJ. 2003. Conformational changes in the Ca2+-regulatory region from soybean calcium-dependent protein kinase-alpha: fluorescence resonance energy transfer studies. Journal of Biological Chemistry. 278(44):43764-43769. doi:10.1074/jbc.M306799200.

Weljie AM, Vogel HJ. 2004. Unexpected structure of the Ca2+-regulatory region from soybean calcium-dependent protein kinase-α. Journal of Biological Chemistry. 279(34):35494-35502. doi:10.1074/jbc.M311520200.

Wenkel S, Turck F, Singer K, Gissot L, Le Gourrierec J, Samach A, Coupland G. 2006. CONSTANS and the CCAAT box binding complex share a functionally important domain and interact to regulate flowering of Arabidopsis. The Plant Cell. 18(11):2971-2984. doi:10.1105/tpc.106.043299.

Wernimont AK, Amani M, Qiu W, Pizarro JC, Artz JD, Lin YH, Lew J, Hutchinson A, Hui R. 2011. Structures of parasitic CDPK domains point to a common mechanism of activation. Proteins: Structure, Function and Bioformatics. 79(3):803-820.

Wernimont AK, Artz JD, Finerty P, Lin YH, Amani M, Allali-Hassani A, Senisterra G, Vedadi M, Tempel W, MacKenzie F et al. 2010. Structures of apicomplexan calcium-dependent protein kinases reveal mechanism of activation by calcium. Nature Structural and Molecular Biology. 17(5):596-601.

Whisstock JC, Lesk AM. 2003. Prediction of protein function from protein sequence and structure. Quarterly Reviews of Biophysics. 36(3):307-340.

White EJ, Venter M, Hiten NF, Burger JT. 2008. Modified Cetyltrimethylammonium bromide method improves robustness and versatility: The benchmark for plant RNA extraction. Biotechnology Journal. 3(11):1424-1428. doi:10.1002/biot.200800207.

Willems E, Leyns L, Vandesompele J. 2008. Standardization of real-time PCR gene expression data from independent biological replicates. Analytical Biochemistry. 379(1):127-129. doi:http://dx.doi.org/10.1016/j.ab.2008.04.036.

Williamson B, Tudzynski B, Tudzynski P, Van Kan JAL. 2007. *Botrytis cinerea*: the cause of grey mould disease. Molecular Plant Pathology. 8(5):561-580. doi:10.1111/j.1364-3703.2007.00417.x.

Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ. 2007. An "Electronic Fluorescent Pictograph" Browser for Exploring and Analyzing Large-Scale Biological Data Sets. PLoS ONE. 2(8):e718. doi:10.1371/journal.pone.0000718.

Wise RP, Caldo RA, Hong L, Shen L, Cannon E, Dickerson JA. 2007. BarleyBase/PLEXdb: A unified expression profiling database for plants and plant pathogens. p. 347-363.

Woodhouse MR, Schnable JC, Pedersen BS, Lyons E, Lisch D, Subramaniam S, Freeling M. 2010. Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. PLoS Biology. 8(6).

Wu P, Wang W, Duan W, Li Y, Hou X. 2017. Comprehensive analysis of the CDPK-SnRK superfamily genes in chinese cabbage and its evolutionary implications in plants. Frontiers in Plant Science. 8. doi:10.3389/fpls.2017.00162.

Wu R-M, Walton EF, Richardson AC, Wood M, Hellens RP, Varkonyi-Gasic E. 2012. Conservation and divergence of four kiwifruit SVP-like MADS-box genes suggest distinct roles in kiwifruit bud dormancy and flowering. Journal of Experimental Botany. 63(2):797-807. doi:10.1093/jxb/err304.

Wurzinger B, Mair A, Pfister B, Teige M. 2011. Cross-talk of calcium-dependent protein kinase and MAP kinase signaling. Plant signaling & behavior. 6(1):8-12.

Xiao XH, Yang M, Sui JL, Qi JY, Fang YJ, Hu SN, Tang CR. 2017. The calcium-dependent protein kinase (CDPK) and CDPK-related kinase gene families in *Hevea brasiliensis*—

comparison with five other plant species in structure, evolution, and expression. FEBS Open Bio. 7(1):4-24. doi:10.1002/2211-5463.12163.

Xin X-F, He SY. 2013. *Pseudomonas syringae* pv. *tomato* DC3000: A model pathogen for probing disease susceptibility and hormone signaling in plants. Annual Review of Phytopathology. 51(1):473-498. doi:10.1146/annurev-phyto-082712-102321.

Xiong L, Zhu JK. 2002. Molecular and genetic aspects of plant responses to osmotic stress. Plant Cell and Environment. 25(2):131-139. doi:10.1046/j.1365-3040.2002.00782.x.

Xu J-H, Messing J. 2008. Organization of the prolamin gene family provides insight into the evolution of the maize genome and gene duplications in grass species. Proceedings of the National Academy of Sciences. 105(38):14330-14335. doi:10.1073/pnas.0807026105.

Xu J, Tian YS, Peng RH, Xiong AS, Zhu B, Jin XF, Gao F, Fu XY, Hou XL, Yao QH. 2010. AtCPK6, a functionally redundant and positive regulator involved in salt/drought stress tolerance in Arabidopsis. Planta. 231(6):1251-1260. doi:10.1007/s00425-010-1122-0.

Xu X, Pan S, Cheng S, Zhang B, Mu D, Ni P, Zhang G, Yang S, Li R, Wang J et al. 2011. Genome sequence and analysis of the tuber crop potato. Nature. 475(7355):189-195.

Ye S, Wang L, Xie W, Wan B, Li X, 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 Molecular Biology. 70(3):311-325. doi:10.1007/s11103-009-9475-0.

Yin X-r, Allan AC, Zhang B, Wu R-m, Burdon J, Wang P, Ferguson IB, Chen K-s. 2009. Ethylene-related genes show a differential response to low temperature during 'Hayward' kiwifruit ripening. Postharvest Biology and Technology. 52(1):9-15. doi:http://dx.doi.org/10.1016/j.postharvbio.2008.09.016.

Yoon GM, Dowd PE, Gilroy S, McCubbin AG. 2006. Calcium-dependent protein kinase isoforms in Petunia have distinct functions in pollen tube growth, including regulating polarity. Plant Cell. 18(4):867-878. doi:10.1105/tpc.105.037135.

Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. Molecular Biology and Evolution. 21(5):809-818. doi:10.1093/molbev/msh075.

Yu XC, Li MJ, Gao GF, Feng HZ, Geng XQ, Peng CC, Zhu SY, Wang XJ, Shen YY, Zhang DP. 2006. Abscisic acid stimulates a calcium-dependent protein kinase in grape berry. Plant Physiology. 140(2):558-579. doi:10.1104/pp.105.074971.

Yu XC, Zhu SY, Gao GF, Wang XJ, Zhao R, Zou KQ, Wang XF, Zhang XY, Wu FQ, Peng CC et al. 2007. Expression of a grape calcium-dependent protein kinase ACPK1 in Arabidopsis thaliana promotes plant growth and confers abscisic acid-hypersensitivity in germination, postgermination growth, and stomatal movement. Plant Molecular Biology. 64(5):531-538. doi:10.1007/s11103-007-9172-9.

Yuasa T, 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. Archives of Biochemistry and Biophysics. 296(1):175-182.

Zhang B, Chen K, Bowen J, Allan A, Espley R, Karunairetnam S, Ferguson I. 2006. Differential expression within the LOX gene family in ripening kiwifruit. Journal of Experimental Botany. 57(14):3825-3836. doi:10.1093/jxb/erl151.

Zhang H, Wei C, Yang X, Chen H, Yang Y, Mo Y, Li H, Zang Y, Ma J, Yang J et al. 2017. Genome-wide identification and expression analysis of calcium-dependent protein kinase and its related kinase gene families in melon (*Cucumis melo* L.). PLoS ONE. 12(4). doi:10.1371/journal.pone.0176352.

Zhang K, Han YT, Zhao FL, Hu Y, Gao YR, Ma YF, Zheng Y, Wang YJ, Wen YQ. 2015. Genome-wide Identification and Expression Analysis of the CDPK Gene Family in Grape, *Vitis* spp. BMC Plant Biology. 15(1):1. doi:10.1186/s12870-015-0552-z.

Zhang M, Liang S, Lu YT. 2005. Cloning and functional characterization of NtCPK4, a new tobacco calcium-dependent protein kinase. Biochimica et Biophysica Acta - Gene Structure and Expression. 1729(3):174-185. doi:10.1016/j.bbaexp.2005.04.006.

Zhang N, Zeng L, Shan H, Ma H. 2012. Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. New Phytologist. 195(4):923-937. doi:10.1111/j.1469-8137.2012.04212.x.

Zhang X, Zong J, Liu J, Yin J, Zhang D. 2010. Genome-wide analysis of WOX gene family in rice, sorghum, maize, arabidopsis and poplar. Journal of Integrative Plant Biology. 52(11):1016-1026. doi:10.1111/j.1744-7909.2010.00982.x.

Zhang XS, Choi JH. 2001. Molecular evolution of calmodulin-like domain protein kinases (CDPKs) in plants and protists. Journal of Molecular Evolution. 53(3):214-224. doi:10.1007/s002390010211.

Zhao LL, Xu QJ, Jiang Y, Li YH. 2008. The mitogen-activated protein kinase signal transduction in plant cell under biotic and abiotic stress conditions. Plant Physiology Communications. 44(1):169-174.

Zhou L, Fu Y, Yang Z. 2009. A genome-wide functional characterization of Arabidopsis regulatory calcium sensors in pollen tubes. Journal of Integrative Plant Biology. 51(8):751-761. doi:10.1111/j.1744-7909.2009.00847.x.

Zhou Y, Yang W, Kirberger M, Lee HW, Ayalasomayajula G, Yang JJ. 2006. Prediction of EF-hand calcium-binding proteins and analysis of bacterial EF-hand proteins. Proteins: Structure, Function and Genetics. 65(3):643-655. doi:10.1002/prot.21139.

Zhu JH, Chen X, Chang WJ, Tian WM, Zhang ZL. 2010. Molecular characterization of HbCDPK1, an ethephon-induced calcium-dependent protein kinase gene of Hevea brasiliensis. Biosci Biotechnol Biochem. 74(11):2183-2188.

Zhu SY, Yu XC, Wang XJ, Zhao R, Li Y, Fan RC, Shang Y, Du SY, Wang XF, Wu FQ et al. 2007. Two calcium-dependent protein kinases, CPK4 and CPK11, regulate abscisic acid signal transduction in Arabidopsis. Plant Cell. 19(10):3019-3036. doi:10.1105/tpc.107.050666.

Zhu X, Dunand C, Snedden W, Galaud J-P. 2015. CaM and CML emergence in the green lineage. Trends in Plant Science. 20(8):483-489. doi:http://dx.doi.org/10.1016/j.tplants.2015.05.010.

Zimmer A, Lang D, Richardt S, Frank W, Reski R, Rensing SA. 2007. Dating the early evolution of plants: detection and molecular clock analyses of orthologs. Molecular Genetics and Genomics. 278(4):393-402. doi:10.1007/s00438-007-0257-6.

Zitter T, Murphy J. 2009. *Cucumber mosaic virus*. The Plant Health Instructor. http://www.apsnet.org/edcenter/intropp/lessons/viruses/Pages/Cucumbermosaic.aspx. doi:10.1094/PHI-I-2009-0518-01.

Zou JJ, Wei FJ, Wang C, Wu JJ, Ratnasekera D, Liu WX, Wu WH. 2010. Arabidopsis calcium-dependent protein kinase cpk10 functions in abscisic acid- and Ca2+-mediated stomatal regulation in response to drought stress. Plant Physiology. 154(3):1232-1243. doi:10.1104/pp.110.157545.

Zuo R, Hu R, Chai G, Xu M, Qi G, Kong Y, Zhou G. 2013. Genome-wide identification, classification, and expression analysis of CDPK and its closely related gene families in poplar (Populus trichocarpa). Molecular Biology Reports. 40(3):2645-2662. doi:10.1007/s11033-012-2351-z.

# Appendices

**Appendix 1.** **IDs and characteristics of CPKs used in the phylogenetic analysis (Chapter 3.3.1).**

Sequences highlighted in red are published as CPKs by previous authors, but were not included in this study because they do not align well with the other CPKs. GreenphylDB and Phytozome IDs of the CPK sequences are both listed in this table; the sequence ID used in the phylogenetic tree are highlighted in yellow.

| Species | CPK name | Gene/EST/TC ID/ Gene models | Chromosome No./ Scaffold Locus | Protein length (AA) | No. of EF Hands | Plant database/ source |
|---|---|---|---|---|---|---|
| *Toxoplasma gondii* | **TgCDPK1** | 162.m00001 | - | 582 | 4 | http://eupathdb.org |
| *Toxoplasma gondii* | **TgCDPK3** | 541.m00134 | - | 537 | 3 | http://eupathdb.org |
| *Plasmodium falciparum* | **PfCDPK3** | PFC0420w | - | 562 | 4 | http://eupathdb.org |
| *Cryptosporidium parvum* | **CpCDPK1** | cgd3_920 | - | 538 | 4 | http://eupathdb.org |
| *Cryptosporidium parvum* | **CpCDPK3** | cgd5_820 | - | 523 | 4 | http://eupathdb.org |
| *Chlamydomonas reinhardtii* | 1 | Cre17.g705000 | 17 | 614 | 4 | Phytozome |
| *Chlamydomonas reinhardtii* | 2 | Cre02.g114750 | 2 | | 4 | Phytozome |
| *Chlamydomonas reinhardtii* | 3 | Cre33.g782750 | scaffold_33: 124706 - 136209 | 565 | 4 | Phytozome |
| *Chlamydomonas reinhardtii* | 4 | Cre88.g796750 | scaffold_88: 3199 - 8579 | 477 | 4 | Phytozome |
| *Chlamydomonas reinhardtii* | 5 | Cre07.g328900 | 7 | 485 | 4 | Phytozome |
| *Chlamydomonas reinhardtii* | 6 | Cre01.g009500 | 1 | 764 | 4 | Phytozome |
| *Chlamydomonas reinhardtii* | 7 | Cre02.g074370 | 2 | | 4 | Phytozome |
| *Chlamydomonas reinhardtii* | 8 | Cre13.g571700 | 13 | 509 | 4 | Phytozome |
| *Volvox carteri* | 1 | Vcarteri41333 (Vocar20008615m.g ) | scaffold_1: 13184881 - 13194476 | 456 | 3 | Phytozome |
| *Volvox carteri* | 2 | Vcarteri74309 (Vocar20010397m.g) | scaffold_4: 1811903 - 1815938 | 613 | 4 | Phytozome |
| *Volvox carteri* | 3 | Vcarteri84165 (Vocar20001731m.g) | scaffold_22: 488110 - 490656 | 486 | 4 | Phytozome |
| *Volvox carteri* | 4 | Vcarteri82146 (Vocar20000488m.g) | scaffold_17: 957834 - 963731 | 469 | 1 | Phytozome |
| *Volvox carteri* | 5 | Vcarteri109867 (Vocar20000362m.g) | scaffold_17: 296545 - 302538 | 485 | 4 | Phytozome |
| *Volvox carteri* | 6 | Vcarteri81022 (Vocar20005450m.g ) | scaffold_11: 752838 - 758649 | 476 | 4 | Phytozome |
| *Volvox carteri* | 7 | Vcarteri80507 (Vocar20008501m.g ) | scaffold_1: 2436838 - 2441117 | 393 | 4 | Phytozome |
| *Volvox carteri* | 8 | Vcarteri62695 (Vocar20014333m.g) | scaffold_6: 1155297 - 1162113 | 489 | 4 | Phytozome |
| *Volvox carteri* | 9 | Vcarteri119030 (Vocar20007067m.g) | scaffold_79: 30662 - 37048 | 437 | 1 | Phytozome |
| *Volvox carteri* | 10 | Vcarteri57999 | - | 467 | 1 | Phytozome |
| *Physcomitrella patens* | 1 | Phypa_130600 | - | 653 | 4 | GreenphylDB |
| *Physcomitrella patens* | | ==Pp1s83_8V6|1807352 1== | - | | | Phytozome |
| *Physcomitrella patens* | 2 | Phypa_90621 | - | 590 | 4 | GreenphylDB |
| *Physcomitrella patens* | | ==Pp1s199_57V6|1805 1297== | - | | | Phytozome |
| *Physcomitrella patens* | 3 | Phypa_185437 | - | 589 | 4 | GreenphylDB |
| *Physcomitrella patens* | | ==Pp1s83_172V6|1807 3410== | - | | | Phytozome |
| *Physcomitrella patens* | 4 | Pp1s370_37V6|1804 2034 | - | 549 | 4 | Phytozome |
| *Physcomitrella patens* | | Phypa_226574 | - | | | GreenphylDB |

| Physcomitrella patens | | Pp1s370_37V6\|18042035 | - | | | Phytozome |
|---|---|---|---|---|---|---|
| Physcomitrella patens | 5 | Phypa_111683 | - | 522 | 4 | GreenphylDB |
| Physcomitrella patens | | Pp1s2_191V6\|18052708 | - | | | Phytozome |
| Physcomitrella patens | 6 | Phypa_92858 | - | 513 | 4 | GreenphylDB |
| Physcomitrella patens | | Pp1s232_44V6\|18070869 | - | | | Phytozome |
| Physcomitrella patens | 7 | Phypa_61370 | - | 526 | 4 | GreenphylDB |
| Physcomitrella patens | | Pp1s364_58V6\|18072372 | - | | | Phytozome |
| Physcomitrella patens | 8 | Phypa_98724 | - | 527 | 4 | GreenphylDB |
| Physcomitrella patens | | Pp1s364_61V6\|18072386 | - | | | Phytozome |
| Physcomitrella patens | 9 | Phypa_158539 | - | 525 | 4 | GreenphylDB |
| Physcomitrella patens | | Pp1s2_156V6\|18052950 | - | | | Phytozome |
| Physcomitrella patens | | Pp1s97_71V6\|18062298 | - | | | Phytozome |
| Physcomitrella patens | 10 | Phypa_234941 | - | 524 | 4 | GreenphylDB |
| Physcomitrella patens | | Pp1s97_71V6\|18062299 | - | | | Phytozome |
| Physcomitrella patens | 11 | Phypa_96987 | - | 545 | 4 | GreenphylDB |
| Physcomitrella patens | | Pp1s316_13V6\|18051059 | - | | | Phytozome |
| Physcomitrella patens | 12 | Phypa_214955 | - | 567 | 4 | GreenphylDB |
| Physcomitrella patens | | Pp1s108_25V6\|18070942 | - | | | Phytozome |
| Physcomitrella patens | 13 | Phypa_225603 | - | 534 | 4 | GreenphylDB |
| Physcomitrella patens | | Pp1s325_31V6\|18045877 | - | | | Phytozome |
| Physcomitrella patens | | Pp1s96_216V6\|18068609 | - | | | Phytozome |
| Physcomitrella patens | 14 | Pp1s96_216V6\|18068610 | - | 549 | 4 | Phytozome |
| Physcomitrella patens | | Phypa_186654 | - | | | GreenphylDB |
| Physcomitrella patens | | Pp1s96_216V6\|18068611 | - | | | Phytozome |
| Physcomitrella patens | 15 | Phypa_214963 | - | 578 | 4 | GreenphylDB |
| Physcomitrella patens | | Pp1s108_32V6\|18070898 | - | | | Phytozome |
| Physcomitrella patens | | Phypa_86218 | - | | | GreenphylDB |
| Physcomitrella patens | | Pp1s143_92V6\|18045769 | - | | | Phytozome |
| Physcomitrella patens | 16 | Pp1s143_92V6\|18045770 | - | 628 | 4 | Phytozome |
| Physcomitrella patens | | Pp1s143_92V6\|18045772 | - | | | Phytozome |
| Physcomitrella patens | | Pp1s143_92V6\|18045771 | - | | | Phytozome |
| Physcomitrella patens | 17 | Phypa_140359 | - | 550 | 4 | GreenphylDB |
| Physcomitrella patens | | Pp1s166_57V6\|18052585 | - | | | Phytozome |
| Physcomitrella patens | 18 | Phypa_171758 | - | 494 | 4 | GreenphylDB |
| Physcomitrella patens | | Pp1s309_91V6\|18052079 | - | | | Phytozome |
| Physcomitrella patens | 19 | Phypa_189881 | - | 496 | 4 | GreenphylDB |
| Physcomitrella patens | | Pp1s138_79V6\|18060521 | - | | | Phytozome |
| Physcomitrella patens | 20 | Phypa_168902 | - | 574 | 4 | GreenphylDB |
| Physcomitrella patens | | Phypa_193793 | - | | | GreenphylDB |
| Physcomitrella patens | | Pp1s205_14V6\|18038182 | - | | | Phytozome |
| Physcomitrella patens | 21 | Pp1s205_14V6\|18038181 | - | 575 | 4 | Phytozome |
| Physcomitrella patens | | Pp1s205_14V6\|18038180 | - | | | Phytozome |
| Physcomitrella patens | | Pp1s205_14V6\|18038179 | - | | | Phytozome |
| Physcomitrella patens | | Pp1s205_14V6\|18038178 | - | | | Phytozome |

| Physcomitrella patens | | Pp1s205_14V6\|180338177 | - | | | Phytozome |
|---|---|---|---|---|---|---|
| Physcomitrella patens | | Pp1s205_14V6\|180338176 | - | | | Phytozome |
| Physcomitrella patens | | Pp1s49_200V6\|18041279 | - | | | Phytozome |
| Physcomitrella patens | 22 | Pp1s49_208V6\|18041343 | - | 491 | 4 | Phytozome |
| Physcomitrella patens | | Phypa_124358 | - | | | GreenphylDB |
| Physcomitrella patens | | Phypa_181460 | - | | | GreenphylDB |
| Physcomitrella patens | | Pp1s187_77V6\|18039254 | - | | | Phytozome |
| Physcomitrella patens | | Phypa_220069 | - | | | GreenphylDB |
| Physcomitrella patens | 23 | Pp1s187_88V6\|18039260 | - | 593 | 4 | Phytozome |
| Physcomitrella patens | | Phypa_192815 | - | | | GreenphylDB |
| Physcomitrella patens | | Pp1s187_88V6\|18039259 | - | | | Phytozome |
| Selaginella moellendorfii | 1 | Smoellindorffii\|92726\|15420653 | - | 504 | 4 | Phytozome |
| Selaginella moellendorfii | | Selmo_92726 | - | | | GreenPhylDB |
| Selaginella moellendorfii | 2 | Selmo_114420 | - | 524 | 4 | GreenPhylDB |
| Selaginella moellendorfii | 3 | Selmo_408188 | - | 536 | 4 | GreenPhylDB |
| Selaginella moellendorfii | | Selmo_164119 | - | | | GreenPhylDB |
| Selaginella moellendorfii | 4 | Smoellindorffii\|164119\|15408038 | - | 543 | 4 | Phytozome |
| Selaginella moellendorfii | | Smoellindorffii\|105020\|15411844 | - | | | Phytozome |
| Selaginella moellendorfii | 5 | Selmo_105020 | - | 531 | 4 | GreenPhylDB |
| Selaginella moellendorfii | | Selmo_125607 | - | | | GreenPhylDB |
| Selaginella moellendorfii | | Smoellindorffii\|105846\|15417957 | - | | | Phytozome |
| Selaginella moellendorfii | 6 | Selmo_105846 | - | 539 | 4 | GreenPhylDB |
| Selaginella moellendorfii | | Selmo_178366 | - | | | GreenPhylDB |
| Selaginella moellendorfii | | Smoellindorffii\|165073\|15410490 | - | | | Phytozome |
| Selaginella moellendorfii | 7 | Selmo_142823 | - | 497 | 4 | GreenPhylDB |
| Selaginella moellendorfii | | Selmo_165073 | - | | | GreenPhylDB |
| Selaginella moellendorfii | | Smoellindorffii\|99178\|15414412 | - | | | Phytozome |
| Selaginella moellendorfii | 8 | Selmo_105709 | - | 494 | 4 | GreenPhylDB |
| Selaginella moellendorfii | | Selmo_99178 | - | | | GreenPhylDB |
| Selaginella moellendorfii | | Smoellindorffii\|96034\|15404581 | - | | | Phytozome |
| Selaginella moellendorfii | 9 | Selmo_236322 | - | 496 | 4 | GreenPhylDB |
| Selaginella moellendorfii | | Selmo_96034 | - | | | GreenPhylDB |
| Selaginella moellendorfii | 10 | Selmo_231639 | - | 504 | 4 | GreenPhylDB |
| Selaginella moellendorfii | 11 | Smoellindorffii\|231127\|15414521 | - | 530 | 4 | Phytozome |
| Selaginella moellendorfii | | Selmo_231127 | - | | | GreenPhylDB |
| Selaginella moellendorfii | 12 | Selmo_177720 | - | 577 | 4 | GreenPhylDB |
| Selaginella moellendorfii | 13 | Smoellindorffii\|152133\|15417480 | - | 486 | 4 | Phytozome |
| Selaginella moellendorfii | | Selmo_152133 | - | | | GreenPhylDB |
| Selaginella moellendorfii | | Smoellindorffii\|118877\|15411727 | - | | | Phytozome |
| Selaginella moellendorfii | 14 | Selmo_118877 | - | 493 | 4 | GreenPhylDB |
| Selaginella moellendorfii | | Selmo_184301 | - | | | GreenPhylDB |
| Oryza sativa | OsCPK01 | Os01g43410 | 1 | 713 | 4 | Phytozome |
| Oryza sativa | OsCPK02 | Os01g59360 | 1 | 515 | 4 | Phytozome |
| Oryza sativa | OsCPK03 | Os01g61590 | 1 | 551 | 4 | Phytozome |
| Oryza sativa | OsCPK04 | Os02g03410 | 2 | 520 | 4 | Phytozome |
| Oryza sativa | OsCPK05 | Os02g46090 | 2 | 512 | 3 | Phytozome |
| Oryza sativa | OsCPK06 | Os02g58520 | 2 | 545 | 1 | Phytozome |
| Oryza sativa | OsCPK07 | Os03g03660 | 3 | 570 | 4 | Phytozome |
| Oryza sativa | OsCPK08 | Os03g59390 | 3 | 538 | 4 | Phytozome |
| Oryza sativa | OsCPK09 | Os03g48270 | 3 | 574 | 4 | Phytozome |
| Oryza sativa | OsCPK10 | Os03g57450 | 3 | 599 | 4 | Phytozome |

| | | | | | | |
|---|---|---|---|---|---|---|
| *Oryza sativa* | OsCPK11 | Os03g57510 | 3 | 576 | 4 | Phytozome |
| *Oryza sativa* | OsCPK12 | Os04g47300 | 4 | 533 | 4 | Phytozome |
| *Oryza sativa* | OsCPK13 | Os04g49510 | 4 | 551 | 4 | Phytozome |
| *Oryza sativa* | OsCPK14 | Os05g41270 | 5 | 528 | 4 | Phytozome |
| *Oryza sativa* | OsCPK15 | Os05g50810 | 5 | 542 | 4 | Phytozome |
| *Oryza sativa* | OsCPK16 | Os05g39090 | 5 | 547 | 4 | Phytozome |
| *Oryza sativa* | OsCPK17 | Os07g06740 | 7 | 568 | 4 | Phytozome |
| *Oryza sativa* | OsCPK18 | Os07g22710 | 7 | 513 | 4 | Phytozome |
| *Oryza sativa* | OsCPK19 | Os07g33110 | 7 | 533 | 4 | Phytozome |
| *Oryza sativa* | OsCPK20 | Os07g38120 | 7 | 550 | 4 | Phytozome |
| *Oryza sativa* | OsCPK21 | Os08g42750 | 8 | 565 | 4 | Phytozome |
| *Oryza sativa* | OsCPK22 | Os09g33910 | 9 | 577 | 4 | Phytozome |
| *Oryza sativa* | OsCPK23 | Os10g39420 | 10 | 534 | 4 | Phytozome |
| *Oryza sativa* | OsCPK24 | Os11g07040 | 11 | 513 | 4 | Phytozome |
| *Oryza sativa* | OsCPK25 | Os11g04170 | 11 | 541 | 3 | Phytozome |
| *Oryza sativa* | OsCPK26 | Os12g03970 | 12 | 541 | 4 | Phytozome |
| *Oryza sativa* | OsCPK27 | Os12g30150 | 12 | 612 | 4 | Phytozome |
| *Oryza sativa* | OsCPK28 | Os12g07230 | 12 | 526 | 4 | Phytozome |
| *Oryza sativa* | OsCPK29 | Os12g12860 | 12 | 563 | 4 | Phytozome |
| *Triticum aestivum* | TaCPK1 | TC252299 | - | 717 | 4 | Li et al. (2008) |
| *Triticum aestivum* | TaCPK2 | TC247204 | - | 564 | 4 | Li et al. (2008) |
| *Triticum aestivum* | TaCPK3 | TC270498 | - | 664 | 4 | Li et al. (2008) |
| *Triticum aestivum* | TaCPK4 | TC266250 | - | 764 | 4 | Li et al. (2008) |
| *Triticum aestivum* | TaCPK5 | TC265323 | - | 653 | 4 | Li et al. (2008) |
| *Triticum aestivum* | TaCPK6 | TC246172 | - | 517 | 4 | Li et al. (2008) |
| *Triticum aestivum* | TaCPK7 | TC252008 | - | 682 | 4 | Li et al. (2008) |
| *Triticum aestivum* | TaCPK8 | TC252083 | - | 524 | 4 | Li et al. (2008) |
| *Triticum aestivum* | TaCPK9 | BE498083 | - | 759 | 4 | Li et al. (2008) |
| *Triticum aestivum* | TaCPK10 | TC255319 | - | 564 | 4 | Li et al. (2008) |
| *Triticum aestivum* | TaCPK11 | TC256213 | - | 430 | 4 | Li et al. (2008) |
| *Triticum aestivum* | TaCPK12 | CV770153 | - | 737 | 4 | Li et al. (2008) |
| *Triticum aestivum* | TaCPK13 | BJ258339 | - | 752 | 4 | Li et al. (2008) |
| *Triticum aestivum* | <mark style="background:red">TaCPK14</mark> | TC386410 | - | - | - | Li et al. (2008) |
| *Triticum aestivum* | TaCPK15 | TC248321 | - | 551 | 4 | Li et al. (2008) |
| *Triticum aestivum* | TaCPK16 | TC242486 | - | 672 | - | Li et al. (2008) |
| *Triticum aestivum* | <mark style="background:red">TaCPK17</mark> | CA744920 | - | - | 4 | Li et al. (2008) |
| *Triticum aestivum* | TaCPK18 | CJ614636 | - | 534 | 4 | Li et al. (2008) |
| *Triticum aestivum* | TaCPK19 | CJ626979 | - | 569 | 4 | Li et al. (2008) |
| *Triticum aestivum* | <mark style="background:red">TaCPK20</mark> | BQ802750 | - | - | - | Li et al. (2008) |
| *Sorghum bicolor* | 1 | <mark style="background:yellow">Sb02g009790|1956517</mark> | 2 | 514 | 4 | Phytozome |
| *Sorghum bicolor* | | Sb02g009790.1_SORBI | | | | GreenphylDB |
| *Sorghum bicolor* | 2 | <mark style="background:yellow">Sb04g002220|1965131</mark> | 4 | 523 | 4 | Phytozome |
| *Sorghum bicolor* | | Sb04g002220|1965130 | | | | Phytozome |
| *Sorghum bicolor* | | Sb04g002220.1_SORBI | | | | GreenphylDB |
| *Sorghum bicolor* | | Sb04g002220.2_SORBI | | | | GreenphylDB |
| *Sorghum bicolor* | 3 | <mark style="background:yellow">Sb08g007660|1978011</mark> | 8 | 569 | 4 | Phytozome |
| *Sorghum bicolor* | | Sb08g007660.1_SORBI | | | | GreenphylDB |
| *Sorghum bicolor* | 4 | <mark style="background:yellow">Sb07g025560|1976652</mark> | 7 | 578 | 4 | Phytozome |
| *Sorghum bicolor* | | Sb07g025560.1_SORBI | | | | GreenphylDB |
| *Sorghum bicolor* | 5 | <mark style="background:yellow">Sb03g043700|1964371</mark> | 3 | 557 | 4 | Phytozome |
| *Sorghum bicolor* | | Sb03g043700.1_SORBI | | | | GreenphylDB |
| *Sorghum bicolor* | 6 | <mark style="background:yellow">Sb01g011630|1950736</mark> | 1 | 586 | 4 | Phytozome |
| *Sorghum bicolor* | | Sb01g011630.1_SORBI | | | | GreenphylDB |

| | | | | | | |
|---|---|---|---|---|---|---|
| Sorghum bicolor | 7 | Sb03g038870\|1963795 | 3 | 545 | 4 | Phytozome |
| Sorghum bicolor | | Sb03g038870.1_SORBI | | | | GreenphylDB |
| Sorghum bicolor | 8 | Sb09g022960\|1981196 | 9 | 543 | 4 | Phytozome |
| Sorghum bicolor | | Sb09g022960.1_SORBI | | | | GreenphylDB |
| Sorghum bicolor | 9 | Sb02g036730\|1959081 | 2 | 543 | 4 | Phytozome |
| Sorghum bicolor | | Sb02g036730.1_SORBI | | | | GreenphylDB |
| Sorghum bicolor | 10 | Sb01g004150\|1949788 | 1 | 532 | 4 | Phytozome |
| Sorghum bicolor | | Sb01g004150.1_SORBI | | | | GreenphylDB |
| Sorghum bicolor | 11 | Sb06g025220\|1973468 | 6 | 533 | 4 | Phytozome |
| Sorghum bicolor | | Sb06g025220.1_SORBI | | | | GreenphylDB |
| Sorghum bicolor | 12 | Sb02g034640.1_SORBI | 2 | 531 | 4 | GreenphylDB |
| Sorghum bicolor | | Sb02g034640\|1958814 | | | | Phytozome |
| Sorghum bicolor | 13 | Sb03g028340\|1962510 | 3 | 525 | 4 | Phytozome |
| Sorghum bicolor | | Sb03g028340.1_SORBI | | | | GreenphylDB |
| Sorghum bicolor | 14 | Sb09g029950\|1982027 | 9 | 541 | 4 | Phytozome |
| Sorghum bicolor | | Sb09g029950.1_SORBI | | | | GreenphylDB |
| Sorghum bicolor | 15 | Sb09g024100\|1981329 | 9 | 527 | 4 | Phytozome |
| Sorghum bicolor | | Sb09g024100.1_SORBI | | | | GreenphylDB |
| Sorghum bicolor | 16 | Sb03g037570\|1963652 | 3 | 462 | 4 | Phytozome |
| Sorghum bicolor | | Sb03g037570.1_SORBI | | | | GreenphylDB |
| Sorghum bicolor | 17 | Sb05g002110\|1969059 | 5 | 538 | 4 | Phytozome |
| Sorghum bicolor | | Sb05g002110.1_SORBI | | | | GreenphylDB |
| Sorghum bicolor | 18 | Sb08g001380\|1977263 | 8 | 574 | 4 | Phytozome |
| Sorghum bicolor | | Sb08g001380.1_SORBI | | | | GreenphylDB |
| Sorghum bicolor | 19 | Sb01g048570\|1954982 | 1 | 544 | 4 | Phytozome |
| Sorghum bicolor | | Sb01g048570.1_SORBI | | | | GreenphylDB |
| Sorghum bicolor | 20 | Sb01g030450\|1952760 | 1 | 538 | 3 | Phytozome |
| Sorghum bicolor | | Sb01g030450.1_SORBI | | | | GreenphylDB |
| Sorghum bicolor | 21 | Sb04g038450\|1968759 | 4 | 580 | 4 | Phytozome |
| Sorghum bicolor | | Sb04g038450.1_SORBI | | | | GreenphylDB |
| Sorghum bicolor | 22 | Sb04g031570\|1967955 | 4 | 490 | 4 | Phytozome |
| Sorghum bicolor | | Sb04g031570.1_SORBI | | | | GreenphylDB |
| Sorghum bicolor | 23 | Sb06g026530\|1973627 | 6 | 555 | 4 | Phytozome |
| Sorghum bicolor | | Sb06g026530.1_SORBI | | | | GreenphylDB |
| Sorghum bicolor | 24 | Sb08g004510\|1977665 | 8 | 515 | 4 | Phytozome |
| Sorghum bicolor | | Sb08g004510.1_SORBI | | | | GreenphylDB |
| Sorghum bicolor | 25 | Sb05g004610\|1969378 | 5 | 515 | 4 | Phytozome |
| Sorghum bicolor | | Sb05g004610.1_SORBI | | | | GreenphylDB |

| Species | ID | Gene | Col | Col | Col | Source |
|---|---|---|---|---|---|---|
| *Sorghum bicolor* | 26 | Sb01g005750\|1949992 | 1 | 585 | 4 | Phytozome |
| *Sorghum bicolor* | | Sb01g005750.1_SORBI | | | | GreenphylDB |
| *Sorghum bicolor* | 27 | Sb02g003500\|1955676 | 2 | 581 | 4 | Phytozome |
| *Sorghum bicolor* | | Sb02g003500.1_SORBI | | | | GreenphylDB |
| *Sorghum bicolor* | 28 | Sb01g005780\|1949995 | 1 | 617 | 4 | Phytozome |
| *Sorghum bicolor* | | Sb01g005780.1_SORBI | | | | GreenphylDB |
| *Sorghum bicolor* | 29 | Sb08g014910\|1978240 | 8 | 645 | 4 | Phytozome |
| *Sorghum bicolor* | | Sb08g014910.1_SORBI | | | | GreenphylDB |
| *Zea mays* | 1 | GRMZM2G365035_P01 | 2 | 512 | 4 | GreenPhylDB |
| *Zea mays* | | GRMZM2G365035 | | | | Phytozome |
| *Zea mays* | 2 | GRMZM2G157068 | 5 | 520 | 4 | Phytozome |
| *Zea mays* | 3 | GRMZM2G053868_P01 | 4 | 522 | 4 | GreenPhylDB |
| *Zea mays* | | GRMZM2G053868 | | | | Phytozome |
| *Zea mays* | 4 | GRMZM2G097533_P01 | 3 | 438 | 1 | GreenPhylDB |
| *Zea mays* | | GRMZM2G097533 | | | | Phytozome |
| *Zea mays* | 5 | GRMZM2G332660_P02 | 4 | 568 | 4 | GreenPhylDB |
| *Zea mays* | | GRMZM2G332660_P01 | | | | GreenPhylDB |
| *Zea mays* | | GRMZM2G332660 | | | | Phytozome |
| *Zea mays* | 6 | GRMZM2G158721 | 2 | 656 | 4 | Phytozome |
| *Zea mays* | 8 | GRMZM2G080871_P02 | 7 | 511 | 3 | GreenPhylDB |
| *Zea mays* | | GRMZM2G080871 | | | | Phytozome |
| *Zea mays* | 9 | GRMZM2G030673_P01 | 8 | 541 | 4 | GreenPhylDB |
| *Zea mays* | | GRMZM2G030673 | | | | Phytozome |
| *Zea mays* | 10 | GRMZM2G088361_P01 | 6 | 540 | 4 | GreenPhylDB |
| *Zea mays* | | GRMZM2G088361 | | | | Phytozome |
| *Zea mays* | 11 | GRMZM2G311220_P03 | 8 | 536 | 4 | GreenPhylDB |
| *Zea mays* | | GRMZM2G311220 | | | | Phytozome |
| *Zea mays* | | GRMZM2G311220_P01 | | | | GreenPhylDB |
| *Zea mays* | | GRMZM2G311220 | | | | Phytozome |
| *Zea mays* | 12 | GRMZM2G104125 | 1 | 536 | 4 | Phytozome |
| *Zea mays* | 13 | AC210013.4_FGP014 | 5 | 538 | 4 | Phytozome |
| *Zea mays* | | AC210013.4_FG014 | | | | Phytozome |
| *Zea mays* | ZmCPK7-2 | AC233871.1_FGP003 | 6 | 539 | 4 | Phytozome |
| *Zea mays* | | AC233871.1_FG003 | | | | Phytozome |
| *Zea mays* | | GRMZM2G028086 | 7 | | | Phytozome |
| *Zea mays* | ZmCPK7-1 | GRMZM2G099425_P01 | 2 | 539 | 4 | GreenPhylDB |
| *Zea mays* | | GRMZM2G099425 | | | | Phytozome |
| *Zea mays* | 14 | GRMZM2G112057_P01 | 10 | 539 | 4 | GreenPhylDB |
| *Zea mays* | | GRMZM2G112057 | | | | Phytozome |
| *Zea mays* | ZmCPK2 | GRMZM2G154489_P01 | 7 | 531 | 4 | GreenPhylDB |
| *Zea mays* | | GRMZM2G154489 | | | | Phytozome |
| *Zea mays* | ZmCPK9 | GRMZM2G168706_P05 | 2 | 531 | 4 | GreenPhylDB |
| *Zea mays* | | GRMZM2G168706 | | | | Phytozome |
| *Zea mays* | 15 | GRMZM2G472311 | 4 | 581 | 4 | Phytozome |
| *Zea mays* | 16 | GRMZM2G365815_P01 | 2 | 552 | 4 | GreenPhylDB |
| *Zea mays* | | GRMZM2G365815 | | | | Phytozome |

| | | | | | | |
|---|---|---|---|---|---|---|
| *Zea mays* | 17 | GRMZM2G340224 | 8 | 614 | 4 | Phytozome |
| *Zea mays* | 18 | GRMZM2G167276_P01 | 3 | 510 | 4 | GreenPhylDB |
| *Zea mays* | | GRMZM2G167276 | | | | Phytozome |
| *Zea mays* | 19 | GRMZM2G058305_P03 | 8 | 539 | 4 | GreenPhylDB |
| *Zea mays* | | GRMZM2G058305_P01 | | | | GreenPhylDB |
| *Zea mays* | | GRMZM2G058305 | | | | Phytozome |
| *Zea mays* | 20 | GRMZM2G025387_P01 | 8 | 530 | 4 | GreenPhylDB |
| *Zea mays* | | GRMZM2G025387 | | | | Phytozome |
| *Zea mays* | 21 | GRMZM5G856738 | 3 | 524 | 4 | Phytozome |
| *Zea mays* | 22 | GRMZM2G040743_P01 | 1 | 540 | 4 | GreenPhylDB |
| *Zea mays* | | GRMZM2G040743 | | | | Phytozome |
| *Zea mays* | 23 | GRMZM2G032852 | | | | Phytozome |
| *Zea mays* | | GRMZM2G032852_P02 | 1 | 544 | 4 | GreenPhylDB |
| *Zea mays* | | GRMZM2G032852 | | | | Phytozome |
| *Zea mays* | 24 | GRMZM2G347047_P01 | 4 | 488 | 4 | GreenPhylDB |
| *Zea mays* | | GRMZM2G347047 | | | | Phytozome |
| *Zea mays* | 25 | GRMZM2G081310_P01 | 4 | 562 | 4 | GreenPhylDB |
| *Zea mays* | | GRMZM2G081310 | | | | Phytozome |
| *Zea mays* | ZmCPK7 | GRMZM2G321239 | 10 | 557 | 4 | Phytozome |
| *Zea mays* | ZmCPK1 | GRMZM2G314396_P01 | 2 | 547 | 4 | GreenPhylDB |
| *Zea mays* | | GRMZM2G314396 | | | | Phytozome |
| *Zea mays* | 26 | GRMZM2G012326 | 2 | 605 | 4 | Phytozome |
| *Zea mays* | 27 | GRMZM2G121228 | 1 | 581 | 4 | Phytozome |
| *Zea mays* | 28 | GRMZM2G027351 | 5 | 585 | 4 | Phytozome |
| *Zea mays* | 29 | GRMZM2G353957_P01 | 3 | 646 | 4 | GreenPhylDB |
| *Zea mays* | | GRMZM2G353957 | | | | Phytozome |
| *Zea mays* | 30 | GRMZM2G320506 | 5 | 621 | 4 | Phytozome |
| *Zea mays* | ZmCPK10 | GRMZM2G028926_P01 | 1 | 608 | 4 | GreenPhylDB |
| *Zea mays* | | GRMZM2G028926 | | | | Phytozome |
| *Zea mays* | 31 | GRMZM2G347226_P01 | 10 | 609 | 4 | GreenPhylDB |
| *Zea mays* | | GRMZM2G347226 | | | | Phytozome |
| *Zea mays* | 32 | GRMZM2G463464_P01 | 3 | 609 | 4 | GreenPhylDB |
| *Zea mays* | | GRMZM2G4463464 | | | | Phytozome |
| *Zea mays* | | GRMZM2G463464 | | | | Phytozome |
| *Zea mays* | 33 | GRMZM2G035843_P01 | 4 | 609 | 4 | GreenPhylDB |
| *Zea mays* | | GRMZM2G035843 | | | | Phytozome |
| *Zea mays* | ZmCPK11 | GRMZM2G047486 | 2 | 609 | 4 | Phytozome |
| *Vitis vinifera* | 2 | GSVIVP00036780001_VITVI | 4 | 558 | 4 | GreenPhylDB |
| *Vitis vinifera* | | GSVIVG01018778001\|chr4 | | | | Phytozome |
| *Vitis vinifera* | 3 | GSVIVP00000071001_VITVI | 17 | 552 | 4 | GreenPhylDB |
| *Vitis vinifera* | | GSVIVG01008077001\|chr17 | | | | Phytozome |
| *Vitis vinifera* | 4 | GSVIVP00038883001_VITVI | 8 | 527 | 4 | GreenPhylDB |
| *Vitis vinifera* | | GSVIVG01011167001\|chr8 | | | | Phytozome |
| *Vitis vinifera* | 5 | GSVIVP00001926001_VITVI | 5 | 554 | 4 | GreenPhylDB |
| *Vitis vinifera* | | GSVIVG01010743001\|chr5 | | | | Phytozome |
| *Vitis vinifera* | 6 | GSVIVP00024598001_VITVI | 6 | 519 | 4 | GreenPhylDB |
| *Vitis vinifera* | | GSVIVG01025249001\|chr6 | | | | Phytozome |

| Species | Name | Chr | Length | Exons | Source |
|---|---|---|---|---|---|
| *Vitis vinifera* | 7 | GSVIVP00025568001_VITVI | 8 | 526 | 4 | GreenPhylDB |
| *Vitis vinifera* | | GSVIVG01033306001\|chr8 | | | | Phytozome |
| *Vitis vinifera* | 8 | GSVIVP00002511001_VITVI | 2 | 528 | 4 | GreenPhylDB |
| *Vitis vinifera* | | GSVIVG01019446001\|chr2 | | | | Phytozome |
| *Vitis vinifera* | 9 | GSVIVP00022286001_VITVI | 6 | 534 | 4 | GreenPhylDB |
| *Vitis vinifera* | | GSVIVG01037295001\|chr6 | | | | Phytozome |
| *Vitis vinifera* | 10 | GSVIVP00015499001_VITVI | 18 | 523 | 4 | GreenPhylDB |
| *Vitis vinifera* | | GSVIVG01008749001\|chr18 | | | | Phytozome |
| *Vitis vinifera* | 11 | GSVIVG01037652001\|chr19 | 19 | 537 | 4 | Phytozome |
| *Vitis vinifera* | | GSVIVG01037652001\|chr19 | | | | ORF checked |
| *Vitis vinifera* | 12 | GSVIVP00010752001_VITVI | 10 | 545 | 4 | GreenPhylDB |
| *Vitis vinifera* | | GSVIVG01012730001\|chr10 | | | | Phytozome |
| *Vitis vinifera* | 13 | GSVIVP00036285001_VITVI | 3 | 561 | 4 | GreenPhylDB |
| *Vitis vinifera* | | GSVIVG01023866001\|chr3 | | | | Phytozome |
| *Vitis vinifera* | VvCPK1/ACPK1 | GSVIVP00003655001_VITVI | 7 | 497 | 4 | GreenPhylDB |
| *Vitis vinifera* | | GSVIVG01000238001\|chr7 | | | | Phytozome |
| *Vitis vinifera* | 15 | GSVIVP00031812001_VITVI | 18 | 489 | 4 | GreenPhylDB |
| *Vitis vinifera* | | GSVIVG01034489001\|chr18 | | | | Phytozome |
| *Vitis vinifera* | 16 | GSVIVP00027136001_VITVI | 8 | 568 | 4 | GreenPhylDB |
| *Vitis vinifera* | | GSVIVG01022524001\|chr8 | | | | Phytozome |
| *Vitis vinifera* | 17 | GSVIVP00003538001_VITVI | 8 | 580 | 4 | GreenPhylDB |
| *Vitis vinifera* | | GSVIVG01022606001\|chr8 | | | | Phytozome |
| *Vitis vinifera* | 18 | GSVIVP00002665001_VITVI | 13 | 569 | 4 | GreenPhylDB |
| *Vitis vinifera* | | GSVIVG01001931001\|chr13 | | | | Phytozome |
| *Glycine max* | 1 | Glyma01g37100.1_SOYBN | 01 | 550 | 4 | GreenPhylDB |
| *Glycine max* | | Glyma01g37100\|16245443 | | | | Phytozome |
| *Glycine max* | 2 | Glyma11g08180.1_SOYBN | 11 | 540 | 4 | GreenPhylDB |
| *Glycine max* | | Glyma11g08180\|16283098 | | | | Phytozome |
| *Glycine max* | 3 | Glyma16g23870\|16302694 | 16 | 554 | 4 | Phytozome |
| *Glycine max* | | Glyma16g23870\|16302693 | | | | Phytozome |
| *Glycine max* | | Glyma16g23870.1_SOYBN | | | | GreenPhylDB |
| *Glycine max* | 4 | Glyma02g05440\|16247072 | 02 | 530 | 4 | Phytozome |
| *Glycine max* | | Glyma02g05440.1_SOYBN | | | | GreenPhylDB |
| *Glycine max* | 5 | Glyma12g05730\|16286536 | 12 | 575 | 4 | Phytozome |
| *Glycine max* | | Glyma12g05730.1_SOYBN | | | | GreenPhylDB |
| *Glycine max* | 6 | Glyma11g13740\|16283788 | 11 | 530 | 4 | Phytozome |
| *Glycine max* | | Glyma11g13740.1_SOYBN | | | | GreenPhylDB |
| *Glycine max* | 7 | Glyma03g29450\|16252722 | 3 | 534 | 4 | Phytozome |

| | | | | | | |
|---|---|---|---|---|---|---|
| *Glycine max* | | Glyma03g29450.1_SOYBN | | | | GreenPhylDB |
| *Glycine max* | 8 | Glyma19g32260\|16313480 | 19 | 535 | 4 | Phytozome |
| *Glycine max* | | Glyma19g32260.1_SOYBN | | | | GreenPhylDB |
| *Glycine max* | 9 | Glyma02g31490\|16248928 | 2 | 525 | 4 | Phytozome |
| *Glycine max* | | Glyma02g31490.1_SOYBN | | | | GreenPhylDB |
| *Glycine max* | 10 | Glyma10g17560\|16279611 | 10 | 568 | 4 | Phytozome |
| *Glycine max* | | Glyma10g17560.1_SOYBN | | | | GreenPhylDB |
| *Glycine max* | 11 | Glyma18g43160\|16310302 | 18 | 531 | 4 | Phytozome |
| *Glycine max* | 12 | Glyma07g18310\|16267680 | 7 | 533 | 4 | Phytozome |
| *Glycine max* | | Glyma07g18310.1_SOYBN | | | | GreenPhylDB |
| *Glycine max* | 13 | Glyma05g01470.1_SOYBN | 5 | 539 | 4 | GreenPhylDB |
| *Glycine max* | | Glyma05g01470\|16257958 | | | | Phytozome |
| *Glycine max* | 14 | Glyma17g10410.1_SOYBN | 17 | 541 | 4 | GreenPhylDB |
| *Glycine max* | | Glyma17g10410\|16305086 | | | | Phytozome |
| *Glycine max* | 15 | Glyma04g34440\|16256738 | 4 | 534 | 4 | Phytozome |
| *Glycine max* | | Glyma04g34440.1_SOYBN | | | | GreenPhylDB |
| *Glycine max* | 16 | Glyma06g20170\|16263586 | 6 | 551 | 4 | Phytozome |
| *Glycine max* | | Glyma06g20170.1_SOYBN | | | | GreenPhylDB |
| *Glycine max* | 17 | Glyma11g02260\|16282406 | 11 | 505 | 4 | Phytozome |
| *Glycine max* | | Glyma11g02260.1_SOYBN | | | | GreenPhylDB |
| *Glycine max* | 18 | Glyma05g37260\|16260996 | 5 | 518 | 4 | Phytozome |
| *Glycine max* | | Glyma05g37260.1_SOYBN | | | | GreenPhylDB |
| *Glycine max* | 19 | Glyma08g02300\|16269679 | 8 | 520 | 4 | Phytozome |
| *Glycine max* | | Glyma08g02300.1_SOYBN | | | | GreenPhylDB |
| *Glycine max* | 20 | Glyma14g04010.1_SOYBN | 14 | 529 | 4 | GreenPhylDB |
| *Glycine max* | 21 | Glyma02g44720\|16250188 | 2 | 527 | 4 | Phytozome |
| *Glycine max* | | Glyma02g44720.1_SOYBN | | | | GreenPhylDB |
| *Glycine max* | 22 | Glyma20g08140\|16315696 | 20 | 531 | 4 | Phytozome |
| *Glycine max* | | Glyma20g08140.1_SOYBN | | | | GreenPhylDB |
| *Glycine max* | 23 | Glyma07g36000\|16268877 | 7 | 510 | 4 | Phytozome |
| *Glycine max* | | Glyma07g36000.1_SOYBN | | | | GreenPhylDB |
| *Glycine max* | 24 | Glyma17g38050\|16307407 | 17 | 576 | 4 | Phytozome |
| *Glycine max* | | Glyma17g38050.1_SOYBN | | | | GreenPhylDB |
| *Glycine max* | 25 | Glyma17g38040\|16307406 | 17 | 536 | 4 | Phytozome |
| *Glycine max* | 26 | Glyma14g40090\|16297015 | 14 | 526 | 4 | Phytozome |
| *Glycine max* | | Glyma14g40090.1_SOYBN | | | | GreenPhylDB |
| *Glycine max* | 27 | Glyma07g39010.1_SOYBN | 7 | 529 | 4 | GreenPhylDB |
| *Glycine max* | | Glyma07g39010\|16269222 | | | | Phytozome |

| | | | | | | |
|---|---|---|---|---|---|---|
| *Glycine max* | 28 | Glyma17g01730\|163 04071 | 17 | 538 | 4 | Phytozome |
| *Glycine max* | | Glyma17g01730.1_S OYBN | | | | GreenPhylDB |
| *Glycine max* | 29 | Glyma14g02680\|162 94118 | 14 | 519 | 4 | Phytozome |
| *Glycine max* | | Glyma14g02680.1_S OYBN | | | | GreenPhylDB |
| *Glycine max* | 30 | Glyma02g46070\|162 50354 | 2 | 528 | 4 | Phytozome |
| *Glycine max* | | Glyma02g46070.1_S OYBN | | | | GreenPhylDB |
| *Glycine max* | 31 | Glyma18g11030\|163 08539 | 18 | 551 | 4 | Phytozome |
| *Glycine max* | | Glyma18g11030.1_S OYBN | | | | GreenPhylDB |
| *Glycine max* | 32 | Glyma08g42850\|162 73633 | 8 | 551 | 4 | Phytozome |
| *Glycine max* | | Glyma08g42850.1_S OYBN | | | | GreenPhylDB |
| *Glycine max* | 33 | Glyma02g48160\|162 50593 | 2 | 549 | 4 | Phytozome |
| *Glycine max* | | Glyma02g48160.1_S OYBN | | | | GreenPhylDB |
| *Glycine max* | 34 | Glyma14g00320\|162 93851 | 14 | 558 | 4 | Phytozome |
| *Glycine max* | | Glyma14g00320.1_S OYBN | | | | GreenPhylDB |
| *Glycine max* | 35 | Glyma06g16920\|162 63217 | 6 | 497 | 4 | Phytozome |
| *Glycine max* | 36 | Glyma04g38150\|162 57133 | 4 | 496 | 4 | Phytozome |
| *Glycine max* | GmCPKa | Glyma08g00840\|162 69497 | 8 | 508 | 4 | Phytozome |
| *Glycine max* | | Glyma08g00840.1_S OYBN | | | | GreenPhylDB |
| *Glycine max* | 37 | Glyma05g33240\|162 60515 | 5 | 507 | 4 | Phytozome |
| *Glycine max* | | Glyma05g33240.1_S OYBN | | | | GreenPhylDB |
| *Glycine max* | GmCPKb | Glyma20g31510\|163 17503 | | 483 | 4 | Phytozome |
| *Glycine max* | | Glyma20g31510.1_S OYBN | | | | GreenPhylDB |
| *Glycine max* | 38 | Glyma10g36090\|162 81140 | 10 | 482 | 4 | Phytozome |
| *Glycine max* | | Glyma10g36090.1_S OYBN | | | | GreenPhylDB |
| *Glycine max* | 39 | Glyma10g36100\|162 81141 | 10 | 492 | 4 | Phytozome |
| *Glycine max* | | Glyma10g36100.1_S OYBN | | | | GreenPhylDB |
| *Glycine max* | 40 | Glyma10g11020\|162 79220 | 10 | 585 | 4 | Phytozome |
| *Glycine max* | | Glyma10g11020.1_S OYBN | | | | GreenPhylDB |
| *Glycine max* | 41 | Glyma10g23620\|162 79817 | 10 | 581 | 4 | Phytozome |
| *Glycine max* | | Glyma10g23620.1_S OYBN | | | | GreenPhylDB |
| *Glycine max* | 42 | Glyma20g17020.1_S OYBN | 20 | 579 | 4 | GreenPhylDB |
| *Glycine max* | | Glyma20g17020\|163 16076 | | | | Phytozome |
| *Glycine max* | | Glyma20g17020\|163 16077 | | | | Phytozome |
| *Glycine max* | 43 | Glyma02g34890\|162 49087 | 2 | 531 | 3 | Phytozome |
| *Glycine max* | | Glyma02g34890.1_S OYBN | | | | GreenPhylDB |
| *Glycine max* | 44 | Glyma19g38890\|163 14250 | 19 | 559 | 3 | Phytozome |
| *Glycine max* | | Glyma19g38890.1_S OYBN | | | | GreenPhylDB |
| *Glycine max* | 45 | Glyma03g36240\|162 53515 | 3 | 479 | 4 | Phytozome |

| | | | | | |
|---|---|---|---|---|---|
| *Glycine max* | | Glyma03g36240.1_S OYBN | | | GreenPhylDB |
| *P. trichocarpa* | 1 | Popal_817213 | 7 | 558 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0007s09580 | | | | Phytozome |
| *P. trichocarpa* | 2 | Popal_796704 | 5 | 557 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0005s11560 | | | | Phytozome |
| *P. trichocarpa* | 3 | Popal_802588 | 7 | 535 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0007s02120 | | | | Phytozome |
| *P. trichocarpa* | 4 | Popal_831925 | 6 | 529 | 4 | Phytozome |
| *P. trichocarpa* | | POPTR_0006s10230 | | | | GreenPhylDB |
| *P. trichocarpa* | 5 | Popal_777845 | 16 | 529 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0016s12460 | | | | Phytozome |
| *P. trichocarpa* | 6 | Popal_570005 | 12 | 556 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0012s07360 | | | | Phytozome |
| *P. trichocarpa* | 7 | Popal_775987 | 15 | 563 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0015s07740 | | | | Phytozome |
| *P. trichocarpa* | 8 | Popal_831839 | 6 | 529 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0006s05140 | | | | Phytozome |
| *P. trichocarpa* | 9 | Popal_835420 | 16 | 533 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0016s05490 | | | | Phytozome |
| *P. trichocarpa* | 10 | Popal_722938 | 9 | 534 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0009s05740 | | | | Phytozome |
| *P. trichocarpa* | 11 | Popal_752237 | 1 | 535 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0001s26430 | | | | Phytozome |
| *P. trichocarpa* | 12 | Popal_767863 | 9 | 526 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0009s07330 | | | | Phytozome |
| *P. trichocarpa* | 13 | Popal_797791 | 1 | 506 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0001s28150 | | | | Phytozome |
| *P. trichocarpa* | 14 | Popal_413635 | 3 | 524 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0003s13380 | | | | Phytozome |
| *P. trichocarpa* | 15 | Popal_829455 | 1 | 516 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0001s10070 | | | | Phytozome |
| *P. trichocarpa* | 16 | Popal_206507 | 5 | 514 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0005s26640 | | | | Phytozome |
| *P. trichocarpa* | 17 | Popal_410888 | 2 | 543 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0002s01850 | | | | Phytozome |
| *P. trichocarpa* | 18 | Popal_196596 | 4 | 533 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0004s01530 | | | | Phytozome |
| *P. trichocarpa* | 19 | Popal_233269 | 21 | 533 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0021s00750 | | | | Phytozome |
| *P. trichocarpa* | 20 | Popal_679968 | 4 | 561 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0004s21710 | | | | Phytozome |
| *P. trichocarpa* | 21 | Popal_803966 | 9 | 566 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0009s16970 | | | | Phytozome |
| *P. trichocarpa* | 22 | Popal_826066 | 19 | 504 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0019s00630 | | | | Phytozome |
| *P. trichocarpa* | 23 | Popal_249682 | 19 | 520 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0019s11290 | | | | Phytozome |
| *P. trichocarpa* | 24 | Popal_287767 | 13 | 522 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0013s11690 | | | | Phytozome |
| *P. trichocarpa* | 25 | Popal_561172 | 6 | 613 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0006s21490 | | | | Phytozome |
| *P. trichocarpa* | 26 | Popal_256178 | 16 | 613 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0016s06700 | | | | Phytozome |
| *P. trichocarpa* | 27 | Popal_417449 | 6 | 599 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0006s21390 | | | | Phytozome |
| *P. trichocarpa* | 28 | Popal_256143 | 16 | 599 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0016s06590 | | | | Phytozome |
| *P. trichocarpa* | 29 | Popal_822684 | 10 | 580 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0010s25090 | | | | Phytozome |
| *P. trichocarpa* | 30 | Popal_720354 | 8 | 580 | 4 | GreenPhylDB |
| *P. trichocarpa* | | POPTR_0008s01530 | | | | Phytozome |

| | | | | | | |
|---|---|---|---|---|---|---|
| *A. thaliana* | AtCPK1 | At5g04870 | 5 | 610 | 4 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK2 | At3g10660 | 3 | 646 | 4 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK3 | At4g23650 | 4 | 529 | 4 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK4 | At4g09570 | 4 | 501 | 4 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK5 | At4g35310 | 4 | 556 | 4 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK6 | At2g17290 | 2 | 544 | 4 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK7 | At5g12480 | 5 | 535 | 3 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK8 | At5g19450 | 5 | 533 | 3 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK9 | At3g20410 | 3 | 541 | 4 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK10 | At1g18890 | 1 | 545 | 3 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK11 | At1g35670 | 1 | 495 | 4 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK12 | At5g23580 | 5 | 490 | 4 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK13 | At3g51850 | 3 | 528 | 2 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK14 | At2g41860 | 2 | 530 | 3 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK15 | At4g21940 | 4 | 562 | 4 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK16 | At2g17890 | 2 | 571 | 4 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK17 | At5g12180 | 5 | 528 | 4 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK18 | At4g36070 | 4 | 562 | 4 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK19 | At1g61950 | 1 | 551 | 3 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK20 | At2g38910 | 2 | 583 | 4 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK21 | At4g04720 | 4 | 531 | 4 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK22 | At4g04710 | 4 | 575 | 5 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK23 | At4g04740 | 4 | 533 | 3 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK24 | At2g31500 | 2 | 582 | 4 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK25 | At2g35890 | 2 | 520 | 1 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK26 | At4g38230 | 4 | 514 | 4 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK27 | At4g04700 | 4 | 485 | 0 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK28 | At5g66210 | 5 | 523 | 4 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK29 | At1g76040 | 1 | 561 | 3 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK30(1a) | At1g74740 | 1 | 541 | 4 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK31 | At4g04695 | 4 | 484 | 4 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK32 | At3g57530 | 3 | 538 | 3 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK33 | At1g50700 | 1 | 521 | 4 | Phytozome, TAIR, Cheng et al. (2002) |
| *A. thaliana* | AtCPK34 | At5g19360 | 5 | 524 | 4 | Phytozome, TAIR, Cheng et al. (2002) |
| *Carica papaya* | 1 | TU.supercontig_17.194\|16410693 | - | 547 | 4 | Phytozome |
| *Carica papaya* | | supercontig_17.175_CARPA | - | | | GreenPhylDB |
| *Carica papaya* | 2 | TU.supercontig_84.46\|16427328 | - | 531 | 4 | Phytozome |
| *Carica papaya* | | supercontig_84.45_CARPA | - | | | GreenPhylDB |

298

| | | | | | | |
|---|---|---|---|---|---|---|
| *Carica papaya* | 3 | TU.supercontig_92.3 1\|16428600 | - | 533 | 4 | Phytozome |
| *Carica papaya* | | supercontig_92.31_C ARPA | - | | | GreenPhylDB |
| *Carica papaya* | 4 | TU.contig_33100\|164 30206 | - | 585 | 4 | Phytozome |
| *Carica papaya* | 5 | TU.supercontig_6.25 4\|16423436 | - | 508 | 4 | Phytozome |
| *Carica papaya* | | supercontig_6.249_C ARPA | - | | | GreenPhylDB |
| *Carica papaya* | 6 | TU.supercontig_122. 15\|16407061 | - | 528 | 4 | Phytozome |
| *Carica papaya* | | supercontig_122.13_ CARPA | - | | | GreenPhylDB |
| *Carica papaya* | 7 | TU.supercontig_157. 56\|16409845 | - | 528 | 4 | Phytozome |
| *Carica papaya* | | supercontig_157.55_ CARPA | - | | | GreenPhylDB |
| *Carica papaya* | 8 | TU.supercontig_26.2 69\|16415215 | - | 517 | 4 | Phytozome |
| *Carica papaya* | | supercontig_26.270_ CARPA | - | | | GreenPhylDB |
| *Carica papaya* | 9 | TU.supercontig_152. 12\|16409594 | - | 549 | 4 | Phytozome |
| *Carica papaya* | | supercontig_152.12_ CARPA | - | | | GreenPhylDB |
| *Carica papaya* | 10 | TU.supercontig_12.3 05\|16406865 | - | 526 | 4 | Phytozome |
| *Carica papaya* | | supercontig_12.308_ CARPA | - | | | GreenPhylDB |
| *Carica papaya* | 11 | TU.supercontig_60.3\| 16423718 | - | 566 | 4 | Phytozome |
| *Carica papaya* | | supercontig_60.4_CA RPA | - | | | GreenPhylDB |
| *Carica papaya* | 12 | TU.supercontig_222. 14\|16413982 | - | 486 | 1 | Phytozome |
| *Carica papaya* | | supercontig_222.16_ CARPA | - | | | GreenPhylDB |
| *Carica papaya* | 13 | TU.supercontig_222. 24\|16413993 | - | 623 | 4 | Phytozome |
| *Carica papaya* | | supercontig_222.27_ CARPA | - | | | GreenPhylDB |
| *Carica papaya* | 14 | TU.supercontig_33.1 22\|16417567 | - | 554 | 4 | Phytozome |
| *Carica papaya* | | supercontig_33.109_ CARPA | - | | | GreenPhylDB |

**Appendix 2.    Multiple sequence alignments of trimmed CPK sequences to remove extremely variable regions (Chapter 3.3.1)**

Please see .fasta file in the attached CD.

**Appendix 3.** **Detailed topology of CPK evolutionary group I, ML tree, using trimmed sequences (Chapter 3.3.1).**

Branch colours match the species colour in Figure 3.2. On the far right, the evolutionary groupings are indicated in Roman numerals. Monocot-dicot clusters are indicated by the green boxes.

**Appendix 4.** **Detailed topology of CPK evolutionary group II ML tree, trimmed sequences (Chapter 3.3.1)**
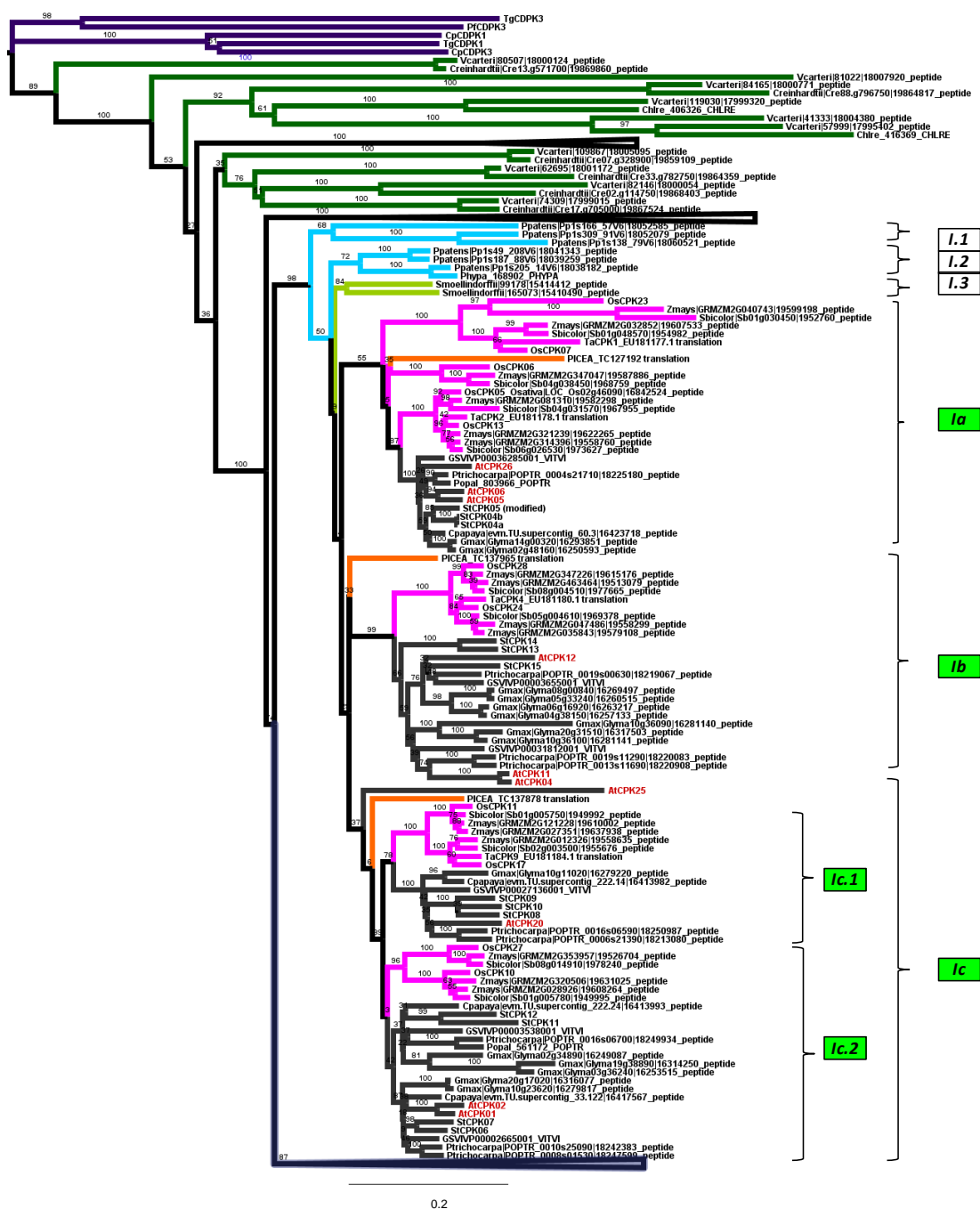
Branch colours match the species colour in Figure 3.2. On the far right, the evolutionary groupings are indicated in Roman numerals. Monocot-dicot clusters are indicated by the green boxes.

**Appendix 5.** **Detailed topology of CPK evolutionary group III ML tree, trimmed sequences (Chapter 3.3.1)**

Branch colours match the species colour in Figure 3.2. On the far right, the evolutionary groupings are indicated in Roman numerals. Monocot-dicot clusters are indicated by the green boxes.

**Appendix 6.** **Detailed topology of CPK evolutionary group IV ML tree, trimmed sequences (Chapter 3.3.1)**

Branch colours match the species colour in Figure 3.2. On the far right, the evolutionary groupings are indicated in Roman numerals. Monocot-dicot clusters are indicated by the green boxes.
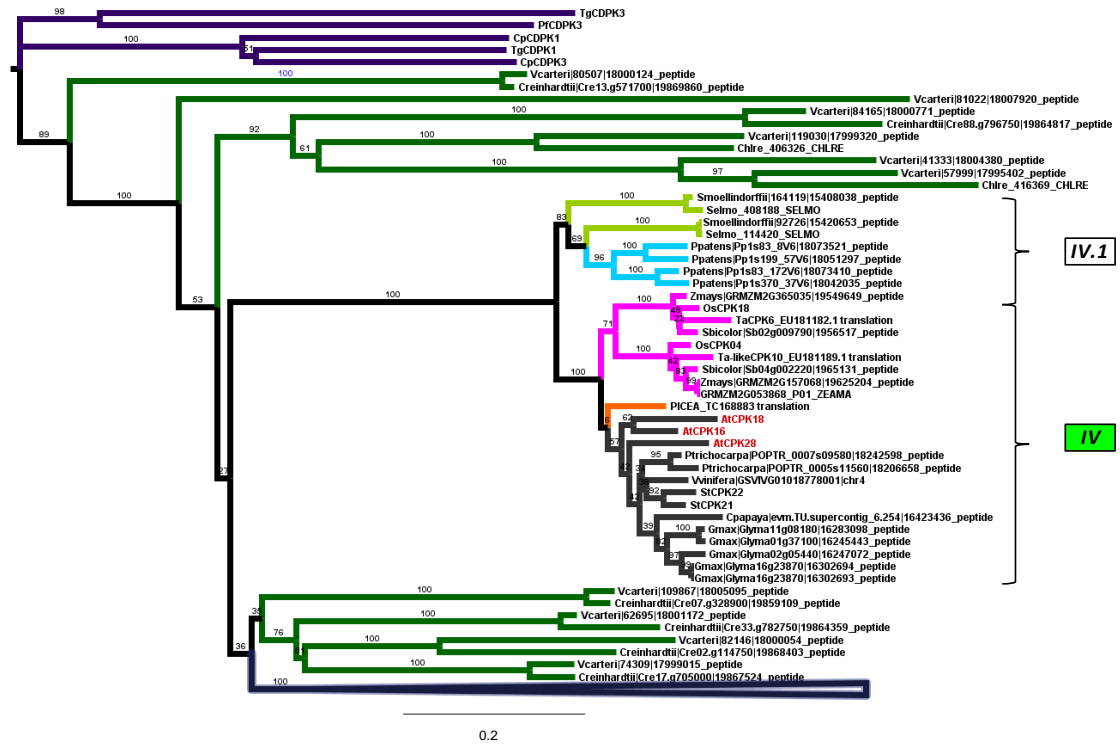
**Appendix 7.** NJ tree of all CPK sequences included in this project, sequences trimmed to include conserved regions only (Chapter 3.3.1)

(tree continued next page)

Phylogenetic tree (tip labels and support values):

Ptrichocarpa|POPTR_0019s00030|18219607_peptide
89 Ppatens|Pp1s187_88V6|18039259_peptide
100 Ppatens|Pp1s49_208V6|18041343_peptide
100 Phypa_168902_PHYPA
Ppatens|Pp1s205_14V6|18038182_peptide
79 Smoellindorfii|99178|15414412_peptide
Smoellindorfii|165073|15410490_peptide
AtCPK25
Cpapaya|evm.TU.supercontig_222.24|16413993_peptide
98 StCPK11
StCPK12
100 Sbicolor|Sb08g014910|1978240_peptide — OsCPK27
95 Zmays|GRMZM2G353957|19526704_peptide
OsCPK10
100 TaCPK16_EU181191.1 translation
51 Zmays|GRMZM2G320506|19631025_peptide
Sbicolor|Sb01g005780|1949995_peptide
Zmays|GRMZM2G028926|19608264_peptide
Cpapaya|evm.TU.supercontig_33.122|16417567_peptide
100 GSVIVP00002665001_VITVI
93 AtCPK01
100 AtCPK02
99 Gmax|Glyma12g3620|16279817_peptide
Gmax|Glyma01g17020|16316077_peptide
Ptrichocarpa|POPTR_0008s01530|18247599_peptide
98 Ptrichocarpa|POPTR_0010s25090|18242383_peptide
StCPK06
StCPK07
59 OsCPK11
100 Zmays|GRMZM2G027351|19637938_peptide
100 Zmays|GRMZM2G121228|19610002_peptide
Sbicolor|Sb01g005750|1949992_peptide
78 Sbicolor|Sb02g003500|1955676_peptide
Zmays|GRMZM2G012326|19558635_peptide
TaCPK9_EU181184.1 translation
85 OsCPK17
AtCPK20
100 GSVIVP00027136001_VITVI
Ptrichocarpa|POPTR_0006s21390|18213080_peptide
100 Ptrichocarpa|POPTR_0016s06590|18250967_peptide
StCPK09
100 StCPK08
StCPK10
94 Cpapaya|evm.TU.supercontig_222.14|16413982_peptide
Gmax|Glyma10g11020|16279220_peptide
70 Gmax|Glyma02g34890|16249087_peptide
100 Gmax|Glyma03g36240|16253515_peptide
53 Gmax|Glyma19g38890|16314250_peptide
GSVIVP00003538001_VITVI
100 Popal_561172_POPTR
Ptrichocarpa|POPTR_0016s06700|18249934_peptide
PICEA_TC127192 translation
56 100 OsCPK06
100 Sbicolor|Sb04g038450|1968759_peptide
Zmays|GRMZM2G347047|19587886_peptide
100 TaCPK1_EU181177.1 translation
100 OsCPK07
Sbicolor|Sb01g048570|1954982_peptide
Zmays|GRMZM2G032852|19607533_peptide
97 OsCPK23
100 Sbicolor|Sb01g030450|1952760_peptide
Zmays|GRMZM2G040743|19599198_peptide
87 Zmays|GRMZM2G321239|19622265_peptide
Zmays|GRMZM2G314396|19558760_peptide
Sbicolor|Sb06g026530|1973627_peptide
OsCPK13
100 TaCPK2_EU181178.1 translation
Sbicolor|Sb04g031570|1967955_peptide
Zmays|GRMZM2G081310|19582298_peptide
60 OsCPK05
GSVIVP00036285001_VITVI
93 AtCPK26
AtCPK05
95 StCPK04a
83 StCPK05 (modified)
78 Ptrichocarpa|POPTR_0004s21710|18225180_peptide
Popal_803966_POPTR
55 Cpapaya|evm.TU.supercontig_60.3|16423718_peptide
Gmax|Glyma02g48160|16250593_peptide
Gmax|Glyma14g00320|16293851_peptide
AtCPK19
97 100 AtCPK22 — AtCPK27
AtCPK31
95 88 AtCPK15
73 AtCPK21 — AtCPK23
Vvinifera|GSVIVG01037652001|chr19edited
GSVIVP00010752001_VITVI
Cpapaya|evm.TU.supercontig_12.305|16406865_peptide
StCPK03_AAQ08324 (modified)
100 Cpapaya|evm.TU.supercontig_152.12|16409594_peptide
100 Gmax|Glyma08g42850|16273633_peptide
Gmax|Glyma18g11030|16308539_peptide
100 Gmax|Glyma02g46070|16250354_peptide
Gmax|Glyma14g02680|16294118_peptide
100 Ptrichocarpa|POPTR_0004s01530|18226252_peptide
Ptrichocarpa|POPTR_0021s00750|18214123_peptide
Gmax|Glyma07g39010|16269222_peptide
60 Gmax|Glyma17g01730|16304071_peptide
100 TaCPK5_EU181181.1 translation
OsCPK19
Sbicolor|Sb02g034640|1958814_peptide
100 Zmays|GRMZM2G154489|19532466_peptide
Zmays|GRMZM2G168706|19557632_peptide
StCPK04 (modified)
100 StCPK02_AF418563_1 (modified)
AtCPK09
AtCPK33
100 OsCPK12
54 TaCPK18_EU181187.1 translation
Sbicolor|Sb06g025220|1973468_peptide
Zmays|GRMZM2G112057|19619532_peptide
AtCPK29
86 StCPK16
67 Vvinifera|GSVIVG01008749001|chr18
93 Cpapaya|evm.TU.supercontig_26.269|16415215_peptide
100 Ptrichocarpa|POPTR_0002s01850|18246149_peptide
Ptrichocarpa|POPTR_0005s26640|18207593_peptide
54 Gmax|Glyma17g38050|16307407_peptide
100 Gmax|Glyma17g38040|16307406_peptide
Gmax|Glyma14g40090|16297015_peptide
100 Smoellindorfii|118877|15411727_peptide
Selmo_177720_SELMO
100 Ppatens|Pp1s316_13V6|18051059_peptide
95 Ppatens|Pp1s108_25V6|18070942_peptide
58 99 Ppatens|Pp1s108_32V6|18070898_peptide
Ppatens|Pp1s143_92V6|18045772_peptide
88 Ppatens|Pp1s325_31V6|18045877_peptide
Ppatens|Pp1s96_216V6|18068611_peptide
84 80 Smoellindorfii|96034|15404581_peptide
100 Selmo_231639_SELMO
Smoellindorfii|231127|15414521_peptide
68 PICEA_TC137067 translation
PICEA_TC148796 translation
100 PICEA_TC167531 translation
PICEA_TC132751 translation
25 PICEA_TC135188 translation
83 StCPK20
92 Cpapaya|evm.TU.supercontig_122.15|16407061_peptide
Gmax|Glyma11g02260|16282406_peptide
98 100 Gmax|Glyma05g37260|16260996_peptide
Gmax|Glyma08g02300|16269679_peptide
52 GSVIVP00002511001_VITVI
Ptrichocarpa|POPTR_0001s10070|18237608_peptide
Ptrichocarpa|POPTR_0003s13380|18216619_peptide
AtCPK03
100 OsCPK15
100 Sbicolor|Sb09g029950|1982027_peptide
Zmays|GRMZM2G058305|19575516_peptide
OsCPK01
99 TaCPK8_EU181183.1 translation
Zmays|GRMZM5G856738|19523505_peptide
Sbicolor|Sb03g028340|1962510_peptide
100 Zmays|GRMZM2G025387|19568531_peptide
OsCPK25
100 OsCPK26
100 Sbicolor|Sb08g001380|1977263_peptide
Zmays|GRMZM2G472311|19578231_peptide
Sbicolor|Sb05g002110|1969059_peptide
Zmays|GRMZM2G365815|19549427_peptide
Sbicolor|Sb03g037570|1963652_peptide
Zmays|GRMZM2G167276|19516616_peptide
TaCPK13_EU181193.1 translation
OsCPK02
OsCPK14
Sbicolor|Sb09g024100|1981329_peptide
Zmays|GRMZM2G340224|19576401_peptide
94 StCPK19
GSVIVP00022286001_VITVI
100 StCPK17
StCPK18
99 AtCPK17
100 AtCPK34
100 Cpapaya|evm.TU.supercontig_157.56|16409845_peptide
100 Ptrichocarpa|POPTR_0001s28150|18237455_peptide
Ptrichocarpa|POPTR_0009s07330|18227524_peptide
96 Gmax|Glyma02g44720|16250188_peptide
Gmax|Glyma14g04010|16294287_peptide
100 Gmax|Glyma07g36000|16268877_peptide
Gmax|Glyma20g08140|16315696_peptide

0.1

**Appendix 8.    Multiple sequence alignments of full CPK sequences (Chapter 3.3.1)**

Please see .fasta file in the attached CD.

(tree continued next page)

**Appendix 10. Detailed topology of CPK evolutionary group I ML tree, full sequences (Chapter 3.3.1)**

Branch colours match the species colour in Figure 3.2. On the far right, the evolutionary groupings are indicated in Roman numerals. Monocot-dicot clusters are highlighted in green.

**Appendix 11.** **Detailed topology of CPK evolutionary group II ML tree, full sequences (Chapter 3.3.1)**

Branch colours match the species colour in Figure 3.2. On the far right, the evolutionary groupings are indicated in Roman numerals. Monocot-dicot clusters are highlighted in green.

## Appendix 12. Detailed topology of CPK evolutionary group III ML tree, full sequences (Chapter 3.3.1)

Branch colours match the species colour in Figure 3.2. On the far right, the evolutionary groupings are indicated in Roman numerals. Monocot-dicot clusters are highlighted in green.

**Appendix 13.** **Detailed topology of CPK evolutionary group IV ML tree, full sequences (Chapter 3.3.1)**

Branch colours match the species colour in Figure 3.2. On the far right, the evolutionary groupings are indicated in Roman numerals. Monocot-dicot clusters are highlighted in green.

Information includes transcript (R) and protein (P) accumulation, enzyme activity (E), mutation lines (M) and overexpression (X) lines. Plus (+), minus (-) and equal (=) signs indicate increase, decrease and no change, respectively, in abundance (transcript or protein), activity (enzyme), sensitivity to hormone/stimuli (mutants/overexpressing plants), resistance to stress (mutants/overexpressing plants), or process (mutants/overexpressing plants).

| Group | CPK Name | Abiotic Stress | | | | | Biotic Stress | | | | Phytohormone interaction/ | | | | | | | Developmental and other signals | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Drought | Salt | Cold | Heat | Wound | Fungi & elicitors | Bacteria and flg22 | Virus | Insects/Herbivory | ABA | GA(3) | IAA | Ethylene | Brassinolide | SA | JA/MeJA | Light response | Stomatal closure | Pollen | Pollen tube dev't/elongation | Seed/early plant dev't | Fruit and tubers | Flower | Root elongation/dev't. | Cotton fibre elongation |
| | AtCPK16 | | | | | | | | | | | | | | | | | | | R+ | | | | | | |
| | AtCPK18 | | | | | | | | R- | M= | | | | | | | | | | | | | | | | |
| IV | OsCPK04 | R= | R= | R+ | | | R+ | | | | | | | | | | | | | | | | | | | |
| | TaCPK10 | R= | R+ | R= | | | R+ | | | | R- | R= | | | | | | | | | | | | | | |
| | OsCPK18 | R= | R- | R= | | | R+ | | | | | | | | | | | | | | | | | | | |
| | TaCPK6 | R+ | R= | | | | R= | | | | R+ | R= | | | | | | | | | | | | | | |
| | LeCPK2_GQ205414 | | | R= | R+ | R+ | | | | | | | R+ | | | R+ | R+ | | | | | | | R+ | | |
| | NtCPK04_AF435451 | | | | | | | | | | | | | | | | | | | R+ | | | | R+ | | |
| | MtCPK01_AY821654 | R+ | | | | R+ | | | | | | | | | | | | | | | | | | | M- | |
| | AtCPK28 | | | | | | | | | R+ | R=, E= | | | | | | | | | | | | | | | |
| IIIa | AtCPK24 | | | | | | | | | | | | | | | | | | | R+ P+, X- | | | | | | |
| | PiCDPK2_DQ147912 | | | | | | | | | | | | | | | | | | | R+, X-, M- | | | | | | |
| | OsCPK29 | | | | | | | | | | | | | | | | | | | R+ | | | | | | |
| | OsCPK21 | R+ | R+, X+ | R+ | | | | | | | X+ | | | | | | | | | R+ | | | | | | |
| | OsCPK22 | R= | R= | R= | | | | | | | | | | | | | | | | R+ | | | | R+ | | |
| IIIb | OsCPK8 | | | | | | R= | | | | | | | | | | | | | | | | | | | |
| | TaCPK12 | R= | R= | R+ | | | R+ | | | | R= | R+ | | | | | | | | | | | | | | |
| | CarCPK02_AY312269 | | R+ | | | | | X+ | | | | | | | | | | | | | | | | R+ | | |
| | OsCPK20 | | | | | | | | | | | | | | | | | | | R+ | | | | | | |
| | TaCPK7 | R= | R= | R+ | | | R+ | | | | R- | R= | | | | | | | | | | | | | | |
| | AtCPK7 | | | | | | | | | M= | | | | | | | | | | | | | | | | |
| | AtCPK8 | | | | | | | | | M= | | | | | | | | | R- | | | | | | | |
| | AtCPK14 | | | | | | R+ | | | | | | | | | | | | | P+ | X- | | | | | |
| | AtCPK32 | | R+ | R= | | | | X+ | R+ | | R= | | | | | | | | | P+ | X- | | | R+ | | |
| | AtCPK13 | | | | | | | X+ | M- | | | | | | | | | | | | | | | | | |
| | TaCPK3 | R= | R= | R- | | | R+ | | | | R+ | R+ | | | | | | | | | | | | | | |
| | OsCPK16 | R- | | | | | | | | | | | | | | | | | | | | | | | | |
| | TaCPK15 | R= | R= | R+ | | | R+ | | | | R= | R+ | | | | | | | | | | | | | | |
| | AtCPK10 | M-, X+ | | | | | | X+ | | M= | M- | | | | | | | | M- | | | | | | | |
| | AtCPK30 | | | | | | | X+ | | | R+ | R+ | | | | | | | | | | | | | | |
| | OsCPK9 | | | | | | R+ | | | | | | | | | | | | | | | | | | | |
| | TaCPK19 | R= | R+ | R= | | | R+ | | | | R= | R= | | | | | | | | | | | | | | |
| I | OsCPK23 | | | | | | | | | | | | | | | | | | | R+ | | | | | | |
| | OsCPK07 | R+, P+ | R+, P+, X+ | R+, X+ | | | M= | | | | R= | | | | | | | | | | | | | | | |
| | TaCPK01 | R+ | R= | R- | | | R- | | | | R= | R= | | | | | | | | | R+ | | | | | |
| | OsCPK05 | R+ | R+ | R+ | | | | | | | | | | | | | | | | | | | | | | |
| | OsCPK13 | R+, P- | P- | R+, P+, X+ | | | | | | | P- | P+ | | | | | | | | | | | | | | |
| | TaCPK02 | R= | R= | | | | R+ | | | | R= | R+ | | | | | | | | | | | | | | |
| | AtCPK26 | | | | | | | X+ | | | | | | | | | | | | | | | | | | |
| | CanCDPK03_AY295081 | R+ | R+ | | | | | R+ | | | R+ | | | | R+ | R+ | R+ | | | | | | | | | |
| | VrCDPK_AAC49405 | R+ | | | | R+ | | | | | R+ | | | | | | | | | | | | | | | |
| | AtCPK05 | | | | | | | X+, M- | M= | | | | | | | | | | | | | | | | | |
| | AtCPK06 | R+ | R+ | | | | | X+, M- | | | M- | | | | | M- | | | M- | | | | | | | |
| | OsCPK27 | R= | R= | | | | | | | | | | | | | | | | | R+ | | | | R+ | | |
| | PaCDPK1_EF555574 | | | R+ | | R+ | R+ | | | | | | | | | | | | | | | | | | | |
| | GhCDPK1_FJ938290 | | | | | | | | | | | | | | | | | | | | | | | | | R+ |
| | AtCPK2 | | | | | | R= | | M= | | | | | | | | | | | | | | | | | |
| | AtCPK01 | | | | | | M-, R+ | M- | | | | | | | | X+ | | | | | | | | | | |
| | IiCPK02_DQ458916 | | R+ | R+ | | | | | | | R+ | | | | | | | | | | | | | | | |
| | OsCPK10 | R+ | R= | R= | | | | | | | | | | | | | | | | | | | | | | |
| | ZmCPK10 | | | | | | R+ | | | | | | | | | | | | | | | | | | | |
| | AtCPK20 | | | | | | R= | | M= | | | | | | | | | | | | | | | | | |
| | OsCPK11 | R= | R= | R= | | | | | | | | | | | | | | P= | | | | | | R+ | | |
| | OsCPK17 | | | | | | R= | | | | | | | | | | | | | | | | | | | |
| | TaCPK09 | R+ | R- | | | | | | | | R+ | | | | | | | | | | | | | | | |
| | AtCPK04 | X+ | M- | | | | | X+, M- | | | M-, X+ | | | | | M= | | | | | | | | | | |
| | AtCPK11 | X+ | M- | | | | | X+, M- | M= | | M-, X+ | | | | | M= | | | R+ | | | | | | | |
| | OsCPK28 | | | | | | | | | | | | | | | | | | | | | | | R+ | | |
| | ZmCPK11 | | | | | E+, R+ | | | | | | | | | | | | | | | | | | R+ | | |
| | OsCPK24 | | | | | | | | | | | | | | | | | | | | | | | R+ | | |
| | TaCPK04 | R= | R+ | R- | | | R+ | | | | R- | R+ | | | | | | | | | | | | | | |
| | VfCDPK1_AY753552 | R+, P+ | R= | | | | | | | | R+ | | | | | | | | | | | | | | | |
| | VvCPK01_AY394009 | | | | | | | | | | P+ | | | | | | | | R+ | | | P+ | | | | |
| IIa | OsCPK12 | R+ | X+, M- | | | | X- | | | | X+ | | | | | | | | | | | | | | | |
| | TaCPK18 | R- | R+ | R= | | | R= | | | | R= | R= | | | | | | | | | | | | | | |
| | OsCPK19 | | | | | | R= | | | | | R+ | R+ | | | | | | | | | | | R- | | |
| | TaCPK05 | R= | R= | R+ | | | R= | | | | R+ | R+ | | | | | | | | | | | | | | |
| | AtCPK33 | | | | | | | | M= | | R+ | | | | | | | | | | | | | | | |
| | AtCPK9 | | | | | | | | | | | | | | | | | | | P- | | | | | | |
| | McCDPK01_AF090835 | R+ | R+ | | | | | | | | | | | | | | | | | | | | | | | |
| | StDPK1_AF115406 | | | | | | | | | | R+ | R+ | R+ | | | R-, E- | | | | | | | | R+ | | |
| | LeCDPK1_AF363784 | | | | | R+ | R+ | | | | | | | | | | | | | | | | | | R+ | R- |
| | StCDPK2_AF418563 | | | | | | | | | | | | | | | E- | | R+ | | | | | | | | |
| | CarCPK01_AY312268 | | R+, P+ | | | | R+, P+ | | | | R= | R= | | | | | | | | | | | | | | |
| | AtCPK19 | | | | | | | | M= | | | | | | | | | | | | | | | | | |
| | AtCPK22 | | | | | | | | M= | | | | | | | | | | | | | | | | | |
| | AtCPK21 | R+, E+, M- | R+, E+, M- | | | | | | M= | | | | | | | | | | | | | | | | | |
| | AtCPK23 | R+, M+, X- | R+, M+, X- | | | | | | | | R+ | R+ | | | | | | | | | M+, X- | | | | | |
| | HbCDPK01_EU581818 | | | | | R+ | | | | | | | | | | R+ | | | | | | | | | | |
| | NtCDPK1_AF072908 | | | | | | | | | | | | E+ | | | | | | | | | | | | | |
| IIb | AtCPK03 | R- | R-, M-, X+ | X+ | X+ | | | X+ | M- | | | | | | | M= | | | M- | | | | | | | |
| | OsCPK01 | R- | R- | | | | | | | | | | | | | | | | | | | | | | | |
| | OsCPK15 | R+ | R+ | | | | R= | | | | | | | | | | | | | | | | | | R- | |
| | MpCDPKA_AB017516 | | | | | | | | | | | | | | | | | | | R+ | | | | | | |
| | PiCDPK01_DQ147913 | | | | | | | | | | | | | | | | | | | R+ | | | R+ | | | |
| | AtCPK17 | | | | | | | | | | | | | | | | | | | P+ | M- | | | | | |
| | AtCPK34 | | | | | | | | | | | | | | | | | | | P+ | M-, Xr | | | | | |
| | OsCPK02 | | | | | | | | | | | | | | | | | P- | | | | | X- | | R+ | |
| | OsCPK14 | | | | | | | | | | | | | | | | | | | | | | | | R+ | |
| | OsCPK25 | | | | | | | | | | | | | | | | | | | R+ | | | | | R+ | |
| | OsCPK26 | | | | | | | | | | | | | | | | | | | R+ | | | | | R+ | |

**Appendix 15. Overview of plant CPK expression patterns, localisation and activity based on published literature (Chapter 3.3.2)**

| CPK Name and source organism | Experimental conditions tested/Tissue or cellular localization | Expression patterns in response to condition | Methodology used (RT-PCR, qPCR, microarray, etc.) | Reference |
|---|---|---|---|---|
| GROUP I | | | | |
| AtCPK1 (*Arabidopsis thaliana*) | Fungal elicitors- *Fusarium oxysporum* | Increase in transcript accumulation upon 30 mins. treatment with elicitor (RT-PCR); Rapid (within 30 mins), but transient (normal after 6 hrs) increase in transcript accumulation (RNA gel blot); Increase in polypeptide levels upon 2 hrs treatment with elicitor (2D-PAGE) | RT-PCR, RNA gel-blot, 2D-PAGE of control & elicitor treated 15-day old plants | (Coca and San Segundo 2010) |
| | Peroxisome association | AtCPK1-GFP was found to be targeting peroxisomes, suggesting CPK involvement in peroxisomal functions in oxidative stress and lipid metabolism | CPK-GFP-fusion proteins (cloned between 35S-derived promoter and a NOS terminator) & fluorescence microscopy in cells near the root tip, membrane fractionation experiments | (Dammann et al. 2003) |
| | Cellular localization | Dual localization of AtCPK1 in lipid bodies and peroxisomes | Transient expression of AtCPK1–GFP fusion gene in suspension-cultured and root cells in Arabidopsis and epidermal cells in onion | (Coca and San Segundo 2010) |
| AtCPK2 (*Arabidopsis thaliana*) | Fungal elicitors- *F .oxysporum* | No significant change in transcript accumulation | RT-PCR control & elicitor treated 15-day old plants. | (Coca and San Segundo 2010) |
| AtCPK4 & 11 (*Arabidopsis thaliana*) | Cellular localization | Nuclear and cytosolic distribution of AtCPK4-GFP is similar to free GFP. AtCPK4 is primarily a soluble protein. | CPK-GFP-fusion proteins (cloned between 35S-derived promoter and a NOS terminator) & fluorescence microscopy in cells near the root tip, membrane fractionation experiments | (Dammann et al. 2003) |
| | | Both have nuclear and cytosolic localization | CPK-GFP-fusion proteins (cloned between 35S-derived promoter and a NOS terminator) | (Boudsocq et al. 2010) |
| | ABA-signalling regulation | CPK 4 and 11 both phosphorylate ABA-responsive transcription factors ABF1 and ABF4 | In-gel kinase and autophosphoryation assays | (Zhu et al. 2007) |
| AtCPK5 & 6 (*Arabidopsis thaliana*) | Cellular localization | Nuclear and cytosolic | CPK-GFP-fusion proteins (cloned between 35S-derived promoter and a NOS terminator) | (Boudsocq et al. 2010) |
| | Development, ubiquitous expression | AtCPK6 transcripts are present in leaf, stem and roots. Proteins are present ubiquitously. Protein level declined during pollen maturation. | Northern, southern and western blots, immunohistochemistry. | Hong et al. 1996 |
| | Salt and drought stress response | AtCPK6 transcript levels increased 6-10 fold within 10 mins after the start of salt or drought stress treatment, peaked at 12-fold by 1 h and 4-fold from 4 to 24 h | RT-qPCR, 3-week-old plants grown on MS solid medium were subjected to salt or drought stress by transferring into MS solution containing 250 mM NaCl or 15% polyethyleneglycol (PEG) 6000 | (Xu et al. 2010) |
| AtCPK20 (*Arabidopsis thaliana*) | Fungal elicitors- F. oxysporum | Transcript levels not detectable in plants even if induced by fungal elicitors | RT-PCR control & elicitor treated 15-day old plants. | (Coca and San Segundo 2010) |
| OsCPK05 (*Oryza sativa*) | Phytohormone response | Increased transcript accumulation in | Microarray | (Ye et al. 2009) |

| | | phytohormones GA3, NAA and KT | | |
|---|---|---|---|---|
| **OsCPK07** *(Oryza sativa)* | Expression levels in development and anatomy | Nearly constitutive throughout development, but slightly upregulated (6- and 14-folds ) in panicles and seeds | Microarray, qPCR | (Ray, et al., 2007) |
| | Endosymbiotic relationship with a fungus *Glomus intraradices* | Moderate to low-level transcripts in response to inoculation (presymbiotic phase). | Semi-quantitative RT-PCR, qRT-PCR, (24, 48, 72 and 96 hours after inoculation) | (Campos-Soriano, et al., 2011 |
| | Cold & salt stress response | OsCPK7 transcript accumulation were increased by cold (4°C) and salt stress (200mM NaCl) treatment but not by ABA application. | Northern blot (8 & 24 h after treatment) | (Saijo et al. 2000) |
| | | High transcript and protein levels in vascular tissues and sclerenchyma in roots under salt and drought stress. Has similar localisation patterns with rab16A, a stress target gene | In situ hybridisation, immunolocalisation, overexpression | (Saijo et al. 2001) |
| **OsCPK10** *(Oryza sativa)* | Expression levels in development and anatomy | 3 to 4-fold increase under desiccation stress (air-dried on a Whatman 3 mm sheet at 28°C for 3 h), 1.5-folds decrease in transcript levels upon salt stress (200 mM NaCl) for 3 h. | Microarray, qPCR | (Ray et al. 2007) |
| | Phytohormone response | Increased transcript accumulation in phytohormones GA3, NAA and KT | Microarray | (Ye et al. 2009) |
| | Endosymbiotic relationship with a fungus *Glomus intraradices* | Moderate to low-level transcripts in response to inoculation (presymbiotic phase). | Semi-quantitative RT-PCR, qRT-PCR, (24, 48, 72 and 96 hours after inoculation) | (Campos-Soriano et al. 2011) |
| **OsCPK11** *(Oryza sativa)* | Expression levels in development and anatomy | Almost undetectable in all levels | Microarray, qPCR | (Ray et al. 2007) |
| | | Predominantly high transcript accumulation in panicle and stamen | Microarray | (Ye et al. 2009) |
| **OsCPK13** *(Oryza sativa)* | Expression levels in development and anatomy; Response to cold, salt, drought & ABA | Nearly constitutive throughout development, but slightly down-regulated (>2-folds) in later seed stages. Transcript levels increase 2.14-fold in cold stress (4°C), 3 to 4-fold increase under desiccation stress (air-dried on a Whatman 3 mm sheet at 28°C for 3 h), 1.84-fold increase in transcript levels upon salt stress (200 mM NaCl) for 3 h | Microarray, qPCR | (Ray et al. 2007) |
| | | High transcript accumulation in leaf, sheath and root tissues, increase in response to phytohormones GA3, NAA and KT | Microarray | (Ye et al. 2009) |
| | | Transcript and protein accumulation were enhanced by cold stress and GA, but suppressed by ABA, drought (withholding water for 2 weeks) and salt (100mM NaCl for 24 h) stress. OsCPK13 transcript levels were higher in cold-tolerant plants than in cold-sensitive plants. | Western blot, northern blot, kinase assays | (Abbasi et al. 2004) |
| | | Highly cold-tolerant rice varieties have higher OsCPK13 protein accumulation than intermediate varieties | 2D-page, western blots, kinase assays. | (Komatsu et al. 2007) |
| | | | | |

| | | | | |
|---|---|---|---|---|
| **OsCPK17** *(Oryza sativa)* **OsCPK24** *(Oryza sativa)* | Endosymbiotic relationship with a fungus *Glomus intraradices* | Moderate to low-level transcripts in response to inoculation (presymbiotic phase). | Semi-quantitative RT-PCR, qRT-PCR, (24, 48, 72 and 96 hours after inoculation) | (Campos-Soriano et al. 2011) |
| **OsCPK23** *(Oryza sativa)* | Expression levels in development and anatomy | High transcript accumulation in and panicle development stage | Microarray, qPCR | (Ray et al. 2007) (Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007)(Ray et al. 2007) |
| | | 87- and 1,724-folds up-regulated in panicle and seed developmental stages, respectively, and higher expression (53-folds) in roots in comparison to mature leaf. | | |
| **OsCPK27** *(Oryza sativa)* | Expression levels in development and anatomy | Almost undetectable in all levels | | |
| | | Predominantly high transcript accumulation in panicle and stamen | Microarray | (Ye et al. 2009) |
| **TaCPK01** *(Triticum aestivum)* | Expression levels in stress response | Increased transcript accumulation in drought and $H_2O_2$; decreased transcript accumulation in cold and *Blumeria graminis tritici* (Bgt); not expressed/no change in other treatments. | Semi-quantitative RT-PCR 7-day-old seedlings were transferred to a Hogland solution containing 200 mM NaCl (salt stress), 16% PEG (drought stress), 10 mM $H_2O_2$ , 5 uM ABA &GA (hormone) or placed at 4°C (cold stress). | (Li et al. 2008b) |
| **TaCPK02** *(Triticum aestivum)* | | Increased transcript accumulation in $H_2O_2$ and *Blumeria graminis tritici* (Bgt); not expressed/no change in other treatments. | | |
| **TaCPK04** *(Triticum aestivum)* | | Increased transcript accumulation in salt, GA and *Blumeria graminis tritici* (Bgt); decreased transcript accumulation in cold and ABA; not expressed/no change in other treatments | | |
| **TaCPK09** *(Triticum aestivum)* | | Increased transcript accumulation in drought, ABA and $H_2O_2$ ; decreased transcript accumulation in salt; | | |

| | | not expressed/no change in other treatments | | |
|---|---|---|---|---|
| **PaCPK1** *(Phalaenopsis amabilis)* | | Transcripts were high in labellum and in periolic flowers but not in leaves and roots. Increased transcription in response to low temperature, wounding, and bacteria (*Erwinia chrysanthemi*) pathogen infection. | GUS fused to the PaCPK1 promoter | (Tsai et al. 2007) |
| **VfCPK1** *(Vicia faba)* | Expression levels in anatomy and stress response | Highest expression in epidermal peels of broad bean leaves. Transcript levels reached its maximum from time point 5 h to 10 h and decreased at 24 h post-drought (PEG 8000) and ABA treatment. Protein levels also increased 5-10 post drought. External application of NaCl and MgCl2 had little effect but Ca increased expression. | Western and RNA blot analyses | (Liu et al. 2006) |
| **VrCPK1** *(Vigna radiata)* | Expression levels in anatomy and stress response | Increase in transcript levels in response to mechanical strain (starting from 30 min and reached max at 60min); 500 uM IAA treatment (response at 6-9 h, peak at 6 h); and salt stress (50mMNaCl; response at 2-6 h) | Northern analysis, 10-day-old plants | (Botella et al. 1996) |
| **VvCPK01** AY394009 *(Vitis vinifera x Vitis labrusca)* | Expression – anatomy and ABA response | Transcripts are undetectable in roots, young stems, or leaves portions, but high in the fleshy mesocarp and seeds with much higher abundance in the mesocarp than in seed. ABA stimulates grape berry CPK transcript and protein accumulation, and increases enzyme activity during fruit development. Other phytohormones GA, IAA, cytokinin and brassinolide do not stimulate this CPK. | Immunoblotting, western, northern and southern blotting, kinase assays | (Yu et al. 2006) |
| **ZmCPK10** *(Zea mays)* | Fungal-*Fusarium moniliforme* infection and elicitors | Transcript levels are higher in infected germinating embryos than control. Transcript rapidly accumulates upon elicitor treatment (5-30 mins peak, followed by decline). | Northern blots | (Murillo et al. 2001) |
| | Regulation of pathogen response genes | ZmCPK10 is present in cell types where *PRms* (pathogenesis-related protein in maize) gene is also present. | Northern blots and in situ hybridisation | (Murillo et al. 2001) |
| **ZmCPK11** *(Zea mays)* | Expression levels in development and anatomy | Ubiquitous, but high in seeds & seedlings and lower in stems, roots & leaves | RT-PCR | (Szczegielniak et al. 2005) |
| | Response to abiotic stress | Cold (4°C) and heat(40°C), 500 mM $H_2O_2$, 300 mM NaCl, desiccation, and 100 mM ABA have no significant effect on transcript levels | RT-PCR | |
| | Response to wounding | Transcript accumulation in leaves after wounding (1, 3, 6 & 24 h). Neighbouring leaves also have elevated ZmCPK11, indicating a systemic response. | RT-PCR | |
| **CanCPK3** AY295081 *(Capsicum annuum)* | Expression in anatomy and stress response | Transcripts are abundant in root tissue from non-stressed plants and detectable in closed and open flower tissues, but not in non-stressed leaves and young fruits. Transcript levels increased after exposure to salt stress (1.5 to 12 h), SA, MeJA, ethephon (ethylene generator) | northern blot analysis | (Chung et al. 2004) |

| | | | | |
|---|---|---|---|---|
| | | and non-compatible bacteria *Xanthomonas axonopodis* infection. | | |
| **GhCPK1** *(Gossypium hirsutum)* | Cotton fibre elongation via ethylene biosynthesis | GhCPK1 protein interacts with GhACS2 (1-aminocyclopropane-1-carboxylic acid synthase), a rate-limiting enzyme in ethylene production. GhACS2 S460 is a possible phosphorylation site for this interaction. GhCPK32 does not perform this interaction. | Wheat germ cell free system, and were subjected to immunoprecipitation analysis. | (Wang et al. 2011b) |
| **IiCPK02** DQ458916 *(Isatis indigotica)* | Expression in stress response | Transcript levels increase in NaCl and GA3 (reached the highest at 16 h then decreased gradually thereafter), as well as in cold stress (16-32h). Transcript levels were not affected by ABA treatment. | semi-quantitative RT-PCR; seedlings were sprayed with solution of 100 µM GA3, 100 µM Abscisic Acid (ABA) and 250 mM NaCl; cold treatment at $4^0$C | (Lu et al. 2006) |
| **GROUP IIA** | | | | |
| **AtCPK9** *(Arabidopsis thaliana)* | Cellular localization | Targeted exclusively to the plasma membrane | CPK-GFP-fusion proteins (cloned between 35S-derived promoter and a NOS terminator) & fluorescence microscopy in cells near the root tip, membrane fractionation experiments | (Dammann et al. 2003) |
| | Development, ubiquitous expression | Transcripts are present in leaf, stem and roots. Proteins are present ubiquitously. Protein level declined during pollen maturation. | Northern, southern and western blots, immunohistochemistry. | Hong et al. 1996 |
| **AtCPK21** *(Arabidopsis thaliana)* | Cellular localization | Targeted exclusively to the plasma membrane | CPK-GFP-fusion proteins (cloned between 35S-derived promoter and a NOS terminator) & fluorescence microscopy in cells near the root tip, membrane fractionation experiments | (Dammann et al. 2003) |
| **OsCPK12** *(Oryza sativa)* | Expression levels in development and anatomy | 3 to 4-fold increase under desiccation stress (air-dried on a Whatman 3 mm sheet at 28°C for 3 h) | Microarray, qPCR | (Ray et al. 2007) |
| **OsCPK19** *(Oryza sativa)* | Expression levels in development and anatomy | Nearly constitutive throughout development, but slightly down-regulated (>2-folds) in later seed stages. | Microarray, qPCR | (Ray et al. 2007; Ye et al. 2009) |
| | Phytohormone response | Increased transcript accumulation in phytohormones GA3 and IAA | Microarray | (Ye et al. 2009) |
| | Endosymbiotic relationship with a fungus *Glomus intraradices* | Moderate to low-level transcripts in response to inoculation (presymbiotic phase). | Semi-quantitative RT-PCR, qRT-PCR, (24, 48, 72 and 96 hours after inoculation) | (Campos-Soriano et al. 2011) |
| **TaCPK5** *(Triticum aestivum)* | Expression levels in stress response | Increased transcript accumulation in $H_2O_2$, GA cold and ABA; not expressed/no change in other treatments | Semi-quantitative RT-PCR 7-day-old seedlings were transferred to a Hogland solution containing 200 mM NaCl (salt stress), 16% PEG (drought stress), 10 mM $H_2O_2$ , 5 uM ABA & GA (hormone) or placed at 4°C (cold stress). | (Li et al. 2008b) |
| **TaCPK18** *(Triticum aestivum)* | Expression levels in stress response | Increased transcript accumulation in salt and $H_2O_2$,; decreased transcript accumulation in drought; not expressed/no change in other treatments | | |
| **StCPK1** *(Solanum tuberosum)* | Tuber development | Increase in mRNA in early stages of tuber development. Hybridisation showed increased expression in induced stolons but not in leaves, shoots, petioles, or stolons. Differentially expressed at the onset of tuberisation. | Northern blot analysis | (Raices et al. 2001) |

| StCPK2 *(Solanum tuberosum)* | Expression-anatomy and phytohormone | Expressed in leaves. Enzyme activity decreased after just 4 hours, increased/ greater exposure to JA increased inhibition and transcript levels (northern blots) were downergulated in a rapid manner (20 hours). | In gel kinase assays, northern blots | (Ulloa et al. 2002b) |
|---|---|---|---|---|
| CarCPK01 AY312268 *(Cicer arietinum)* | Expression-anatomy & stress response | Transcript found in roots, shoots, leaves, but NOT in fruits or flowers. Transcript levels are highest in roots. Protein kinase activity matched the transcript levels.<br><br>Transcript & protein accumulation and protein kinase activity increased in response to salt (max transcript increase 8-fold at 2 h and down to 4-fold by 4 and 8 h), and *Aspergillus* fungal spores (max transcript increase 10-fold at 1h and basal by 12 h); did not change in response to dehydration, GA and IAA. Localised in the leaf mesophyll cells in the chloroplast and stem xylem parnchyma cells in the PM. | qPCR, protein gel blots, immuneprecipitation and kinase assays. Stress treatments are 100mM NaCl (salt), excising leaves and placing into Whatman paper (dehydration) and *Aspergillus* spores suspension (fungi) | (Syam Prakash and Jayabaskaran 2006) |
| HbCPK01 EU581818 *Hevea brasiliensis* | Expression-anatomy | Present in EST library of ethephon-responsive genes. Transcripts present in all tissues sampled highest in latex (laticifer preferential). Transcript accumulation strongly induced by mechanical wounding, jasmonic acid (JA), and ethephon | Semi-quantitative RT-PCR expression analysis post treatment | (Zhu et al. 2010) |
| LeCPK1 *(Solanum lycopersicum,* formerly *Lycopersicum esculentum)* | Expression in anatomy and stress response | Transcript ubiquitous, but were lower in roots and higher in leaves and flowers. Transcript levels increase in response to chitosan (derived from pathogen cell walls; rapidly and remained high 1 – 4 h); Polygalacturonide [PGA] (derived from plant cell walls; max. at 4 h and rapidly declined down at 8 h); $H_2O_2$ (rapidly (1 h) and remained high up to 4 h); and wounding of leaves (rapid, from 0-4 h, maintained until 8 h and declined thereafter. Neighbouring unwounded leaves had a 4 hour delay but with an identical response | Northern blot analysis, detached leaf assays | (Chico et al. 2002) |
| McCPK01 *(Mesembryanthe-mum crystallinum)* | Stress response | Transcript levels increase in salinity and drought-induced; protein co-localises to the nuclei with CSP1 (CPK-substrate protein 1) under salt stress and phosphorylates it. The physiological response of this phosphorylation is unknown | RT-PCR, east-2 hybrid, wheat germ interaction assay, in vitro kinase assay | (Patharkar and Cushman 2000) |
| NtCPK1 [AF072908] *(Nicotiana tabacum)* | Transcriptional regulation in response to GA | Interacts in a Ca2+-dependent manner *in vivo* and *in vitro* with RSG ('repression of shoot growth'), a transcriptional activator important in GA feedback | In-gel kinase assay, immunoblot, immunoprecipitation | (Ishida et al. 2008) |
| **GROUP IIB** | | | | |
| AtCPK3 *(Arabidopsis thaliana)* | Tissue localization | Transcript was identified in phloem exudates (sap). Phloem-derived RNAs in this study were | Microarray analysis of sieve elements (transcripts | (Deeken et al. 2008) |

| | | considered as mobile, potential long-distance signals. | from phloem sap) and parenchyma cells. | |
|---|---|---|---|---|
| | Phosphorylation targets | CPK3 phosphorylates ERF1, HsfB2a, and CZF1/ZFAR1 in the presence of $Ca^{2+}$. CPK3-derived phosphorylation of a heat shock factor HsfB2a promotes PDF1.2 transcriptional activation in the defense response | *In vitro* kinase assays | (Kanchiswamy et al. 2010) |
| | Kinase activity | HA-epitope tagged CPK3 in protoplasts showed increased protein kinase activity 15 min after treatment with salt (150 mM NaCl), cold (4°C), heat (37°C), 1mM $H_2O_2$, 15nM flagellin and 1mM laminarin. In planta, CPK3 is constitutively active in roots and leaves | Immunocomplex kinase assays in protoplasts (histone III-S) and 6-week old plants (using a CPK3 C-terminal-specific antibody) | (Mehlmer et al. 2010) |
| | Cellular localization | Predominantly localized in the nucleus and plasma membranes; seems to co-localise with chloroplast but shown to be unimportable to chloroplasts in other studies. N-myristoylation is required for plasma membrane and vacuolar localization. | Subcellular fractionation followed by western blot; YFP fusion and microscopy | (Mehlmer et al. 2010) |
| | | Nuclear and cytosolic distribution of AtCPK3-GFP is similar to free GFP. AtCPK3 is primarily a soluble protein. | CPK-GFP-fusion proteins (cloned between 35S-derived promoter and a NOS terminator) & fluorescence microscopy in cells near the root tip, membrane fractionation experiments | (Dammann et al. 2003) |
| **AtCPK17 & 34** *(Arabidopsis thaliana)* | Tissue localization | Preferentially expressed in mat pollen | Microarray data (TAIR) | (Zhou et al. 2009) |
| | | | GUS-promoter activity (CPK34 only) | (Myers et al. 2009) |
| | Cellular localization | Localized uniformly in cytoplasm | AtCPK-GFP fusion proteins, overexpressed by particle bombardment. CPKs were tagged with GFP in the N-terminus. | (Zhou et al. 2009) |
| | | Targeted to the plasma membran | AtCPK-YFP fusion proteins (controlled by ACA9 promoter), confocal microscopy. CPKs were tagged with YFP in the C-terminus. | (Myers et al. 2009) |
| **OsCPK01** *(Oryza sativa)* | Salt and desiccation stress | Significant decrease in transcript accumulation | Microarray, qPCR | (Ray et al. 2007) |
| | Expression levels in development and anatomy | High transcript accumulation in and panicle development stage, >2-folds decrease in transcript levels upon salt stress (200 mM NaCl) for 3 h | Microarray, qPCR | |
| | Phytohormone response | Increased transcript accumulation in phytohormone IAA | Microarray | (Ye et al. 2009) |
| **OsCPK15** *(Oryza sativa)* | Expression levels in development and anatomy | Nearly constitutive throughout development, but slightly down-regulated (>2-folds) in later seed stages. 3 to 4-fold increase under desiccation stress (air-dried on a Whatman 3 mm sheet at 28°C for 3 h). | Microarray, qPCR | (Ray et al. 2007) |
| | Endosymbiotic relationship with a fungus *Glomus intraradices* | Moderate to low-level transcripts in response to inoculation (presymbiotic phase). | Semi-quantitative RT-PCR, qRT-PCR, (24, 48, 72 and 96 hours after inoculation) | (Campos-Soriano et al. 2011) |
| **OsCPK2 & 14** *(Oryza sativa)* | Expression levels in development and anatomy | OsCPK2 & 14 are predominantly high transcript accumulation in panicle and stamen | Microarray | (Ye et al. 2009) |

| | | | | |
|---|---|---|---|---|
| | | OsCPK2 protein levels are at low levels during early seed development; increased and maintained in later stages; and increased during flower development | western and dot blots, in situ hybridisation | (Frattini et al. 1999) |
| | Light response | OsCPK2 protein levels are very low in leaves exposed in light but very high in the dark. | western and dot blots, in situ hybridisation | |
| | | Etiolated plants have high levels of OsCPK2 protein | southern, western and dot blot | (Morello, et al., 2000) |
| **OsCPK25 OsCPK26** (*Oryza sativa*) | Expression levels in development and anatomy | Predominantly high transcript accumulation in panicle and stamen | Microarray | (Ye et al. 2009) |
| **MpCPKA** (*Marchantia polymorpha*) | Expression levels- anatomy | Transcript accumulated in the male sexual organ than the other organs | RT-PCR | (Nishiyama et al. 1999) |
| **PiCPK1** (*Petunia inflate*) | Expression levels- anatomy | Transcripts are highly abundant in pollen; accumulation commences in 15- to 20-mm flower buds, peaks in mature pollen, and remains high after germination.<br><br>Pi CPK1-GFP localized mainly to a thin layer at the periphery of the pollen tube, consistent with a plasma membrane location, though some fluorescence was also visible in the cytoplasm. | RNA gel blot analysis; GFP-fusion proteins and fluorescence microscopy | (Yoon et al. 2006) |
| **GROUP IIIA** | | | | |
| **AtCPK24** (*Arabidopsis thaliana*) | Cellular localization | Localized to the vegetative nucleus and the generative cell/sperms | AtCPK-GFP fusion protein and staining nucleus-sperm organelles with 4'6'-diamidino-2-phenylindole dihydrochloride (DAPI) in tobacco pollen tubes | (Zhou et al. 2009) |
| | Tissue localization | Preferentially expressed in mat pollen | Microarray data (TAIR) | (Zhou et al. 2009) |
| **OsCPK21** (*Oryza sativa*) | Expression levels in development and anatomy | No detectable transcripts in vegetative tissues; transcripts increased 12- and 44-folds during panicle and seed development stages. 3 to 4-fold increase under desiccation stress (air-dried on a Whatman 3 mm sheet at 28°C for 3 h). | Microarray, qPCR | (Ray et al. 2007) |
| | | High transcript accumulation in stamen | Microarray | (Ye et al. 2009) (Ye et al. 2009) (Ye et al. 2009) |
| **OsCPK22** (*Oryza sativa*) | | Predominantly high transcript accumulation in panicle and stamen | | |
| **OsCPK29** (*Oryza sativa*) | | Predominantly high transcript accumulation in panicle and stamen | | |
| **PiCPK2** (*Petunia inflate*) | Expression levels- anatomy | Transcripts are highly abundant in pollen; accumulation commences in 15- to 20-mm flower buds, peaks in mature pollen, and remains high after germination.<br><br>Pi CPK2 does not localize to the plasma membrane but in unknown internal membrane compartments. | RNA gel blot analysis; GFP-fusion proteins and fluorescence microscopy | (Yoon et al. 2006) |
| **GROUP IIIB** | | | | |
| **AtCPK10 & 30** (*Arabidopsis thaliana*) | Tissue localization | AtCPK10 transcript was detectable in all plant tissues or organs; high promoter activities detected in stomata | qRT-PCR, GUS-promoter fusion | (Zou et al. 2010) |
| | Protein interaction | AtCPK10 interacts with heat shock protein 1 (HSP-1) in a $Ca^{2+}$-dependent manner | Yeast two-hybrid methods, GUS staining, and in vitro pull-down assay with His-CPK10 and GST-HSP-1 | |

| | | AtCPK10 & 30 interacts with ABF4 (ABA-responsive element binding factor) | Yeast 2 hybrid screens | (Choi et al. 2005) |
|---|---|---|---|---|
| **AtCPK8** *(Arabidopsis thaliana)* | Development, ubiquitous expression | Transcripts are present in leaf, stem and roots. Proteins are present ubiquitously. Protein level declined during pollen maturation. | Northern, southern and western blots, immunohistochemistry. | Hong et al. 1996 |
| **AtCPK13** *(Arabidopsis thaliana)* | Cellular localization | Nuclear, cytosolic, and plasma membrane localization | CPK-GFP-fusion proteins (35S-derived promoter) & fluorescence microscopy in transiently expressing onion skins | (Kanchiswamy et al. 2010) |
| **AtCPK14 & 32** *(Arabidopsis thaliana)* | Tissue localization | Preferentially expressed in mat pollen | Microarray data (TAIR) | (Zhou et al. 2009) |
| | Cellular localization | Localized uniformly in cytoplasm | AtCPK-GFP fusion proteins, overexpressed by particle bombardment. CPKs were tagged with GFP in the N-terminus. | (Zhou et al. 2009) |
| | Protein interaction | CPK32 interacts with ABF4 (ABA-responsive element binding factor) and the conserved region of ABF1, 2 and 3. | Yeast 2 hybrid screens, GST pulldown assay, northern blot analysis, in vitro kinase assay, overexpression, RTPCR | (Choi et al. 2005) |
| **OsCPK9** *(Oryza sativa)* | Expression levels in development and anatomy | Expressed preferentially in vegetative tissues, highest in mature leaves | Microarray, qPCR | (Ray et al. 2007; Ye et al. 2009) |
| | Rice blast fungus response | Increase in transcript accumulation 12-24 hrs after infection | RNA gel blot analysis | (Asano et al. 2005) |
| **OsCPK8** *(Oryza sativa)* **OsCPK16** *(Oryza sativa)* | Endosymbiotic relationship with a fungus *Glomus intraradices* | Moderate to low-level transcripts in response to inoculation (presymbiotic phase). | Semi-quantitative RT-PCR, qRT-PCR, (24, 48, 72 and 96 hours after inoculation) | (Campos-Soriano et al. 2011) |
| **OsCPK20** *(Oryza sativa)* | Expression levels in development and anatomy | High transcript accumulation in panicle development stage | Microarray, qPCR | (Ray et al. 2007) |
| **TaCPK12** *(Triticum aestivum)* | Expression levels in stress response | Increased transcript accumulation in cold, $H_2O_2$, GA, and in *Blumeria graminis tritici* (Bgt); not expressed/no change in other treatments | Semi-quantitative RT-PCR 7-day-old seedlings were transferred to a Hogland solution containing 200 mM NaCl (salt stress), 16% PEG (drought stress), 10 mM $H_2O_2$ , 5 uM ABA & GA (hormone) or placed at 4°C (cold stress). | (Li et al. 2008b) |
| **TaCPK15** *(Triticum aestivum)* | Expression levels in stress response | Increased transcript accumulation in cold, GA, and in *Blumeria graminis tritici* (Bgt) | | |
| **TaCPK19** *(Triticum aestivum)* | Expression levels in stress response | Increased transcript accumulation in salt and *Blumeria graminis tritici* (Bgt); not expressed/no change in other treatments | | |
| **TaCPK3** *(Triticum aestivum)* | Expression levels in stress response | Increased transcript accumulation in *Blumeria graminis tritici* (Bgt), ABA and GA; decreased transcript accumulation in cold; not expressed/no change in other treatments. | | |
| **TaCPK7** *(Triticum aestivum)* | Expression levels in stress response | Increased transcript accumulation in cold, $H_2O_2$, and in *Blumeria graminis tritici* (Bgt); decreased transcript accumulation in ABA; not expressed/no change in other treatments | | |
| **CarCPK2** *(Cicer arietinum)* | Expression-anatomy & stress response | Transcript found in roots, shoots, leaves, but NOT in fruits or flowers. Transcript levels are highest in roots. Protein kinase activity matched the transcript levels. Transcript & protein accumulation and protein | qPCR, protein gel blots, immuneprecipitation and kinase assays. Stress treatments are 100mM NaCl (salt), excising leaves and placing into Whatman paper (dehydration) and *Aspergillus* spores suspension (fungi) | (Syam Prakash and Jayabaskaran 2006) |

| | | | | |
|---|---|---|---|---|
| | | kinase activity increased in response to salt (max transcript increase 4-fold at 2 h and maintained up to 4 and 8 h), and dehydration (max transcript increase 10-fold at 1h and decreased to 4-fold by 12 h), but did not change in response to *Aspergillus* fungal spores, GA and IAA. Immunolocalisation of CACPk1 and CaCPK2 in the leaf mesophyll cells in the chloroplast and stem xylem parnchyma cells in the PM. | | |
| **AtCPK7 & 8** (*Arabidopsis thaliana*) | Cellular localization | targeted exclusively to the plasma membrane | CPK-GFP-fusion proteins (cloned between 35S-derived promoter and a NOS terminator) & fluorescence microscopy in cells near the root tip, membrane fractionation experiments | (Dammann et al. 2003) |
| **GROUP IV** | | | | |
| **AtCPK16** (*Arabidopsis thaliana*) **AtCPK28** (*Arabidopsis thaliana*) | Cellular localization | targeted exclusively to the plasma membrane | CPK-GFP-fusion proteins (cloned between 35S-derived promoter and a NOS terminator) & fluorescence microscopy in cells near the root tip, membrane fractionation experiments | (Dammann et al. 2003) |
| | Protein interaction | Does not interact with ABF4 (ABA-responsive element binding factor) | Yeast 2 hybrid screens | (Choi et al. 2005) |
| **OsCPK4 & 18** (*Oryza sativa*) | Expression levels in development and anatomy | Both have high transcript accumulation in panicle and root tissues, OsCPK4 transcript level is also high in leaves & sheath tissues | Microarray | (Ye et al. 2009) |
| | | Transcript increased 3.02-folds under cold stress (4°C) | Microarray, qPCR | (Ray et al. 2007) |
| | Endosymbiotic relationship with a fungus *Glomus intraradices* | OsCPK4 and 18 are transcriptionally activated by inoculation (presymbiotic phase) and upregulated by secreted molecules from this fungus. | Semi-quantitative RT-PCR, qRT-PCR, (24, 48, 72 and 96 hours after inoculation) | (Campos-Soriano et al. 2011) |
| **TaCPK6** (*Triticum aestivum*) | Expression levels in stress response | Increased transcript accumulation in drought and ABA, decreased transcript accumulation in cold; not expressed/no change in other treatments | Semi-quantitative RT-PCR 7-day-old seedlings were transferred to a Hogland solution containing 200 mM NaCl (salt stress), 16% PEG (drought stress), 10 mM $H_2O_2$, 5 uM ABA & GA (hormone) or placed at 4°C (cold stress). | (Li et al. 2008b) |
| **TaCPK10** (*Triticum aestivum*) | | Increased transcript accumulation in salt, $H_2O_2$, and in *Blumeria graminis tritici* (Bgt); decreased transcript accumulation in ABA; not expressed/no change in other treatments | | |
| **MtCPK01** (*Medicago truncatula*) | Expression levels in anatomy and stress response | Transcript levels are moderate in flowers, stems, leaves, and roots. In roots, transcript levels increased in response to desiccation and wounding. GUS-CPK1 promoter showed highest expression in the root elongation zone, and in emerging and elongating root hairs | RNA gel blot analysis; GUS-fused with CPK1 promoter | (Ivashuta et al. 2005) |
| **LeCPK2** (*Solanum lycopersicum*, formerly *Lycopersicum esculentum*) | Expression levels in anatomy and stress response | Transcript present in roots, stems, leaves, flowers and fruits but highest in flowers. Increased transcript levels in response to wounding, and phytohormones ethylene | Semi-quantitative RT-PCR | (Chang et al. 2009) |

| | | (ethephon), methyl-jasmonate and salicyclic acid. | | |
|---|---|---|---|---|
| **NtCPK04** *(Nicotiana tabacum)* | Expression levels in anatomy and development | Transcripts are present in all roots, stems, leaves and flowers, but were highest in flowers (also in developing ovules). RNA in-situ studies showed mRNA accumulation at the zones of cell division and the vascular bundle, shoot apical meristem, and lateral branch primordia. | Northern blot analysis ( to untranslated 3' mRNA transctripts); in situ hybridisation | (Zhang et al. 2005) |

**Appendix 16. Overview of plant CPK functional information based on mutation experiments (Chapter 3.3.2)**

| CPK Name and source organism | Mutation/ Constructs | Type of response/ experimental conditions tested | Phenotype | Reference |
|---|---|---|---|---|
| Group I | | | | |
| AtCPK1 (*Arabidopsis thaliana*) | *cpk1-1* (TDI in CAD) *cpk1-2*(TDI in AJ) | *Fusarium oxysporum, Botyritis cinerea,* and *Pseudomonas syringae* infection | Homozygous *cpk1* mutants have complete abolishment of *F. oxysporum* elicitor-induced CPK1 activation (based on RT-PCR). Higher susceptibility to *F. oxysporum, B. cinerea,* and *P. syringae* infection of *cpk1-1* mutants compared to wild type | (Coca and San Segundo 2010) |
| | *AtCPK1* overexpressor (CaMV 35S promoter) and *cpk1* antisense lines (against N-VD) | | Overexpressor plants are less susceptible to all 3 pathogens compared to wildtype. Antisense plants are more susceptible to all 3 pathogens compared to wildtype. Overexpression leads to accumulation of salicylic acid (SA) and constitutive expression of SA-regulated pathogen defence genes (based on microarray & RT-PCR) | |
| AtCPK2 (*Arabidopsis thaliana*) | *cpk2*(TDI) | Plant defensin gene transcription upon insect attack | *cpk2* mutants did not show significant change in transcript levels of PDF1.2 compared to WT plants after 24 hrs attack with *Spodoptera littoralis* larvae (based on RT-qPCR). | (Kanchiswamy et al. 2010) |
| AtCPK4 and 11 (*Arabidopsis thaliana*) | *cpk4-1, cpk11-2* single and *cpk4-1/cpk11-2* double mutant | MeJA-induced stomatal closure | MeJA-induced stomatal closure is not impaired | Munemasa, et al., 2011 |
| | *cpk4-1, cpk11-1* & *cpk11-2* single; *cpk4-1/cpk11-1* and *cpk4-1/cpk11-2* double mutants; and *CPK4* & *CPK11* overexpressors (CaMV 35S promoter) | ABA-signalling regulation | Single mutants exhibit ABA insensitivity in seed development & stomatal movement, salt insensitivity in seed germination, and decreased tolerance of seedlings to salt stress. Double mutants have more severe insensitivity phenotypes while overexpressing plants have enhanced ABA sensitivity in seedling growth and stomata and water conservation capacity. | (Zhu et al. 2007) |
| | Mesophyll protoplasts with constitutively active CPK4 and CPK11 (AJ & CAD domains deleted) | Induction of flg22 response | Constitutively active AtCPK4 and 11 could increase promoter activity of a flg22-responsive gene NHL10-LUC more than ten-fold (CPK4) and more than 15-fold (CPK11) | (Boudsocq et al. 2010) |
| | *cpk4* and *cpk11*(TDI) | | Unaltered flg22 response and pathogen susceptibility in single mutants | |
| | *cpk4/cpk5/cpk6/cpk11* quadruple mutant (virus-induced gene silencing [VIGS]) | | Reduction in oxidative burst induce ed by flg22 and transcript levels of flg22-induced genes PHI-1, NHL10, PER62 and PER4 in *cpk5/cpk6/cpk11* triple and *cpk4/cpk5/cpk6/cpk11* quadruple mutants. Increased susceptibility to *Pseudomonas syringae* pv. *tomato* Pto DC3000 in *cpk5/cpk6/cpk11* triple mutants | |
| | *cpk11* single mutant | Plant defensin gene transcription upon insect attack | *cpk11* mutants did not show significant change in transcript levels of PDF1.2 compared to WT plants after 24 hrs attack with *Spodoptera littoralis* larvae (based on RT-qPCR). | (Kanchiswamy et al. 2010) |

| | | | | |
|---|---|---|---|---|
| AtCPK5 & 6 (*Arabidopsis thaliana*) | Mesophyll protoplasts with constitutively active CPK5 & 6 (AJ & CAD domains deleted) | Induction of flg22 response | Constitutively active CPK5 & 6 could increase promoter activity of a flg22-responsive gene NHL10-LUC almost 20-fold and ten-fold, respectively | (Boudsocq et al. 2010) |
| | *cpk5/cpk6* double & *cpk5/cpk6/cpk11* triple mutants (genetic crosses of TDI lines); *cpk4 cpk5/cpk6/cpk11* quadruple mutant (virus-induced gene silencing [VIGS]) | | Reduction in oxidative burst induced by flg22 and transcript levels of flg22-induced genes PHI-1, NHL10, PER62 and PER4 in double, triple and quadruple mutants <br><br> Increased susceptibility to *Pseudomonas syringae* pv. to*mato* Pto DC3000 in double and triple mutants | |
| | *cpk5* (TDI) <br> *cpk6* (TDI) | | Unaltered flg22 response and pathogen susceptibility | |
| | | Plant defensin gene transcription upon insect attack | *cpk5* mutants did not show significant change in transcript levels of PDF1.2 compared to WT plants after 24 hrs attack with *Spodoptera littoralis* larvae (based on RT-qPCR). | (Kanchiswamy et al. 2010) |
| | *cpk6-1, cpk6-2; cpk3-1/cpk6-1* double mutants | MeJA-induced stomatal closure | MeJA-insensitivity in stomatal closure of *cpk3-1/cpk6-1* double mutants. MeJA-induced stomatal closure, MeJA activation of $I_{Ca}$ channels and S-type anion channels are impaired in *cpk6-1* and *cpk6-2*, while MeJA-induced ROS and NO production are not reduced (experiment done on *cpk1* only). This suggests the role of AtCPK6 as positive regulator of MeJA signalling in Arabidopsis guard cells, downstream of ROS and NO production. | (Munemasa et al. 2011) |
| | *cpk6-1, cpk6-2* single mutants and *cpk3-1/cpk6-1* & *cpk3-2/cpk6-2* double mutants | ABA regulation of guard cell channels & stomatal closure | *cpk3-1/cpk6–1* and *cpk3-2/cpk6-2* double mutant plants showed a slight growth delay (2 d in 4-wk old plants); ABA and $Ca^{2+}$ activation of slow-type anion channels, ABA activation of plasma membrane $Ca^{2+}$-permeable channels, and stomatal closure were impaired in independent alleles of single and double *cpk3/cpk6* mutant guard cells | (Mori et al. 2006) |
| | *AtCPK6* sense mutants (overexpressors, CaMV 35S promoter), and *cpk6-1* (TDI) | Salt & drought tolerance | Leaves of overexpressing lines have higher water capacity than *cpk6* mutants and WT 1 h after harvest. Salt stress (watering with 250 mM NaCl) and drought stress (watering with PEG) for 2 weeks caused death to almost all WT and *cpk6* mutants but >60% of the overexpressors survived. | (Xu et al. 2010) |
| AtCPK20 (*Arabidopsis thaliana*) | *cpk20* (TDI) | Plant defensin gene transcription upon insect attack | *cpk20* mutants did not show significant change in transcript levels of PDF1.2 compared to WT plants after 24 hrs attack with *Spodoptera littoralis* larvae (based on RT-qPCR). | (Kanchiswamy et al. 2010) |
| AtCPK26 (*Arabidopsis thaliana*) | Mesophyll protoplasts with constitutively active CPK26 (AJ & CAD domains deleted) | Induction of flg22 response | Constitutively active CPK 26 could increase promoter activity of a flg22-responsive gene NHL10-LUC more than five-fold. | (Boudsocq et al. 2010) |
| OsCPK7 (*Oryza sativa*) | *OsCPK7* overexpressors (CaMV 35S promoter) with varying expression levels | Cold and salt stress response | Overexpressing plants are cold and salt stress tolerant, unlike wild-type. Tolerance level to cold and salt/drought increases as OsCPK7 expression increases. Overexpression of OsCPK7 enhances induction of stress-related genes rab16A, salT, wsi18 and LEA3 under salt | (Saijo et al. 2000; Saijo et al. 2001) |

| | | | stress but not in cold stress (based on Northern blot analyses) | |
|---|---|---|---|---|
| OsCPK13 (*Oryza sativa*) | *OsCPK13* sense (overexpressors) and antisense mutants (CaMV 35S promoter) | Response to cold, salt, drought & ABA | Sense plants have higher recovery rates after cold treatment. Antisense plants have dwarf phenotypes. | (Abbasi et al. 2004) |
| | | Response to cold | Overexpression of OsCPK13 and calreticulin interacting protein 1 (CRTintP1) in rice confers cold tolerance. | (Komatsu et al. 2007) |
| VvCPK1 (*Vitis vinifera*) | Grape berry VvCPK1 gene overexpressed in Arabidopsis (*SUPER* promoter) | ABA response | Expression of grape berry CPK in Arabidopsis promotes plant growth (faster) and enhances ABA sensitivity in germination, seedling growth and stomatal movement | (Yu et al. 2007) |
| **Group IIa** | | | | |
| AtCPK9 & 33 (*Arabidopsis thaliana*) | *cpk9* (TDI), *cpk33* (TDI) | Plant defensin gene transcription upon insect attack | Single mutants did not show significant change in transcript levels of PDF1.2 compared to WT plants after 24 hrs attack with *Spodoptera littoralis* larvae (based on RT-qPCR). | (Kanchiswamy et al. 2010) |
| AtCPK19 (*Arabidopsis thaliana*) | *cpk19* (TDI) | | | |
| AtCPK21 (*Arabidopsis thaliana*) | *cpk21* (TDI) | | | |
| | *cpk21-1* mutants (TDI) and *CPK21* overexpressors (CaMV 35S promoter) | Hyperosmotic stress response | *cpk21-1* has enhanced tolerance to hyperosmotic stress (300 mM mannitol). Overexpressor mutants show accumulation of DREB1a, COR15A, and Rd29A in 300mM mannitol. | (Franz et al. 2011) |
| AtCPK23 (*Arabidopsis thaliana*) | *cpk23* (TDI) and *CPK23* overexpressor (*SUPER* promoter) | Drought & salt stress response, stomatal movement | *cpk23* showed enhanced tolerance to drought and salt stress, and reduced stomatal apertures. Overexpressing plants show more sensitivity to drought and salt stresses and increased stomatal apertures. | (Ma and Wu 2007) |
| AtCPK22 (*Arabidopsis thaliana*) | *cpk22* (TDI) | Plant defensin gene transcription upon insect attack | *cpk22* mutants did not show significant change in transcript levels of PDF1.2 compared to WT plants after 24 hrs attack with *Spodoptera littoralis* larvae (based on RT-qPCR). | (Kanchiswamy et al. 2010) |
| OsCPK12 (*Oryza sativa*) | *OsCPK12-OX* (CaMV 35S promoter); *oscpk12* retrotransposon insertion (Tos17) mutants and RNAi mutants | Salt stress and blast disease *Magnaporthe grisea* (fungal infection) resistance | Overexpressing plants exhibit increased tolerance to salt stress, less $H_2O_2$ accumulation in leaves, higher expression levels of ROS scavenging enzyme encoding genes, increased sensitivity to ABA (seedling growth inhibition) and increased susceptibility to blast fungus. Loss-of-function mutants (Tos17 and RNAi-silenced) were more sensitive to salinity. | (Asano et al. 2012) |
| **Group IIb** | | | | |
| AtCPK3 (*Arabidopsis thaliana*) | *cpk3-1* (TDI) | MeJA-induced stomatal closure | MeJA-induced stomatal closure is not impaired | (Munemasa et al. 2011) |
| | *cpk3-2* (TDI) | Promotes plant defensin gene transcription upon insect attack | *cpk3-2* mutants show significantly lower transcript levels of PDF1.2 compared to WT plants after 24 hrs attack with *Spodoptera littoralis* larvae (based on RT-qPCR). | (Kanchiswamy et al. 2010) |
| | *cpk3-1, cpk3-2* single mutants and *cpk3-1/cpk6-1* & *cpk3-2/cpk6-2* double mutants (TDI) | ABA regulation of guard cell channels & stomatal closure | *cpk3-1/cpk6–1* and *cpk3-2/cpk6–2* double mutant plants showed a slight growth delay (2 d in 4-wk old plants). ABA and $Ca^{2+}$ activation of slow-type anion channels, ABA activation of plasma membrane $Ca^{2+}$-permeable channels, and stomatal closure were impaired in independent alleles of single | (Mori et al. 2006) |

| | | | and double *cpk3/cpk6* mutant guard cells | |
|---|---|---|---|---|
| | Mesophyll protoplasts with constitutively active CPK3 (AJ & CAD domains deleted) | Induction of flg22 response | Constitutively active CPK 3 could increase promoter activity of a flg22-responsive gene NHL10-LUC more than ten-fold, although the increase may be associated with the high endogenous expression of CPK3 | (Boudsocq et al. 2010) |
| | *cpk3-1, cpk3-2* single mutants (TDI) and *cpk3-3, CPK3-1* & *CPK3-2* overexpressors | Salt stress acclimation independent of MAPK | Under salt stress (150 mM NaCl), germination rate is decreased in *cpk3-1* and *cpk3-2* mutants and increased in *cpk3-3* overexpression mutants. Transcriptional induction of salt stress and MAPK dependent marker genes are not significantly different between wild type, *cpk3-2* mutants, and *CPK3-1* & *CPK3-2* overexpressor lines based on semi-quantitative RT-PCR | (Mehlmer et al. 2010) |
| AtCPK17 & 34 (*Arabidopsis thaliana*) | *cpk17* (TDI) | Plant defensin gene transcription upon insect attack | *cpk17* mutants did not show significant change in transcript levels of PDF1.2 compared to WT plants after 24 hrs attack with *Spodoptera littoralis* larvae (based on RT-qPCR). | (Kanchiswamy et al. 2010) |
| | Transiently overexpressed (particle-bombardment-mediated) and GFP-tagged CPK17 & 34 in tobacco pollen tube | Pollen tube development/ elongation | Transient overexpression of GFP-tagged CPK34 induced depolarization of pollen tube growth (causes reduced elongation), but not CPK17. | (Zhou et al. 2009) |
| | *cpk17, cpk34* single and *cpk17-2/cpk34-2* and *cpk17-5/cpk34-1* double mutants | | Single mutant plants have no detectable phenotypic difference with wild type. *cpk17/cpk34* double mutants have normal pollen tube morphology, but exhibit 350-fold reduction in pollen transmission efficiency, 3-fold reduction in tube growth rate, and >90% failure to locate and fertilize ovules | (Myers et al. 2009) |
| OsCPK02 (*Oryza sativa*) | *OsCPK2* overexpressors (maize ubiquitin-1 promoter, NOS terminator) | Light response | Overexpressing plants are normal but seed development is disrupted. Light exposure represses overexpression. | (Morello et al. 2000) |
| PiCPK1 (*Petunia inflata*) | Transient overexpression of normal (*PiCPK1*), constitutively active (*PiCPK1/CA*) (gene overexpression) and dominant-negative (*PiCPK1/DN*) (gene knockout) constructs by microprojectile bombardment in pollen tubes | Pollen development | *PiCPK1* show loss of growth polarity in transformed tubes, resulting in extremely short tubes with almost spherical tips. *PiCPK1/DN* exhibited loss of growth polarity, but less severely than the wild-type protein. (*PiCPK1/CA*) are severely inhibited in both pollen germination and tube growth (both length and width). | (Yoon et al. 2006) |
| **Group IIIa** | | | | |
| AtCPK24 (*Arabidopsis thaliana*) | Transiently overexpressed (particle-bombardment-mediated) and GFP-tagged CPK24 in tobacco pollen tubes | Pollen tube development/ elongation | Transient overexpression of GFP-tagged CPK24 slightly inhibited pollen tube elongation but had no effect on pollen tube expansion | (Zhou et al. 2009) |
| PiCPK2 (*Petunia inflate*) | Transient overexpression of *PiCPK2* constructs by microprojectile bombardment in pollen tubes | Pollen development | *PiCPK2* show inhibition of pollen tube extension but no effect in growth polarity or germination rates, resulting in short tubes with normal morphology (did not expand). | (Yoon et al. 2006) |
| **Group IIIb** | | | | |
| AtCPK7 & 8 (*Arabidopsis thaliana*) | *cpk7* (TDI); *cpk8* (TDI) | Plant defensin gene transcription upon insect attack | *cpk7* mutants did not show significant change in transcript levels of PDF1.2 compared to WT plants after 24 hrs attack with | (Kanchiswamy et al. 2010) |

| | | | *Spodoptera littoralis* larvae (based on RT-qPCR). | |
|---|---|---|---|---|
| AtCPK10 & 30 *(Arabidopsis thaliana)* | Mesophyll protoplasts with constitutively active CPK 10 & 30 (AJ & CAD domains deleted) | Induction of flg22 response | Constitutively active CPK 10 & 30 could increase promoter activity of a flg22-responsive gene NHL10-LUC more than five-fold. | (Boudsocq et al. 2010) |
| | *cpk10* mutants (TDI), *CPK10* overexpressors (Super1300 vector) | Plant defensin gene transcription upon insect attack | *cpk10* mutants did not show significant change in transcript levels of PDF1.2 compared to WT plants after 24 hrs attack with *Spodoptera littoralis* larvae (based on RT-qPCR) | (Kanchiswamy et al. 2010) |
| | | Drought response | *cpk10* mutants are more sensitive to drought (20-day withholding irrigation on 1-week old seedlings) than WT and insensitive to ABA Induction of stomatal closure and inhibition of stomatal opening. Overexpression mutants have enhanced tolerance to drought. | (Zou et al. 2010) |
| AtCPK13 *(Arabidopsis thaliana)* | Mesophyll protoplasts with constitutively active CPK13 (AJ & CAD domains deleted) | Induction of flg22 response | Constitutively active CPK 13 could increase promoter activity of a flg22-responsive gene NHL10-LUC more than five-fold. | (Boudsocq et al. 2010) |
| | *cpk13-1* and *cpk13-2* | Promotes plant defense genes transcription upon insect attack | *cpk13* mutants show significantly lower transcript levels of PDF1.2 compared to WT plants after 24 hrs attack with *Spodoptera littoralis* larvae (based on RT-qPCR). | (Kanchiswamy et al. 2010) |
| AtCPK14 *(Arabidopsis thaliana)* | Transiently overexpressed (particle-bombardment-mediated), GFP-tagged CPK14 in tobacco pollen tubes | Pollen tube development/ elongation | Transient overexpression of GFP-tagged CPK14 induced depolarization of pollen tube growth (causes reduced elongation). | (Zhou et al. 2009) |
| AtCPK32 *(Arabidopsis thaliana)* | Mesophyll protoplasts with constitutively active CPK32 (AJ & CAD domains deleted) | Induction of flg22 response | Constitutively active CPK32 could increase promoter activity of a flg22-responsive gene NHL10-LUC about five-fold. | (Boudsocq et al. 2010) |
| | Transiently overexpressed (particle-bombardment-mediated) and GFP-tagged CPK32 in tobacco pollen tubes | Pollen tube development/ elongation | Transient overexpression of GFP-tagged CPK32 induced severe depolarization of pollen tube growth (causes reduced elongation). | (Zhou et al. 2009) |
| | *CPK32* overexpressors (CaMV 35S promoter) | ABA signalling and regulation | Overexpression of CPK32 promotes ABA and salt sensitivities during germination. It also regulates the expression of ABF4-regulated genes and ABA responsive genes rd29A, rab18, and rd29B. | (Choi et al. 2005) |
| **Group IV** | | | | |
| AtCPK18 *(Arabidopsis thaliana)* | *cpk18* (TDI) | Plant defensin gene transcription upon insect attack | *cpk18* mutants did not show significant change in transcript levels of PDF1.2 compared to WT plants after 24 hrs attack with *Spodoptera littoralis* larvae (based on RT-qPCR). | (Kanchiswamy et al. 2010) |
| MtCPK1 *(Medicago truncatula)* | RNAi-silenced plants | Root development | RNAi silenced composite plants whose roots have been transformed by A. rhizogenes carrying the pRNAi1444-1 plasmid (henceforth termed CPKi roots) have stunted roots and short root hairs. CPKi roots were impaired in their ability to form a symbiotic association with Glomus versiforme fungus and have altered ROS accumulation in roots and root hairs | (Ivashuta et al. 2005) |

**Appendix 17. Expression analysis of AtCPKs using publicly available microarray data (Chapter 3.3.2)**

(a) Summary of transcript accumulation upon exposure to biotic, abiotic, and phytohormone treatments, based on TAIR (www.arabidopsis.org) and Plant Expression Database (www.plexdb.org) microarray data. Up and down arrows indicate upregulation and down regulation, respectively. Intensity of shading indicates speed of response, with darkest indicating the most rapid response. Empty white boxes means no data was available.

| CPK Name | Evo Grp | Abiotic Stress | | | | | | | | Biotic Stress | | | Phytohormone | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Drought (hyperosmotic stress) | Salt | Osmotic stress (Mannitol) | Cold | Oxidative stress | Heat | UV-B[1] | Wounding | Fungi & elicitors | Bacteria (Pseudomonas) & flg22 | Phytophthora | ABA | GA | IAA | JA/MeJA |
| AtCPK28 | | ↑ | ↑[2] | ↑ | ↑ | = | ? | ↑ | ↑ | = | ↑ | = | = | ↓ | | ↑ |
| AtCPK18 | IV | | | | | | | | | | | | | | | |
| AtCPK16 | | = | = | = | = | = | = | = | = | | | | | | | |
| AtCPK24 | IIIA | | | | | | | | | | | | | | | |
| AtCPK13 | | = | = | = | ↑[1] | ↑ | ? | = | = | = | = | ↑ | = | = | = | = |
| AtCPK30 | | = | = | ↑ | = | = | = | = | = | = | = | = | ↑ | | | |
| AtCPK10 | | ↑ | ↑ | ↑ | ↑ | ↑ | ? | ↑ | ↑ | ↑ | ↑ | ↑ | ↑ | = | = | = |
| AtCPK08 | IIIB | ↑ | ↑ | ↑ | = | ? | = | = | = | ↑ | = | ↑ | = | = | = | ↑ |
| AtCPK07 | | ? | ↑ | = | ? | ? | ? | ↑ | ? | ↑ | ↑ | ↑ | = | = | = | = |
| AtCPK32 | | ↑ | ↑[2] | ↑ | ↑ | = | ↑[3] | ↑ | ↑ | ↑ | ↑ | ? | | | | |
| AtCPK14 | | | | | | | | | | | | | | | | |
| AtCPK25 | | | | | | | | | | | | | | | | |
| AtCPK12 | | = | = | = | = | = | ? | = | = | | | | = | | = | ↑ |
| AtCPK11 | | ↑[2] | ? | ↑ | ↑ | ↑[2] | ↓ | ? | = | = | = | ↑ | = | = | = | ↑ |
| AtCPK04 | | ↑ | ↑ | ↑ | ↑ | = | ↓ | ↑ | ↑ | ↑ | ↑ | ↑ | ↓ | = | ↑ | = |
| AtCPK26 | | = | = | = | = | = | = | ? | = | | | | = | | | = |
| AtCPK06 | I | ↑ | ↑ | ↑ | ↑ | = | ? | ↑ | ↑ | ↑ | ↑ | ↑ | = | = | ↑ | = |
| AtCPK05 | | ? | ↑ | = | ↑ | ? | ↓ | ↑ | = | ↑ | ↑ | | ↓ | = | ↑ | = |
| AtCPK20 | | | = | | | | = | | = | | | | | | | |
| AtCPK02 | | = | ↓ | ↓ | = | ↓ | = | = | ↓ | | | | | | | = |
| AtCPK01 | | ↑ | ↓ | = | ↑ | = | ? | ↓ | ↑ | ↑ | = | ↑ | ↓ | ↑ | | ↓ |
| AtCPK29 | | ↑ | ↑[2] | ↑ | = | = | = | ↑[1] | = | = | ↑ | = | = | | | = |
| AtCPK33 | | = | = | = | = | = | ↑[3] | = | = | = | | | | | | |
| AtCPK09 | | = | ↓ | ↑ | = | = | ↑ | ↑ | = | ↑ | ↑ | ↑ | = | = | = | = |
| AtCPK19 | | | | | | | | | | | | | | | | |
| AtCPK22 | IIA | ? | ↓ | ↓ | = | = | ? | = | = | = | = | = | = | = | = | = |
| AtCPK31 | | | | | | | | | | | | | | | | |
| AtCPK27 | | | | | | | | | | | | | | | | |
| AtCPK15 | | = | ↑[1] | ↑[1] | ↑[1] | = | ↑[3] | = | ↑[1] | ↑ | ↑ | | ↑ | = | | |
| AtCPK23 | | = | = | ↑[1] | ↓ | = | = | ↓[1] | = | = | = | | = | | | |
| AtCPK21 | | = | = | ↓ | ? | = | = | = | = | = | = | | = | | | |
| AtCPK03 | | = | ↓ | ↓ | = | ↓ | ? | ? | = | ↑ | ↑ | ↑ | = | ? | = | |
| AtCPK34 | IIB | | | | | | | | | | | | | | | |
| AtCPK17 | | | | | | | | | | | | | | | | |

| | | |
|---|---|---|
| 0-1hrs | ↑ | ↓ |
| 1-3 | ↑ | ↓ |
| 3-6 | ↑ | ↓ |
| 6-12 | ↑ | ↓ |
| 12-24 | ↑ | ↓ |
| 24-48 | ↑ | ↓ |
| 2 days-1 week | ↑ | ↓ |

(b) Anatomy-specific expression pattern of Arabidopsis CPKs based on publicly available microarray data, generated using the Genevestigator Meta-profile analysis tool (www.genevestigator.com). CPKs are arranged based on their evolutionary groups.
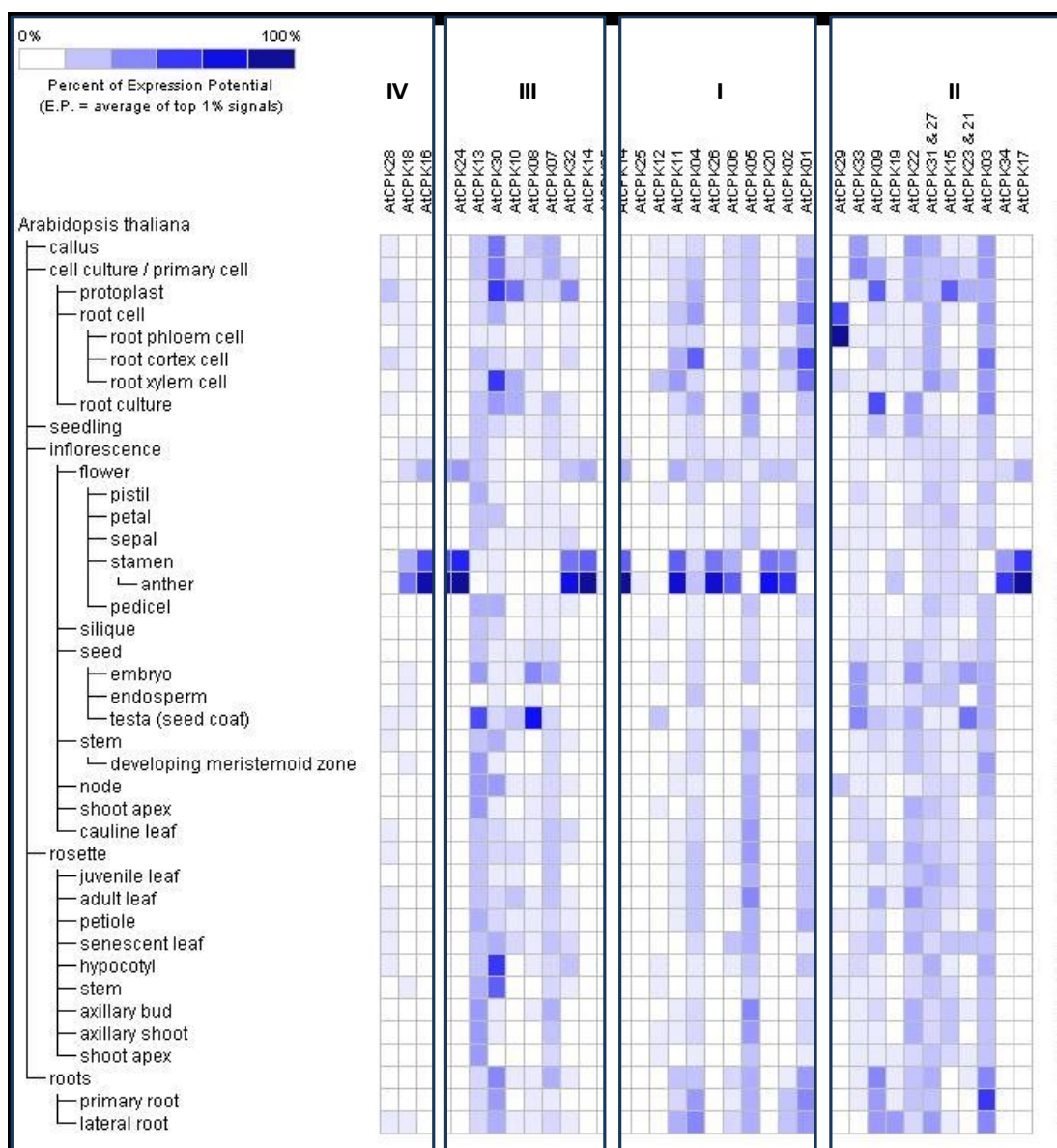
**Appendix 18. Detailed $K_a/K_s$ tree of AtCPKs (Chapter 3.3.3)**

Please see .pdf file in the attached CD

**Appendix 19. $K_a/K_s$ analysis of Arabidopsis CPKs (Chapter 3.3.3)**

A.       Table of $K_a/K_s$ ratios, $K_a$ and $K_s$ values for each node and branch

| Node# | $K_a/K_s$ Branch1 | $K_a$ Branch1 | $K_s$ Branch1 | $K_a/K_s$ Branch2 | $K_a$ Branch2 | $K_s$ Branch2 |
|---|---|---|---|---|---|---|
| 1 | 0.2301 | 0.05957 | 0.2589 | 0.1597 | 0.04222 | 0.2643 |
| 2 | 0.1199 | 0.02909 | 0.2427 | 0.1231 | 0.02702 | 0.2196 |
| 3 | 0.2027 | 0.0752 | 0.3709 | 0.1498 | 0.05951 | 0.3973 |
| 4 | 0.1611 | 0.03345 | 0.2075 | 0.1477 | 0.03179 | 0.2153 |
| 5 | 0.2896 | 0.1037 | 0.3579 | 0.2415 | 0.08975 | 0.3716 |
| 6 | 0.2319 | 0.07989 | 0.3444 | 0.1822 | 0.05561 | 0.3052 |
| 7 | 0.2146 | 0.06649 | 0.3098 | 0.171 | 0.04158 | 0.2431 |
| 8 | 0.1957 | 0.06296 | 0.3217 | 0.1583 | 0.06304 | 0.3983 |
| 9 | 0.1411 | 0.03191 | 0.2262 | 0.5323 | 0.3629 | 0.6817 |
| 10 | 0.1075 | 0.02596 | 0.2415 | 0.1058 | 0.02424 | 0.229 |
| 11 | 0.6682 | 0.3113 | 0.4659 | 0.2562 | 0.09596 | 0.3746 |
| 12 | 0.3841 | 0.1446 | 0.3764 | 0.2257 | 0.08126 | 0.36 |
| 13 | 0.04557 | 0.00862 | 0.1891 | 0.05811 | 0.01269 | 0.2184 |
| 14 | 0.2018 | 0.07836 | 0.3884 | 0.4175 | 0.1795 | 0.43 |
| 15 | 0.2516 | 0.07904 | 0.3142 | 0.1703 | 0.04779 | 0.2807 |
| 16 | 0.1059 | 0.02223 | 0.21 | 0.117 | 0.02577 | 0.2202 |
| 17 | 0.332 | 0.1512 | 0.4553 | 0.1245 | 0.06246 | 0.5019 |
| 18 | 0.2113 | 0.05715 | 0.2705 | 0.108 | 0.03206 | 0.2969 |
| 19 | 0.2425 | 0.06275 | 0.2587 | 0.2502 | 0.07331 | 0.293 |
| 20 | 0.2083 | 0.06663 | 0.3199 | 0.2304 | 0.07109 | 0.3086 |
| 21 | 0.0746 | 0.01943 | 0.2604 | 0.08534 | 0.02031 | 0.238 |
| 22 | 0.3421 | 0.1487 | 0.4346 | 0.2385 | 0.1194 | 0.5008 |
| 23 | 0.3334 | 0.1188 | 0.3563 | 0.1661 | 0.05895 | 0.3548 |
| 24 | 0.298 | 0.1735 | 0.5823 | 0.2039 | 0.05996 | 0.294 |
| 25 | 0.2003 | 0.03967 | 0.1981 | 0.1398 | 0.03231 | 0.2311 |
| 26 | 0.2063 | 0.07453 | 0.3612 | 0.2211 | 0.08274 | 0.3742 |
| 27 | 0.443 | 0.1396 | 0.3152 | 0.1632 | 0.03923 | 0.2403 |
| 28 | 0.6739 | 0.08299 | 0.1232 | 0.1677 | 0.01391 | 0.08291 |
| 29 | 0.1092 | 0.01785 | 0.1635 | 0.1659 | 0.02772 | 0.1671 |
| 30 | 0.5264 | 0.04248 | 0.0807 | 0.7558 | 0.05577 | 0.07379 |
| 31 | 0.1998 | 0.03757 | 0.1881 | 0.4546 | 0.1063 | 0.2338 |
| 32 | 0.1912 | 0.05543 | 0.2899 | 0.1852 | 0.06204 | 0.3349 |
| 33 | 0.3175 | 0.09038 | 0.2846 | 0.3011 | 0.08151 | 0.2707 |

B.    $K_a/K_s$ ratio in each branch leading to CPK sequence

| Group | AtCPK | $K_a/K_s$ ratio |
|-------|-------|-----------------|
| I | AtCPK2 | 0.1411 |
| I | AtCPK1 | 0.1058 |
| I | AtCPK11 | 0.0455 |
| I | AtCPK4 | 0.0581 |
| I | AtCPK12 | 0.2018 |
| I | AtCPK25 | 0.4175 |
| I | AtCPK6 | 0.1059 |
| I | AtCPK5 | 0.117 |
| I | AtCPK20 | 0.332 |
| I | AtCPK26 | 0.1245 |
| IIa | AtCPK29 | 0.3421 |
| IIa | AtCPK19 | 0.298 |
| IIa | AtCPK9 | 0.2039 |
| IIa | AtCPK33 | 0.1398 |
| IIa | AtCPK22 | 0.443 |
| IIa | AtCPK15 | 0.1632 |
| IIa | AtCPK23 | 0.6739 |
| IIa | AtCPK21 | 0.1677 |
| IIa | AtCPK31 | 0.5264 |
| IIa | AtCPK27 | 0.7558 |
| IIb | AtCPK34 | 0.0745 |
| IIb | AtCPK17 | 0.0853 |
| IIb | AtCPK3 | 0.2385 |
| IIIa | AtCPK24 | 0.5323 |
| IIIb | AtCPK14 | 0.2301 |
| IIIb | AtCPK32 | 0.1597 |
| IIIb | AtCPK8 | 0.1199 |
| IIIb | AtCPK7 | 0.1231 |
| IIIb | AtCPK30 | 0.1611 |
| IIIb | AtCPK10 | 0.1477 |
| IIIb | AtCPK13 | 0.2415 |
| IV | AtCPK18 | 0.2146 |
| IV | AtCPK16 | 0.171 |
| IV | AtCPK28 | 0.1583 |

C.    Average $K_a/K_s$ ratios by group

| Group | Mean | Standard deviation | SE |
|-------|------|--------------------|----|
| I | 0.1649 | 0.120287 | 0.038038 |
| IIa | 0.3714 | 0.22141 | 0.070016 |
| IIb | 0.1328 | 0.091727 | 0.052959 |
| IIIa | 0.5323 | 0 | 0 |
| IIIb | 0.1690 | 0.039949 | 0.015099 |
| IV | 0.1813 | 0.029529 | 0.017049 |

**Appendix 20.    Detailed $K_a/K_s$ tree of AtCPKs, OsCPKs, VvCPKs, StCPKs and PpCPKs (Chapter 3.3.3)**

Please see .pdf file in the attached CD

**Appendix 21. $K_a/K_s$ analysis of $K_a/K_s$ analysis of Arabidopsis, rice, grape, potato and moss CPKs (Chapter 3.3.3)**

A.       Table of $K_a/K_s$ ratios, $K_a$ and $K_s$ values for each node and branch

| Node# | $K_a/K_s$ Branch1 | $K_a$ Branch1 | $K_s$ Branch1 | $K_a/K_s$ Branch2 | $K_a$ Branch2 | $K_s$ Branch2 |
|---|---|---|---|---|---|---|
| 1 | 0.1082 | 0.0310 | 0.2860 | 0.1295 | 0.0402 | 0.3104 |
| 2 | 0.1033 | 0.0400 | 0.3873 | 0.1208 | 0.0549 | 0.4547 |
| 3 | 0.1154 | 0.0287 | 0.2486 | 0.0706 | 0.0143 | 0.2028 |
| 4 | 0.2215 | 0.0818 | 0.3695 | 0.1462 | 0.0574 | 0.3927 |
| 5 | 0.1150 | 0.0423 | 0.3679 | 0.1393 | 0.0665 | 0.4772 |
| 6 | 0.0565 | 0.0206 | 0.3641 | 0.0823 | 0.0198 | 0.2407 |
| 7 | 0.0710 | 0.0269 | 0.3787 | 0.0917 | 0.0327 | 0.3561 |
| 8 | 0.0720 | 0.0438 | 0.6079 | 0.1161 | 0.0204 | 0.1761 |
| 9 | 0.1684 | 0.0195 | 0.1160 | 0.1286 | 0.0365 | 0.2841 |
| 10 | 0.0228 | 0.0034 | 0.1511 | 0.0266 | 0.0038 | 0.1444 |
| 11 | 0.1446 | 0.0270 | 0.1870 | 0.0850 | 0.0178 | 0.2096 |
| 12 | 0.0445 | 0.0008 | 0.0172 | 0.0000 | 0.0000 | 0.0067 |
| 13 | 0.1379 | 0.0332 | 0.2406 | 0.2284 | 0.0696 | 0.3047 |
| 14 | 0.0224 | 0.0063 | 0.2820 | 0.0347 | 0.0088 | 0.2540 |
| 15 | 0.0573 | 0.0385 | 0.6715 | 0.0779 | 0.0282 | 0.3618 |
| 16 | 0.0392 | 0.0163 | 0.4175 | 0.0363 | 0.0157 | 0.4327 |
| 17 | 0.0954 | 0.0369 | 0.3871 | 0.1485 | 0.0561 | 0.3776 |
| 18 | 0.2686 | 0.1272 | 0.4734 | 0.1439 | 0.0660 | 0.4586 |
| 19 | 0.1704 | 0.0533 | 0.3130 | 0.0724 | 0.0329 | 0.4540 |
| 20 | 1.1357 | 0.0368 | 0.0324 | 0.1664 | 0.0054 | 0.0326 |
| 21 | 0.0955 | 0.0284 | 0.2978 | 0.1504 | 0.0601 | 0.3993 |
| 22 | 0.6523 | 0.0436 | 0.0669 | 0.5033 | 0.0506 | 0.1005 |
| 23 | 0.1708 | 0.1288 | 0.7543 | 0.3745 | 0.1149 | 0.3069 |
| 24 | 0.5097 | 0.1598 | 0.3134 | 0.1507 | 0.0377 | 0.2503 |
| 25 | 0.1006 | 0.0084 | 0.0839 | 0.7421 | 0.0946 | 0.1275 |
| 26 | 0.0830 | 0.0176 | 0.2122 | 0.0994 | 0.0156 | 0.1569 |
| 27 | 0.2264 | 0.0489 | 0.2162 | 0.1540 | 0.0378 | 0.2455 |
| 28 | 0.1623 | 0.0548 | 0.3376 | 0.1177 | 0.0374 | 0.3172 |
| 29 | 0.0781 | 0.0166 | 0.2130 | 0.0602 | 0.0122 | 0.2029 |
| 30 | 0.1178 | 0.0525 | 0.4459 | 0.1370 | 0.0440 | 0.3208 |
| 31 | 0.1057 | 0.0321 | 0.3039 | 0.1211 | 0.0349 | 0.2884 |
| 32 | 0.1612 | 0.0282 | 0.1750 | 0.1306 | 0.0284 | 0.2176 |
| 33 | 0.0703 | 0.0139 | 0.1981 | 0.1392 | 0.0369 | 0.2652 |
| 34 | 0.0545 | 0.0147 | 0.2701 | 0.0366 | 0.0129 | 0.3535 |
| 35 | 0.1138 | 0.0255 | 0.2242 | 0.0880 | 0.0253 | 0.2879 |
| 36 | 0.1101 | 0.0352 | 0.3193 | 0.1704 | 0.0564 | 0.3309 |
| 37 | 0.1286 | 0.0238 | 0.1849 | 0.1098 | 0.0224 | 0.2036 |
| 38 | 0.2181 | 0.1213 | 0.5561 | 0.0723 | 0.0222 | 0.3068 |
| 39 | 0.0550 | 0.0134 | 0.2442 | 0.0669 | 0.0151 | 0.2255 |
| 40 | 0.1818 | 0.0468 | 0.2572 | 0.1459 | 0.0553 | 0.3787 |
| 41 | 0.1398 | 0.0514 | 0.3678 | 0.1354 | 0.0486 | 0.3591 |
| 42 | 0.1258 | 0.0263 | 0.2087 | 0.0613 | 0.0147 | 0.2405 |
| 43 | 0.1179 | 0.0300 | 0.2543 | 0.0733 | 0.0208 | 0.2842 |
| 44 | 0.0557 | 0.0119 | 0.2142 | 0.0910 | 0.0194 | 0.2132 |
| 45 | 0.0911 | 0.0567 | 0.6226 | 0.0824 | 0.0288 | 0.3494 |
| 46 | 0.0725 | 0.0240 | 0.3308 | 0.0596 | 0.0180 | 0.3021 |
| 47 | 0.0487 | 0.0144 | 0.2953 | 0.0671 | 0.0372 | 0.5545 |
| 48 | 0.0000 | 0.0000 | 0.0000 | 0.8900 | 0.0009 | 0.0000 |
| 49 | 0.0507 | 0.0151 | 0.2974 | 0.0782 | 0.0282 | 0.3606 |
| 50 | 0.0659 | 0.0150 | 0.2277 | 0.0425 | 0.0105 | 0.2471 |
| 51 | 0.2125 | 0.0642 | 0.3020 | 0.2103 | 0.0667 | 0.3172 |
| 52 | 0.1262 | 0.0185 | 0.1463 | 0.0971 | 0.0146 | 0.1507 |
| 53 | 0.3096 | 0.0835 | 0.2698 | 0.1474 | 0.0424 | 0.2874 |
| 54 | 0.2111 | 0.0740 | 0.3507 | 0.2221 | 0.1058 | 0.4763 |
| 55 | 0.0755 | 0.0432 | 0.5722 | 0.1276 | 0.0343 | 0.2686 |
| 56 | 0.1203 | 0.0286 | 0.2378 | 0.0847 | 0.0217 | 0.2558 |
| 57 | 0.0959 | 0.0368 | 0.3840 | 0.1701 | 0.0396 | 0.2326 |
| 58 | 0.1901 | 0.0444 | 0.2336 | 0.1129 | 0.0261 | 0.2307 |
| 59 | 0.0830 | 0.0241 | 0.2898 | 0.0878 | 0.0378 | 0.4302 |
| 60 | 0.1677 | 0.0272 | 0.1621 | 0.0999 | 0.0256 | 0.2562 |
| 61 | 0.0801 | 0.0163 | 0.2028 | 0.0721 | 0.0144 | 0.1991 |
| 62 | 0.3974 | 0.1907 | 0.4800 | 0.1315 | 0.0453 | 0.3446 |
| 63 | 0.0689 | 0.0297 | 0.4313 | 0.0785 | 0.0246 | 0.3131 |
| 64 | 0.0690 | 0.0135 | 0.1958 | 0.0490 | 0.0102 | 0.2072 |
| 65 | 0.1132 | 0.0288 | 0.2548 | 0.1762 | 0.0413 | 0.2344 |
| 66 | 0.0907 | 0.0315 | 0.3468 | 0.1450 | 0.0457 | 0.3150 |
| 67 | 0.1379 | 0.0223 | 0.1618 | 0.1121 | 0.0171 | 0.1529 |
| 68 | 0.0662 | 0.0156 | 0.2361 | 0.1517 | 0.0462 | 0.3042 |
| 69 | 0.0764 | 0.0187 | 0.2447 | 0.0364 | 0.0084 | 0.2309 |

| 70 | 0.0473 | 0.0104 | 0.2192 | 0.0401 | 0.0101 | 0.2518 |
|----|--------|--------|--------|--------|--------|--------|
| 71 | 0.2176 | 0.0501 | 0.2302 | 0.0733 | 0.0183 | 0.2495 |
| 72 | 0.0426 | 0.0167 | 0.3923 | 0.0832 | 0.0246 | 0.2955 |
| 73 | 0.0842 | 0.0233 | 0.2764 | 0.1408 | 0.0373 | 0.2647 |
| 74 | 0.2704 | 0.0747 | 0.2763 | 0.2681 | 0.0762 | 0.2843 |
| 75 | 0.0815 | 0.0150 | 0.1837 | 0.0634 | 0.0122 | 0.1917 |
| 76 | 0.1976 | 0.0511 | 0.2586 | 0.0520 | 0.0145 | 0.2795 |
| 77 | 0.0411 | 0.0168 | 0.4100 | 0.0974 | 0.0340 | 0.3491 |
| 78 | 0.0944 | 0.0175 | 0.1858 | 0.0654 | 0.0126 | 0.1919 |
| 79 | 0.0982 | 0.0457 | 0.4659 | 0.1195 | 0.0465 | 0.3891 |
| 80 | 0.0559 | 0.0207 | 0.3710 | 0.1743 | 0.0668 | 0.3833 |
| 81 | 0.1620 | 0.0627 | 0.3872 | 0.0950 | 0.0283 | 0.2980 |
| 82 | 0.1384 | 0.0442 | 0.3191 | 0.1470 | 0.0498 | 0.3387 |
| 83 | 0.0467 | 0.0135 | 0.2884 | 0.1088 | 0.0342 | 0.3146 |
| 84 | 0.0330 | 0.0102 | 0.3108 | 0.0533 | 0.0172 | 0.3222 |
| 85 | 0.0668 | 0.0037 | 0.0549 | 0.1082 | 0.0395 | 0.3654 |
| 86 | 0.0676 | 0.0017 | 0.0248 | 0.0636 | 0.0037 | 0.0578 |
| 87 | 0.2905 | 0.1314 | 0.4525 | 0.3088 | 0.1166 | 0.3776 |
| 88 | 0.1782 | 0.0703 | 0.3945 | 0.2564 | 0.0715 | 0.2789 |
| 89 | 0.1452 | 0.0389 | 0.2682 | 0.3240 | 0.0806 | 0.2488 |
| 90 | 0.1812 | 0.0829 | 0.4576 | 0.1725 | 0.0837 | 0.4849 |
| 91 | 0.5769 | 0.3353 | 0.5812 | 0.0613 | 0.0421 | 0.6878 |
| 92 | 0.1403 | 0.0428 | 0.3050 | 0.1910 | 0.0623 | 0.3263 |
| 93 | 0.2411 | 0.0788 | 0.3266 | 0.2023 | 0.0778 | 0.3847 |
| 94 | 0.2027 | 0.0597 | 0.2946 | 0.1452 | 0.0456 | 0.3142 |
| 95 | 0.0984 | 0.0360 | 0.3662 | 0.2633 | 0.0969 | 0.3681 |
| 96 | 0.1074 | 0.0196 | 0.1828 | 0.1139 | 0.0163 | 0.1427 |
| 97 | 0.0661 | 0.0149 | 0.2258 | 0.2562 | 0.0479 | 0.1870 |
| 98 | 0.0706 | 0.0277 | 0.3919 | 0.1121 | 0.0359 | 0.3204 |
| 99 | 0.0835 | 0.0282 | 0.3373 | 0.0710 | 0.0261 | 0.3675 |
| 100 | 0.1823 | 0.0662 | 0.3630 | 0.0996 | 0.0399 | 0.4009 |
| 101 | 0.1092 | 0.0233 | 0.2139 | 0.0261 | 0.0071 | 0.2731 |
| 102 | 0.3049 | 0.1351 | 0.4432 | 0.2107 | 0.0985 | 0.4674 |
| 103 | 0.1375 | 0.0503 | 0.3661 | 0.2174 | 0.0627 | 0.2882 |
| 104 | 0.1008 | 0.0405 | 0.4015 | 0.3203 | 0.2185 | 0.6822 |
| 105 | 0.1084 | 0.0256 | 0.2364 | 0.0710 | 0.0216 | 0.3035 |
| 106 | 0.0400 | 0.0090 | 0.2254 | 0.0876 | 0.0149 | 0.1697 |
| 107 | 0.0839 | 0.0310 | 0.3702 | 0.0667 | 0.0196 | 0.2936 |
| 108 | 0.0822 | 0.0340 | 0.4133 | 0.1304 | 0.0311 | 0.2382 |
| 109 | 0.1102 | 0.0263 | 0.2387 | 0.2207 | 0.0536 | 0.2430 |
| 110 | 0.0830 | 0.0288 | 0.3465 | 0.0830 | 0.0318 | 0.3832 |
| 111 | 0.1219 | 0.0261 | 0.2144 | 0.1142 | 0.0229 | 0.2003 |
| 112 | 0.0665 | 0.0202 | 0.3031 | 0.1020 | 0.0309 | 0.3030 |
| 113 | 0.1778 | 0.0734 | 0.4131 | 0.1769 | 0.0665 | 0.3759 |
| 114 | 0.2126 | 0.0647 | 0.3043 | 0.0622 | 0.0241 | 0.3868 |
| 115 | 0.1191 | 0.0324 | 0.2722 | 0.1228 | 0.0329 | 0.2677 |
| 116 | 0.0926 | 0.0445 | 0.4802 | 0.1484 | 0.0397 | 0.2676 |
| 117 | 0.1949 | 0.0542 | 0.2780 | 0.1815 | 0.0560 | 0.3087 |

B.      $K_a/K_s$ ratio in each branch leading to CPK sequence

| Group | CPK | $K_a/K_s$ ratio |
|-------|-----|-----------------|
| I | ATCPK01 [39] | 0.04903954 |
| I | ATCPK02 [40] | 0.06903492 |
| I | ATCPK04 [72] | 0.07225198 |
| I | ATCPK05 [86] | 0.0910302 |
| I | ATCPK06 [87] | 0.05566451 |
| I | ATCPK11 [73] | 0.06693704 |
| I | ATCPK12 [77] | 0.2181 |
| I | ATCPK20 [36] | 0.1315 |
| I | ATCPK26 [82] | 0.09111942 |
| I | GSVIVP000026650 [42] | 0.1517 |
| I | GSVIVP000035380 [41] | 0.06617948 |
| I | GSVIVP000036550 [79] | 0.06131798 |
| I | GSVIVP000271360 [45] | 0.08782005 |
| I | GSVIVP000318120 [76] | 0.1398 |
| I | GSVIVP000362850 [85] | 0.07820747 |
| I | OsCPK05 [80] | 0.08238886 |
| I | OsCPK06 [92] | 0.2107 |
| I | OsCPK07 [64] | 0.1474 |
| I | OsCPK10 [66] | 0.08474447 |
| I | OsCPK11 [100] | 0.191 |
| I | OsCPK13 [81] | 0.05958769 |
| I | OsCPK17 [67] | 0.1203 |

| | | |
|---|---|---|
| I | OsCPK23 [65] | 0.3096 |
| I | OsCPK24 [70] | 0.09705939 |
| I | OsCPK27 [69] | 0.07547422 |
| I | OsCPK28 [71] | 0.1262 |
| I | Pp1s309_91V6 [35] | 0.3049 |
| I | Pp1s364_58V6 [89] | 0.06361761 |
| I | Pp1s364_61V6 [90] | 0.06682301 |
| I | Pp1s49_200V6 [91] | 0.2174 |
| I | Pp1s97_71V6 [88] | 0.1082 |
| I | StCDPK4a [83] | 0.89 |
| I | StCDPK4b [84] | 0 |
| I | StCPK10 [46] | 0.09989551 |
| I | StCPK11 [37] | 0.1762 |
| I | StCPK12 [38] | 0.1132 |
| I | StCPK13 [74] | 0.1098 |
| I | StCPK14 [75] | 0.1286 |
| I | StCPK15 [78] | 0.1354 |
| I | StCPK6 [43] | 0.1121 |
| I | StCPK7 [44] | 0.1379 |
| I | StCPK8 [48] | 0.1677 |
| I | StCPK9 [47] | 0.08302367 |
| IIa | ATCPK09 [23] | 0.0601591 |
| IIa | ATCPK15 [19] | 0.1507 |
| IIa | ATCPK19 [15] | 0.2686 |
| IIa | ATCPK21 [21] | 0.1006 |
| IIa | ATCPK22 [18] | 0.5097 |
| IIa | ATCPK23 [20] | 0.7421 |
| IIa | ATCPK25 [33] | 0.3974 |
| IIa | ATCPK27 [16] | 0.5033 |
| IIa | ATCPK29 [29] | 0.1439 |
| IIa | ATCPK31 [17] | 0.6523 |
| IIa | ATCPK33 [24] | 0.07806677 |
| IIa | GSVIVG010087490 [30] | 0.0723613 |
| IIa | GSVIVG010376520 [27] | 0.1211 |
| IIa | GSVIVP000107520 [28] | 0.1057 |
| IIa | OsCPK12 [34] | 0.3745 |
| IIa | OsCPK19 [22] | 0.137 |
| IIa | StCDPK1-DM [25] | 0.1306 |
| IIa | StCDPK2-DM [26] | 0.1612 |
| IIa | StCPK16 [31] | 0.1664 |
| IIa | StCPK16[114] | 0.08300516 |
| IIa | StCPK28 | |
| IIb | ATCPK03 [49] | 0.1393 |
| IIb | ATCPK17 [59] | 0.02243838 |
| IIb | ATCPK34 [58] | 0.03471783 |
| IIb | GSVIVP000025110 [52] | 0.08233432 |
| IIb | GSVIVP000222860 [60] | 0.07204035 |
| IIb | OsCPK01 [50] | 0.07061928 |
| IIb | OsCPK02 [56] | 0.08502095 |
| IIb | OsCPK14 [57] | 0.1446 |
| IIb | OsCPK15 [51] | 0.1154 |
| IIb | OsCPK25 [54] | 0 |
| IIb | OsCPK26 [55] | 0.04445621 |
| IIb | Pp1s316_13V6 [12] | 0.1208 |
| IIb | Pp1s325_31V6 [13] | 0.1295 |
| IIb | Pp1s96_216V6 [14] | 0.1082 |
| IIb | StCPK17 [62] | 0.1684 |
| IIb | StCPK18 [63] | 0.1161 |
| IIb | StCPK19 [61] | 0.1286 |
| IIb | StCPK20 [53] | 0.115 |
| IIIa | ATCPK24 [93] | 0.2023 |
| IIIa | GSVIVP000019260 [94] | 0.1452 |
| IIIa | OsCPK21 [99] | 0.2564 |
| IIIa | OsCPK29 [96] | 0.3088 |
| IIIa | StCPK25 [95] | 0.2411 |
| IIIb | ATCPK07 [105] | 0.2562 |
| IIIb | ATCPK08 [104] | 0.06607796 |
| IIIb | AtCPK10[112] | 0.1142 |
| IIIb | AtCPK13 [117] | 0.08385611 |
| IIIb | ATCPK14 [103] | 0.2207 |
| IIIb | ATCPK30 [111] | 0.1219 |
| IIIb | ATCPK32 [102] | 0.1102 |
| IIIb | GSVIVG010252490 [108] | 0.1823 |
| IIIb | GSVIVP000255680 [109] | 0.02613261 |

| | | |
|---|---|---|
| IIIb | GSVIVP000388830[113] | 0.08296085 |
| IIIb | GSVIVP000388830[118] | 0.06673855 |
| IIIb | OsCPK03[115] | 0.04000534 |
| IIIb | OsCPK08 [68] | 0.1276 |
| IIIb | OsCPK09 [97] | 0.2905 |
| IIIb | OsCPK16[116] | 0.08760659 |
| IIIb | OsCPK20 [101] | 0.1403 |
| IIIb | OsCPK22 [98] | 0.1782 |
| IIIb | StCPK23 [106] | 0.1139 |
| IIIb | StCPK24 [110] | 0.09958807 |
| IIIb | StCPK27 [107] | 0.1074 |
| IV | ATCPK16 [3] | 0.05199072 |
| IV | ATCPK18 [4] | 0.1976 |
| IV | ATCPK28 [5] | 0.162 |
| IV | GSVIVG010187780 [6] | 0.09495912 |
| IV | OsCPK04 [1] | 0.147 |
| IV | OsCPK18 [2] | 0.1384 |
| IV | Pp1s370_37V6 [10] | 0.06543465 |
| IV | Pp1s83_172V6 [11] | 0.09442316 |
| IV | Pp1s83_8V6 [9] | 0.1195 |
| IV | StCPK21 [7] | 0.06343455 |
| IV | StCPK22 [8] | 0.08147911 |

C. Average $K_a/K_s$ ratios by group

| Group | Mean | Standard deviation | SE |
|---|---|---|---|
| I | 0.1360 | 0.136743 | 0.020853 |
| IIa | 0.2479 | 0.207861 | 0.046479 |
| IIb | 0.0943 | 0.046071 | 0.010859 |
| IIIa | 0.2308 | 0.061207 | 0.027372 |
| IIIb | 0.1258 | 0.06892 | 0.015411 |
| IV | 0.1106 | 0.046305 | 0.013962 |

**Appendix 22. Preparation of culture media and buffers (Chapters 4.2.2, 5.2.5 and 5.2.6)**

A. **½ Murashige and Skoog (½ MS agar)**

| Component | Amount | Final Concentration |
|---|---|---|
| Murashige & Skoog medium including vitamins (Sigma) | 2.2 g | 0.22% (w/v) |
| Sucrose | 20 g | 2% (w/v) |
| Phytagel™ (Sigma) | 2.5 g | 0.25% (w/v) |
| distilled water | To a final volume of 1 L | |

The pH was adjusted to pH 5.7. The ½ MS agar was autoclaved for 15 min at 121°C. Upon cooling to about 55 °C, the media was supplemented with the appropriate antibiotic and poured to set into 98 mm sterile growth pots (Alto Packaging).

B. **Potato Dextrose agar (PDA medium)**

| Component | Amount | Final Concentration |
|---|---|---|
| Potato dextrose agar (Difco) | 39 g | 3.9% (w/v) |
| distilled water | To a final volume of 1 L | |

The PDA medium was autoclaved for 15 min at 121°C. Upon cooling to about 55 °C, the media was poured to set into standard 90 mm petri plates.

C. **Luria-Bertani broth (LB broth)**

| Component | Amount | Final Concentration |
|---|---|---|
| LB broth, Miller (Difco) | 25 g | 2.5% (w/v) |
| distilled water | To a final volume of 1 L | |

The LB broth was autoclaved for 15 min at 121°C and cooled to room temperature upon use.

D. **Virus inoculation buffer**

| Component | Amount | Final Concentration |
|---|---|---|
| Potassium phosphate ($K_2HPO_4$) | 13.6 g | 0.1 M, pH 7.4 at 4°C |
| distilled water | To a final volume of 1 L | |
| anhydrous sodium sulphite ($Na_2SO_3$) | 1 mg per 10 mL added to required volume prior to use (usually 10 to 50 mL) | 0.1% (w/v) |

E. **Kings medium B Agar**

| Component | Amount | Final Concentration |
|---|---|---|
| Proteose peptone (BD) | 20 g | 2% (w/v) |
| Glycerol | 10g | 1% (w/v) |
| Potassium phosphate ($K_2HPO_4$) | 1.5 g | 11 mM |
| distilled water | To a final volume of 1 L | |

The pH was adjusted to pH 7.2. The Kings B medium was autoclaved for 15 min at 121°C. Upon cooling to about 55 °C, sterile 1M $MgSO_4$ was added. The medium was poured to set into standard 90 mm petri plates.

F. **Co-cultivation medium M1**

| Component | Amount | Final Concentration |
|---|---|---|
| Murashige & Skoog medium including vitamins (Sigma) | 2.2 | 0.22% (w/v) |
| Sucrose | 30 g | 3% (w/v) |
| Phytagel™ (Sigma) | 2.5 g | 0.25% (w/v) |
| Zeatin | 3 mg | 3 µg/mL |
| 6-benzylaminopurine | 1 mg | 1 µg/mL |
| acetosyringone | 9.8 mg | 50 µM |
| distilled water | To a final volume of 1 L | |

## G.  Shoot elongation medium M3

| Component | Amount | Final Concentration |
|---|---|---|
| Murashige & Skoog medium including vitamins (Sigma) | 2.2 g | 0.22% (w/v) |
| Sucrose | 20 g | 2% (w/v) |
| Agar (Germantown) | 7 g | 0.7% (w/v) |
| Kanamycin | 100 mg | 100 µg/mL |
| Indolebutrytic acid (IBA) | 0.1 mg | 0.1 µg/mL |
| Timentin | 150 mg | 150 µg/mL |
| distilled water | To a final volume of 1 L per batch | |

## H.  Root elongation medium M4

| Component | Amount | Final Concentration |
|---|---|---|
| Murashige & Skoog medium including vitamins (Sigma) | 2.2 g | 0.22% (w/v) |
| Sucrose | 20 g | 2% (w/v) |
| Agar (Germantown) | 7 g | 0.7% (w/v) |
| Kanamycin | 50 mg | 50 µg/mL |
| Indolebutrytic acid (IBA) | 1 mg | 1 µg/mL |
| Timentin | 150 mg | 150 µg/mL |
| distilled water | To a final volume of 1 L per batch | |

## I.   Hoagland's Medium

| Component | Amount | Final Concentration |
|---|---|---|
| Potassium nitrate ($KNO_3$) | 5 mL of 1 M stock solution | 5 mM |
| Calcium nitrate ($Ca(NO_3)_2$) | 5 mL of 1 M stock solution | 5 mM |
| Potassium phosphate ($K_2HPO_4$) | 1 mL of 1 M stock solution | 1 mM |
| Magnesium sulfate ($MgSO_4$) | 2 mL of 1 M stock solution | 2 mM |
| Micronutrient stock | 1 mL | * |
| Iron EDDHA (Fe $(NO_3)_3$) | 10 mL of 10 mM stock | 0.1 mM |
| distilled water | To a final volume of 1 L per batch | |
| *Micronutrient stock<br>Boric acid (H3BO3)<br>Manganese chloride (MnCl2•H20)<br>Zinc sulfate ZnS04<br>Copper sulfate<br>Molybdic acid | g per 1 L of $H_2O$<br>2.86<br>1.81<br>0.22<br>0.080<br>0.02 | Stock concentration<br>46 mM<br>9 mM<br>0.765 mM<br>0.32 mM<br>0.111 mM |

## J.   Liquid pollen growth medium

| Component | Amount | Final Concentration |
|---|---|---|
| Sucrose | 180 g | 18% (w/v) |
| Boric Acid | 10 g | 0.01% (w/v) |
| Calcium chloride | 1 mL of 1 M stock solution | 1 mM |
| Calcium nitrate | 1 mL of 1 M stock solution | 1 mM |
| Magnesium sulfate | 1 mL of 1 M stock solution | 1 mM |
| distilled water | To a final volume of 1 L per batch | |

**Appendix 23. Information about pathogens used in this thesis (Chapter 4.2.2)**

| Pathogen Scientific name | Common name/ Abbreviation | Lineage (based on EntrezTaxonomy http://www.ncbi.nlm.nih.gov/taxonomy) | Reported host organism(s) | Description/ Symptoms | Reference(s) |
|---|---|---|---|---|---|
| *Pseudomonas syringae pv. tomato* DC3000 | Pto DC3000 or Pto DC3000 Bacterial canker or bacterial blast | Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas | Tomato Arabidopsis | Causes bacterial speck of tomato and yellow spots on Arabidopsis. Is a relatively weak epiphyte but highly aggressive once inside host tissues. A model pathogen for probing disease susceptibility and hormone signalling in plants. Genome completely sequenced. Pathogenicity relies on a type III secretion system (TTSS) to inject virulence effector proteins into host cells. | (Buell et al. 2003; Katagiri et al. 2002; Xin and He 2013) |
| *Botrytis cinerea* | Grey-mould | Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Leotiomycetes; Helotiales; Sclerotiniaceae; Botrytis | Over 200 crops including vegetables and fruits such as cabbages, legumes, grapes and kiwifruit. Arabidopsis | Affects different plant organs causing soft rots, collapse of parenchymal tissue and formation of grey conidial masses. Airborne with a necrotrophic lifestyle. One of the model systems in molecular phytopathology. Genome completely sequenced for strains B05.10 and T4. Pathogenicity involves production of a range of cell-wall-degrading enzymes, toxins and other compounds, and induction of programmed cell death. | (Staats and van Kan 2012; Williamson et al. 2007) |
| *Cauliflower mosaic virus* | CaMV | Viruses; Retro-transcribing viruses; Caulimoviridae; Caulimovirus; Cauliflower mosaic virus | Wide range of cruciferous plants and few | Causes local and/or systemic lesions and chlorosis in leaves. | (Cecchini et al. 1998; Martinière et al. 2009; Melcher |

| | | | | | |
|---|---|---|---|---|---|
| | | | plants in other families. Arabidopsis | In Arabidopsis, symptoms include vein clearing, chlorotic spotting, stunting and decreased seed yield.<br>Distinct pathogenic determinants are associated with particular regions of the viral genome.<br>Pathogenicity include accumulation of subcellular inclusion bodies and translational transactivation.<br>Transmitted by aphids. | 1989; Schoelz and Shepherd 1988) |
| *Tobacco mosaic virus* | TMV | Viruses; ssRNA viruses; ssRNA positive-strand viruses, no DNA stage; Virgaviridae; Tobamovirus | Over 200 species including tobacco and other solanaceous plants. Arabidopsis | Causes local and/or systemic lesions, chlorosis and necrosis in leaves. Also causes stunting, mottling, and leaf curling.<br>Symptoms vary depending on host plant.<br>No visible symptoms in many Arabidopsis ecotypes.<br>Pathogenicity is influenced by the speed of virus cell-to-cell movement and by the induction of pathogenesis-related genes.<br>Transmitted by direct contact or rubbing but not by insects, fungi or nematodes. | (Dardick et al. 2000; Scholthof 2004) |
| *Tomato spotted wilt virus* | TSWV | Viruses; ssRNA viruses; ssRNA negative-strand viruses; Bunyaviridae; Tospovirus | Over 1000 species in 85 families including vegetables, legumes and tobacco. Arabidopsis | Causes stunting, chlorotic or necrotic rings in leaves or fruits and discolouration of seeds.<br>Symptoms vary depending on host plant.<br>Symptoms visible in many Arabidopsis ecotypes.<br>Pathogenicity is influenced by the speed of virus cell-to-cell movement via the plasmodesmata and by RNA silencing suppressor activity<br>Transmitted by thrips. | (Cândido et al. 2006; Sherwood et al. 2003) |
| *Turnip mosaic virus* | TuMV | Viruses; ssRNA viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae; Potyvirus | Wide range of species including Brassicaceae, various crops and ornamentals | Causes stunting, distortion, chlorosis or necrosis in leaves.<br>Symptoms in Arabidopsis include vein clearing.<br>HC-Pro region acts as a suppressor of RNA silencing<br>Transmitted by aphids. | (Kim et al. 2008; Ohshima et al. 2002; Shiboleth et al. 2007) |

| | | | | | |
|---|---|---|---|---|---|
| *Turnip yellow mosaic virus* | TYMV | Viruses; ssRNA viruses; ssRNA positive-strand viruses, no DNA stage; Tymovirales; Tymoviridae; Tymovirus | Narrow host range, restricted to Cruciferae | Causes diffuse chlorotic local lesions and systemic yellow mosaic symptoms in leaves. Symptoms visible in Arabidopsis Characterised by transfer RNA mimicry, RNA replication in chloroplasts and a C-rich genome Transmissible by mechanical inoculation or by flea beetles. | (Dreher 2004) |
| *Magnaporthe grisea* | Rice blast fungus | Eukaryota; Fungi; Dikarya; Ascomycota; Pezizomycotina; Sordariomycetes; Sordariomycetidae; Magnaporthales; Magnaporthaceae; Magnaporthe | Wide range of gramineous plants including rice and other grasses | Causes rupture of cuticle and lesions on all parts of the plant. Infects plants as a spore releases an adhesive, attached to the plant tissue and germinates, producing an appressorium that ruptures the cuticle and allows invasion of the tissue. Resistant cultivars can inhibit growth of this pathogen. | (Dean et al. 2005; TeBeest et al. 2007) |
| *Pseudomonas syringae pv. syringae* | Pss Tomato leaf spot | Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales; Pseudomonadaceae; Pseudomonas | Tomato Arabidopsis | Symptoms often difficult to distinguish from Pto infection | (Gullino et al. 2009) |
| *Cymbidium mosaic virus* | CymMV | Viruses; ssRNA viruses; ssRNA positive-strand viruses, no DNA stage; Tymovirales; Alphaflexiviridae; Potexvirus | Orchids Poaceae | Causes chlorotic to necrotic sunken patches on leaves and necrosis on flowers. Rice species are experimental hosts of this virus | (Hu et al. 1994; Lapierre and Signoret 2004) |
| *Cucumber mosaic virus* | CMV | Viruses; ssRNA viruses; ssRNA positive-strand viruses, no DNA stage; Bromoviridae; Cucumovirus | Over 1200 species in over 100 families of monocots and dicots | May cause severe epinasty, petiole bending, leaf reduction and malformation, plant stunting, and roughness in fruits. Symptoms vary between species and cultivars. May remain symptomless in some. Transmitted by aphid seeds, neighbouring plants | (Zitter and Murphy 2009) |

**Appendix 24. Maps of cloning vectors used in this study (Chapter 4.2.2)**

**A. pDONR™/Zeo (Gateway®)- vector used to make the entry clone for AtCPK3**
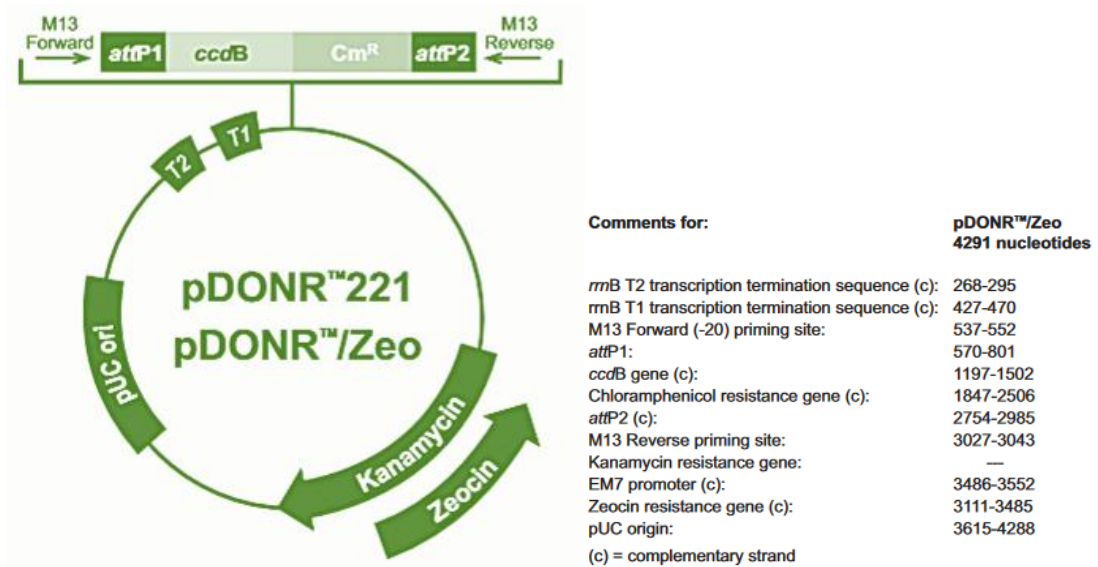


Map taken from the Gateway® pDONR™ Vectors User's guide
(https://tools.thermofisher.com/content/sfs/manuals/gateway_pdonr_vectors.pdf)

**B. pHEX2 (PFR) – used as the destination vector for AtCPK3**
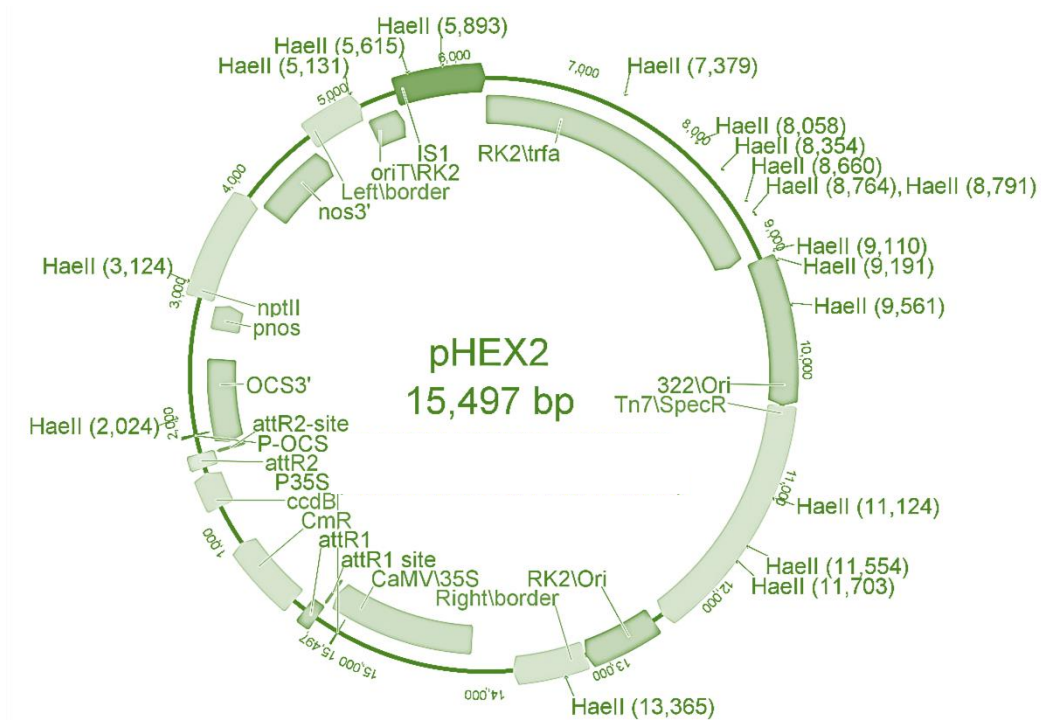


Map obtained from Sakuntala Karunairetnam (Breeding and Genomics Team, PFR)

**C. pSAK778S – binary vector for AcCPK16 overexpression**
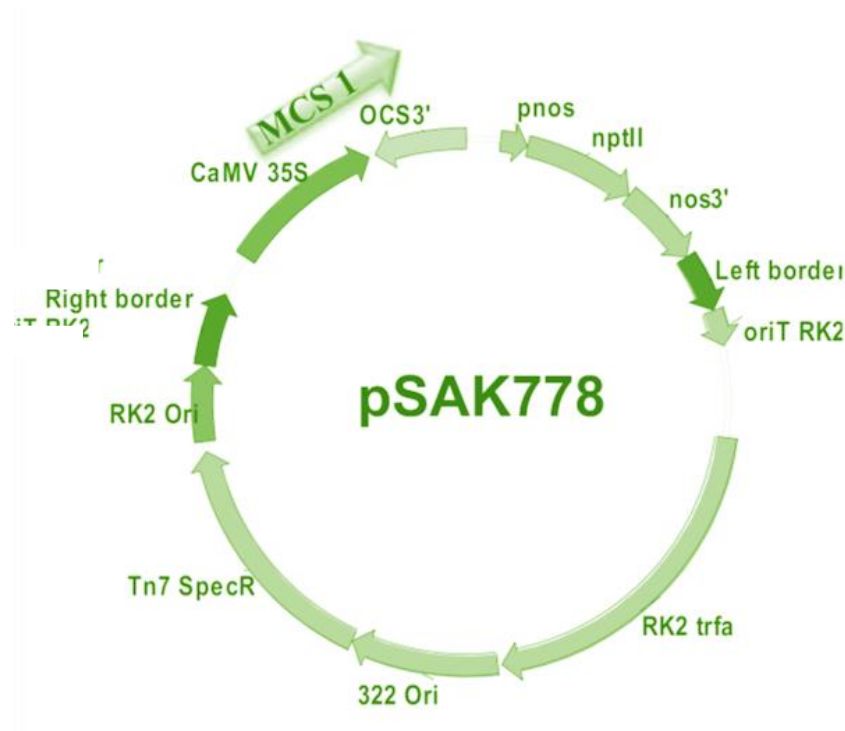


Map obtained from Daisy Wang (Breeding and Genomics Team, PFR)

**D. pTKO2S - binary vector for AcCPK16 knock out**



Map obtained from Daisy Wang (Breeding and Genomics Team, PFR)

**Appendix 25. Phenotype measurement parameters (Chapter 4.2.2)**

Date _____    Day _____    DPI: _____    Data collector: _____

| Virus infection /treatment | Plant Genetic background | Unique Plant Number | Height of primary inflorescence | Number of nodes / secondary inflorescences | Virus symptom severity score | Number of siliques | Lengths of secondary inflorescences* | | | Rosette leaf petiole lengths* | | | Notes: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |
| | | | | | | | | | | | | | |

*Enter a dash or leave the square blank do not enter 0

**Virus symptom severity score**

**0**-no symptoms          **NB: This excludes inoculated leaves, only examine systemic leaves**

**1**-A few leaves showing symptoms

**2**-Most of leaves showing symptoms

**3**-All leaves showing symptoms

**4**-Dead and/or dying

**Please also note types of symptoms observed as appropriate; e.g. yellowing or curling or mosaics etc.**
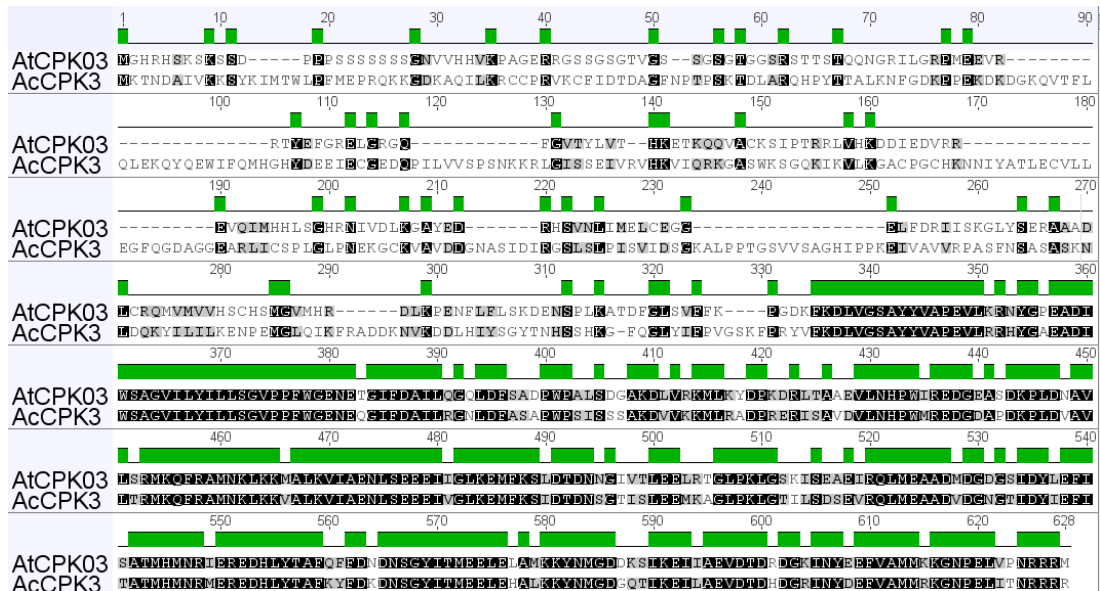
**Appendix 26. Alignment of AcCPK16, AcCPK3 and AtCPK3 (Chapter 4.3.1)**

Black shading indicates amino acid identity. Grey shading indicates amino acid similarity. Absence of shading indicate amino acid difference. Green plots indicate regions of amino acid identity.
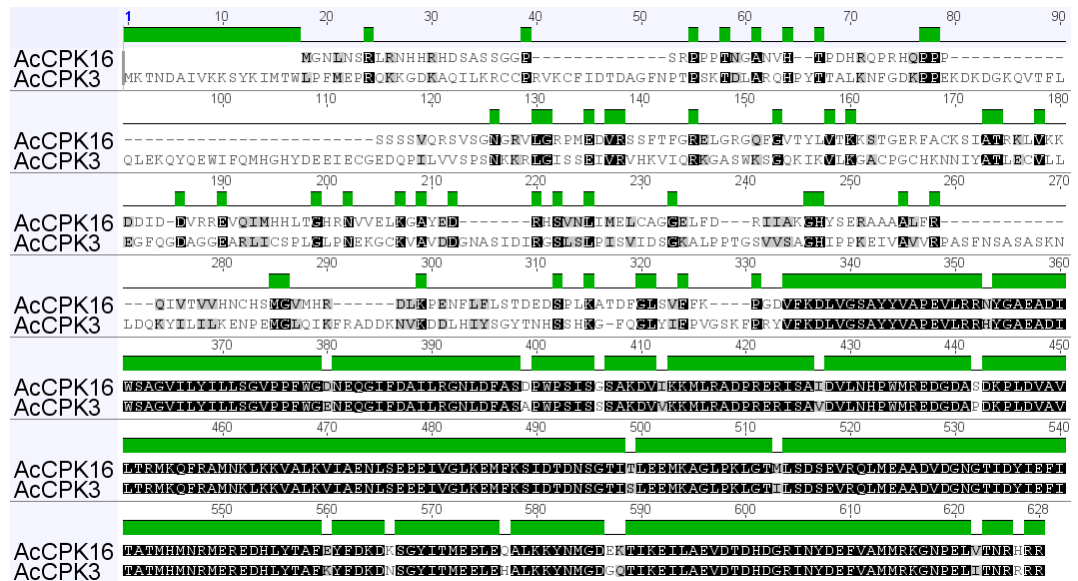
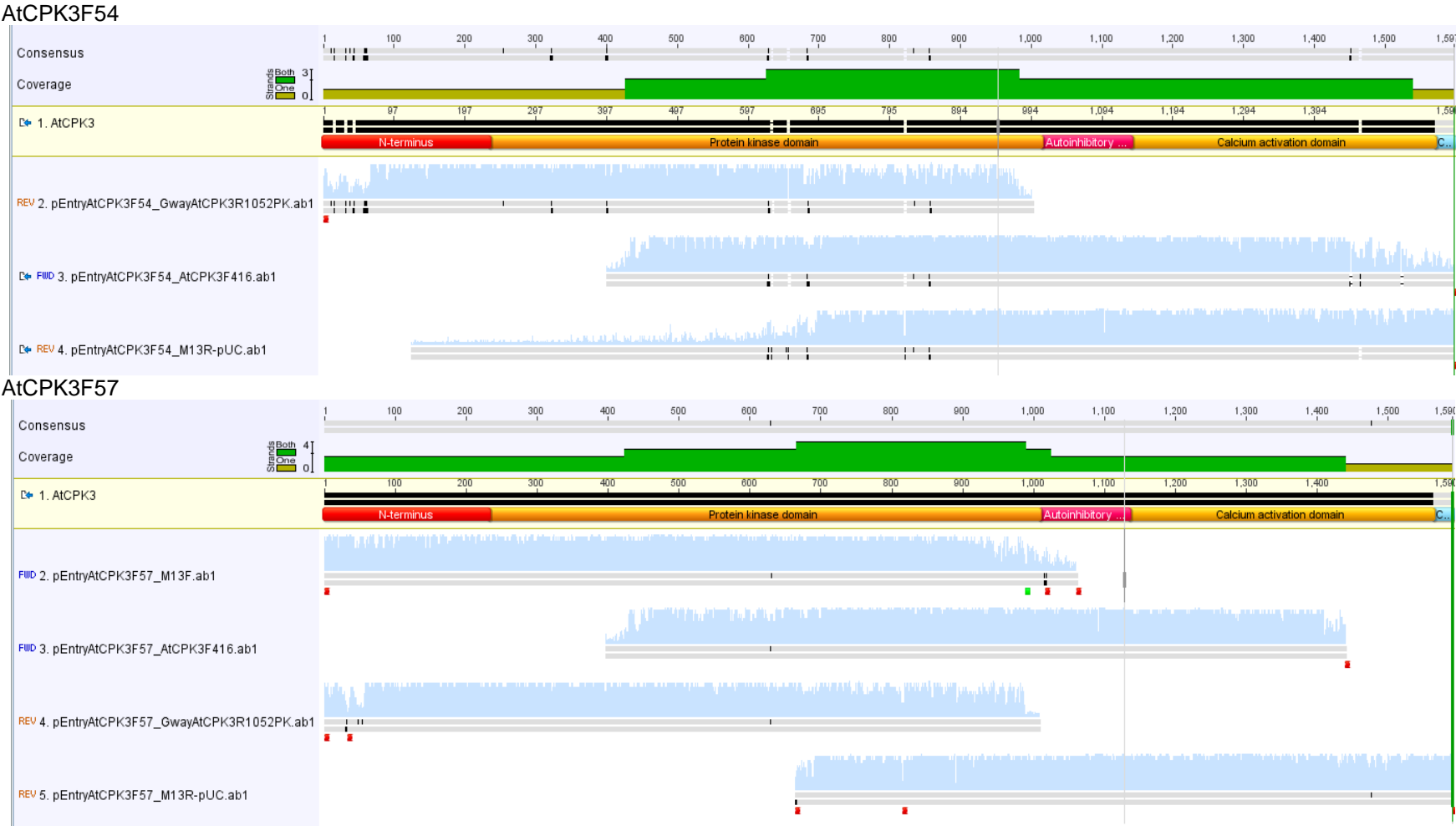A.   Alignment of AtCPK3 and AcCPK16
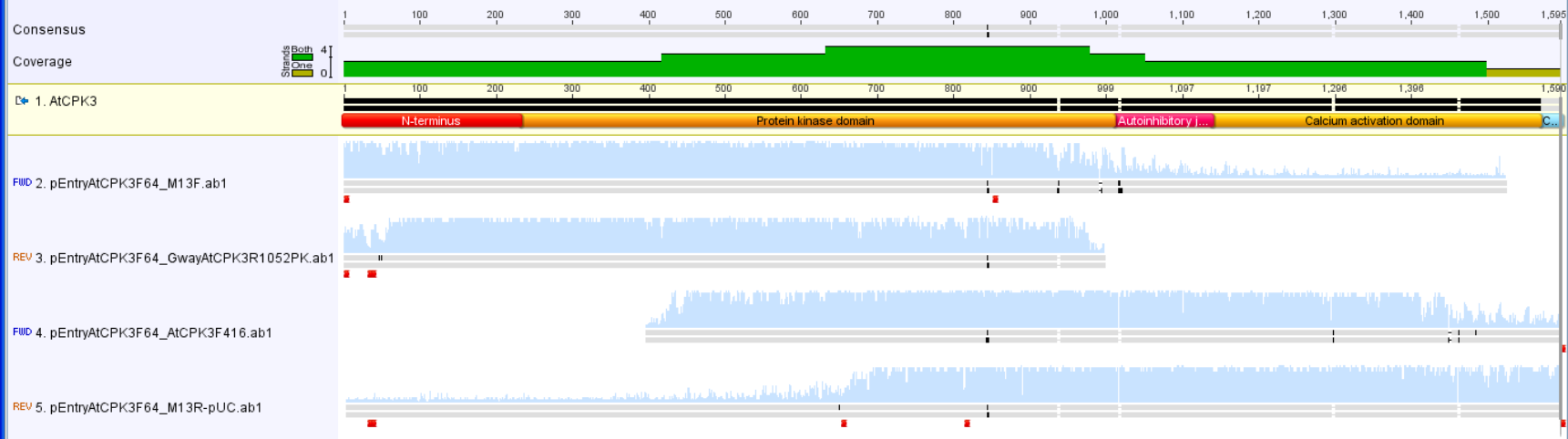


B.   Alignment of AtCPK3 and AcCPK3

C. Alignment of AtCPK16 and AcCPK3
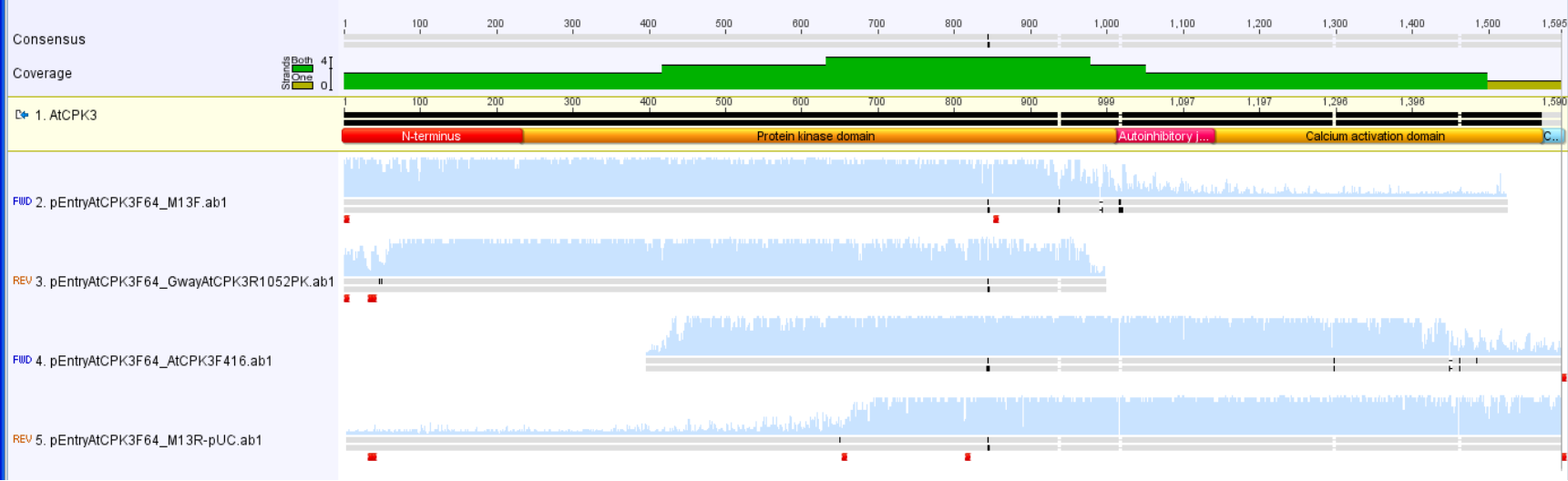
**Appendix 27. Alignment of Gateway AtCPK3 entry clones screened with mismatches to the published sequence. (Chapter 4.3.3)**
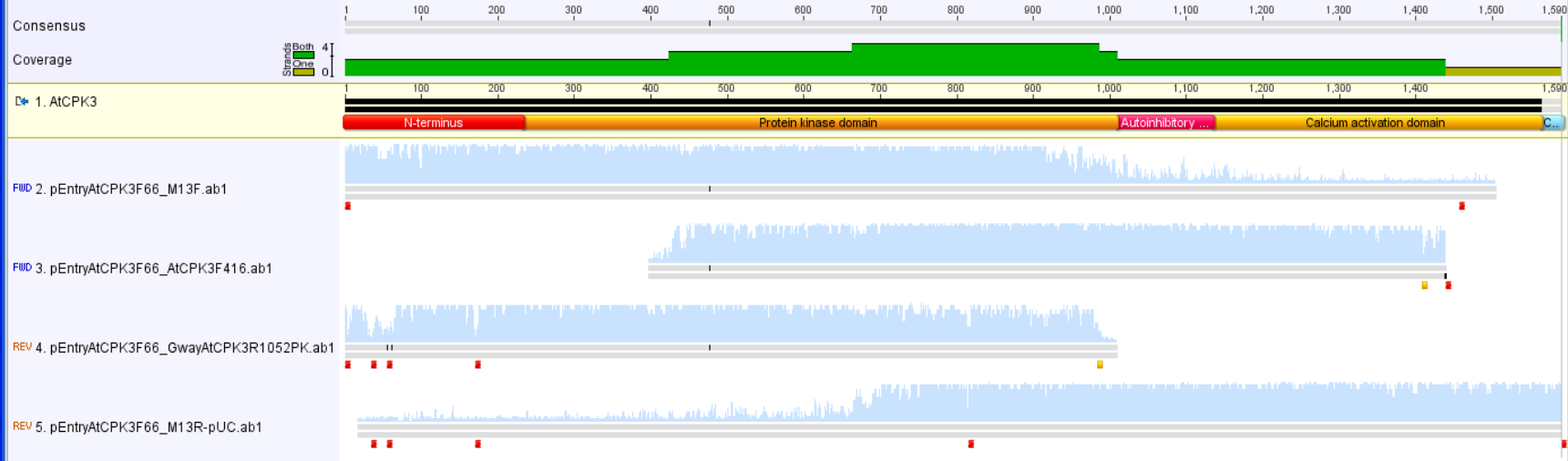
AtCPK3F54
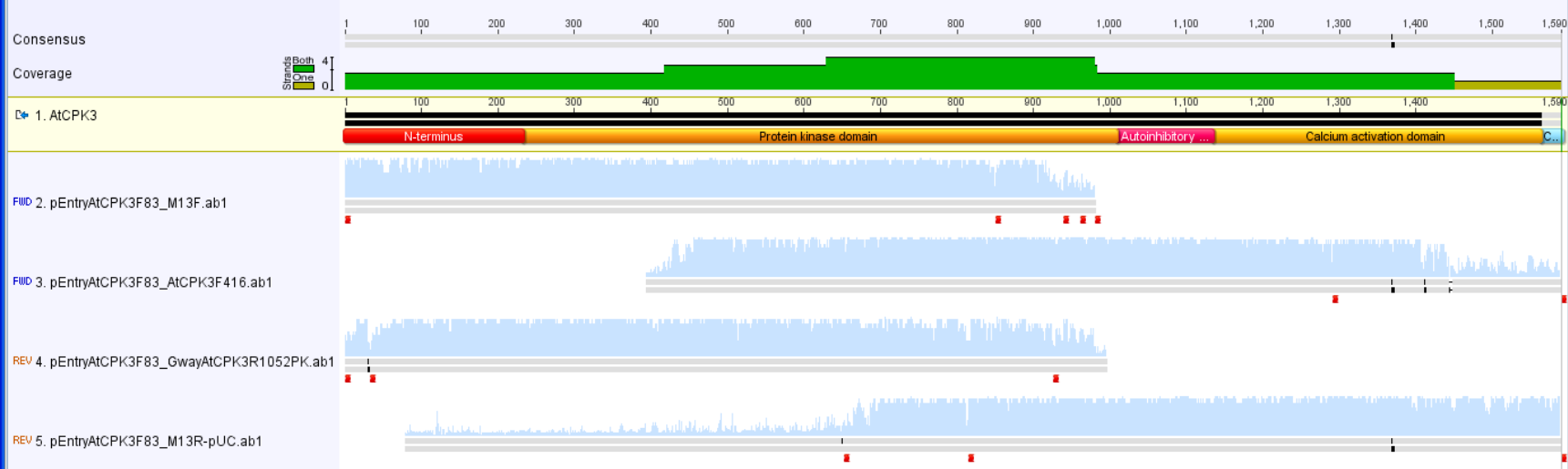


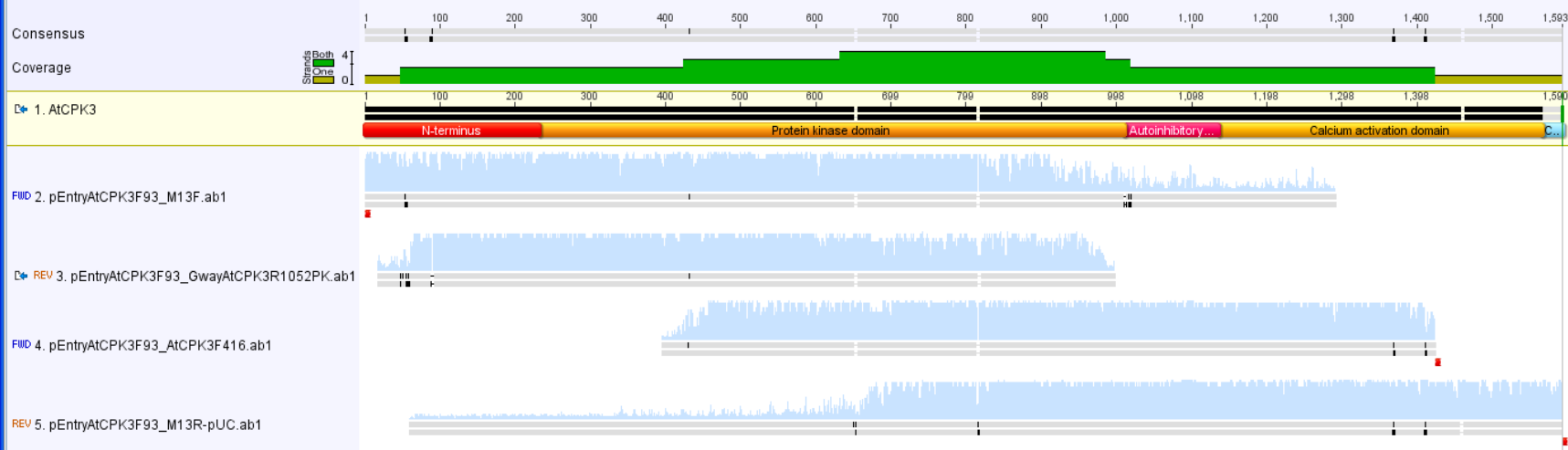AtCPK3F57

## AtCPK3F59



## AtCPK3F64

AtCPK3F66



AtCPK3F83

AtCPK3F93

**Appendix 28. Western blot analysis of AtCPK3 antibody (Chapter 4.3.3)**

The predicted weight for the AtCPK3 protein was 59.34 KDa, but the antibody detected a higher molecular weight proteins at around 70 KDa and lower molecular weight proteins at around 30 KDa. Antibody used was at 1:500x dilution and incubated for 24 hrs. Secondary Antibody was at 1:5000x dilution.



a                                                                b
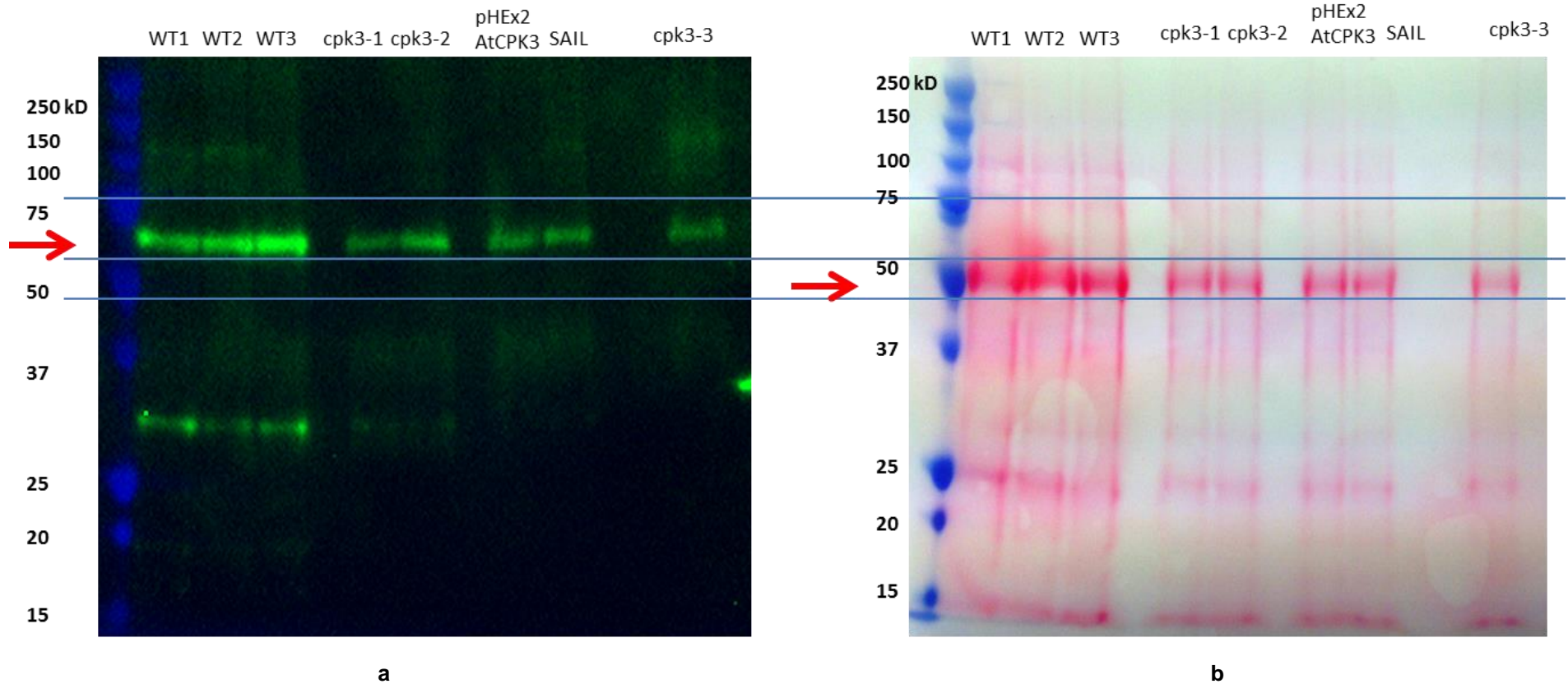
**Appendix 29. Swiss-Model and ITASSER statistical support for predicted tertiary structures of Group IIb CPKs. (Chapter 5.3.1)**

Please find the folder and html files attached in the CD.


**Appendix 30. Predicted tertiary structures of Group IIb AtCPKs and OsCPKs in twelve different angles at 360⁰ rotation (Chapter 5.3.1)**

Please see the .ppt file in the attached CD


**Appendix 31. Group IIb CPKs tertiary structures from Arabidopsis and rice: Pymol graphics system file. (Chapter 5.3.1)**

Please see .pse file in the attached CD

This can be opened using the Pymol software (free download at https://www.pymol.org/)


**Appendix 32. Transcription factor binding sites of AtCPK3, 17 and 34 predicted using MatInspector (Chapter 5.3.2)**

Please see files in the attached CD


**Appendix 33. List of papers that have been published/ will be published from this thesis (Chapter 6)**

1. Valmonte GR, Arthur K, Higgins CM, MacDiarmid RM. 2014. Calcium-dependent protein kinases in plants: Evolution, expression and function. Plant and Cell Physiology. 55(3):551-569. – Findings from Chapter 3
2. Valmonte GR, Lilly, S, Pearson M, Higgins CM, MacDiarmid RM. 2016. Can plants tell you if they are infected viruses? The use of AtSGS3 and AtCPK3 as molecular indicators for plant virus infection. – Findings from Chapter 4
3. Valmonte GR, Arthur K, Higgins CM, MacDiarmid RM. 2016. AtCPK3 and its orthologues in rice and kiwifruit are important in abiotic and biotic stress responses. – Findings from Chapter 4
4. Valmonte GR, Arthur K, Higgins CM, MacDiarmid RM. 2016. What influences CPK functional specificity? A closer look at CPKs belonging to evolutionary group IIb – Findings from Chapter 5

**Appendix 34. List of recent papers that have cited the paper: Valmonte GR, Arthur K, Higgins CM, MacDiarmid RM. 2014. Calcium-dependent protein kinases in plants: Evolution, expression and function. Plant and Cell Physiology. 55(3):551-569. (Chapter 5.3.1)**

Assmann SM, Jegla T. 2016. Guard cell sensory systems: recent insights on stomatal responses to light, abscisic acid, and $CO_2$. Current Opinion in Plant Biology. 33:157-167. doi:10.1016/j.pbi.2016.07.003.

Cakr O, Turgut-Kara N, Arl S, Zhang B. 2015. De novo transcriptome assembly and comparative analysis elucidate complicated mechanism regulating astragalus chrysochlorus response to selenium stimuli. PLoS ONE. 10(10). doi:10.1371/journal.pone.0135677.

Delaux PM, Radhakrishnan GV, Jayaraman D, Cheema J, Malbreil M, Volkening JD, Sekimoto H, Nishiyama T, Melkonian M, Pokorny L et al. 2015. Algal ancestor of land plants was preadapted for symbiosis. Proceedings of the National Academy of Sciences of the United States of America. 112(43):13390-13395. doi:10.1073/pnas.1515426112.

Dubrovina AS, Kiselev KV, Khristenko VS, Aleynova OA. 2015. VaCPK20, a calcium-dependent protein kinase gene of wild grapevine *Vitis amurensis* Rupr., mediates cold and drought stress tolerance. Journal of Plant Physiology. 185:1-12. doi:10.1016/j.jplph.2015.05.020.

Dubrovina AS, Kiselev KV, Khristenko VS, Aleynova OA. 2016. VaCPK21, a calcium-dependent protein kinase gene of wild grapevine *Vitis amurensis* Rupr., is involved in grape response to salt stress. Plant Cell, Tissue and Organ Culture. 124(1):137-150. doi:10.1007/s11240-015-0882-4.

Gupta B, Tripathi AK, Joshi R, Pareek A, Singla-Pareek SL. 2015. Designing climate-smart future crops employing signal transduction components. Elucidation of abiotic stress signalling in plants: functional genomics perspectives, Volume 2. p. 393-413.

Hettenhausen C, Sun G, He Y, Zhuang H, Sun T, Qi J, Wu J. 2016. Genome-wide identification of calcium-dependent protein kinases in soybean and analyses of their transcriptional responses to insect herbivory and drought stress. Scientific Reports. 6. doi:10.1038/srep18973.

Hu Z, Li T, Zheng J, Yang K, He X, Leng P. 2015. Ca2+ signal contributing to the synthesis and emission of monoterpenes regulated by light intensity in *Lilium* 'siberia'. Plant Physiology and Biochemistry. 91:1-9. doi:10.1016/j.plaphy.2015.03.005.

Hui R, El Bakkouri M, Sibley LD. 2015. Designing selective inhibitors for calcium-dependent protein kinases in apicomplexans. Trends in Pharmacological Sciences. 36(7):452-460. doi:10.1016/j.tips.2015.04.011.

Jin X, Yang R, Guo L, Wang X, Yan X, Gu Z. 2017. iTRAQ analysis of low-phytate mung bean sprouts treated with sodium citrate, sodium acetate and sodium tartrate. Food Chemistry. 218:285-293. doi:10.1016/j.foodchem.2016.09.029.

Kanchiswamy CN, Maffei ME. 2015. Calcium signalling preceding the emission of plant volatiles in plant-insect interactions. Journal of the Indian Institute of Science. 95(1):15-23.

Kawamoto N, Sasabe M, Endo M, Machida Y, Araki T. 2015. Calcium-dependent protein kinases responsible for the phosphorylation of a bZIP transcription factor FD crucial for the florigen complex formation. Scientific Reports. 5:8341. doi:10.1038/srep08341.

Long S, Wang Q, Sibley LD. 2016. Analysis of noncanonical calcium dependent protein kinases in Toxoplasma gondii by targeted gene deletion using CRISPR/Cas9. Infection and Immunity. 84(5):1262-1273. doi:10.1128/IAI.01173-15.

Mansour MMF, Salama KHA, Allam HYH. 2015. Role of the Plasma Membrane in Saline Conditions: Lipids and Proteins. Botanical Review. 81(4):416-451. doi:10.1007/s12229-015-9156-4.

Marco F, Bitrián M, Carrasco P, Rajam MV, Alcázar R, Tiburcio AF. 2015. Genetic engineering strategies for abiotic stress tolerance in plants. Plant Biology and Biotechnology: Volume II: Plant Genomics and Biotechnology. p. 579-609.

Miransari M. 2016. Soybean, protein, and oil production under stress. Environmental stresses in soybean production: Soybean Production. p. 157-176.

Mohanta TK, Mohanta N, Mohanta YK, Bae H. 2015. Genome-wide identification of calcium dependent protein kinase gene family in plant lineage shows presence of novel D-x-D and D-E-L motifs in EF-hand domain. Frontiers in Plant Science. 6(DEC). doi:10.3389/fpls.2015.01146.

Plattner H. 2015. Molecular aspects of calcium signalling at the crossroads of unikont and bikont eukaryote evolution - The ciliated protozoan Paramecium in focus. Cell Calcium. 57(3):174-185. doi:10.1016/j.ceca.2014.12.002.

Plattner H, Verkhratsky A. 2015. The ancient roots of calcium signalling evolutionary tree. Cell Calcium. 57(3):123-132. doi:10.1016/j.ceca.2014.12.004.

Ranty B, Aldon D, Cotelle V, Galaud JP, Thuleau P, Mazars C. 2016. Calcium sensors as key hubs in plant responses to biotic and abiotic stresses. Frontiers in Plant Science. 7(MAR2016). doi:10.3389/fpls.2016.00327.

Simeunovic A, Mair A, Wurzinger B, Teige M. 2016. Know where your clients are: Subcellular localization and targets of calcium-dependent protein kinases. Journal of Experimental Botany. 67(13):3855-3872. doi:10.1093/jxb/erw157.

Wang JL, Huang SY, Zhang NZ, Chen J, Zhu XQ. 2015a. Genome-wide expression patterns of calcium-dependent protein kinases in Toxoplasma gondii. Parasites and Vectors. 8(1). doi:10.1186/s13071-015-0917-z.

Wang JP, Munyampundu JP, Xu YP, Cai XZ. 2015b. Phylogeny of plant calcium and calmodulin-dependent protein kinases (CCaMKs) and functional analyses of tomato CCaMK in disease resistance. Frontiers in Plant Science. 6(DEC). doi:10.3389/fpls.2015.01075.

Wang M, Zhang X, Liu JH. 2015c. Deep sequencing-based characterization of transcriptome of trifoliate orange (*Poncirus trifoliata* (L.) Raf.) in response to cold stress. BMC Genomics. 16(1). doi:10.1186/s12864-015-1629-7.

Wei M, Wang S, Dong H, Cai B, Tao J. 2016. Characterization and comparison of the CPK gene family in the apple (*Malus* x *domestica*) and other rosaceae species and its response to *Alternaria alternata* infection. PLoS ONE. 11(5). doi:10.1371/journal.pone.0155590.

Zebelo SA, Maffei ME. 2015. Role of early signalling events in plant-insect interactions. Journal of Experimental Botany. 66(2):435-448. doi:10.1093/jxb/eru480.

Zhu X, Dunand C, Snedden W, Galaud JP. 2015. CaM and CML emergence in the green lineage. Trends in Plant Science. 20(8):483-489. doi:10.1016/j.tplants.2015.05.010.

**Appendix 35. Statistical Analyses for qPCR, phenotype analysis, seed and pollen germination experiments (Chapter 4.3.2, 4.3.3, 5.2.5 and 5.2.6)**

Please see files in the attached CD.